

Karl-Rudolf Koch

**Parameterschätzung
und
Hypothesentests
in linearen Modellen**

Vierte, bearbeitete Auflage

Bonn 2004

<http://www.geod.uni-bonn.de>

Ehemals Ferd. Dummlers Verlag, Bonn

©2004 Professor Dr.-Ing., Dr.-Ing. E.h. mult. Karl-Rudolf Koch
Institut für Theoretische Geodäsie
der Rheinischen Friedrich-Wilhelms-Universität Bonn
Nußallee 17
D-53115 Bonn
e-mail: koch@theor.geod.uni-bonn.de

Vorwort zur vierten Auflage

Für die Veröffentlichung dieses Buches im Internet wurden bekannt gewordene Fehler beseitigt. Außerdem wurde auf einige neuere Entwicklungen durch zusätzliche Literaturstellen hingewiesen.

Bonn, im November 2004

Karl-Rudolf Koch

Vorwort zur dritten Auflage

Für die notwendig gewordene dritte Auflage wurde dieses Buch noch einmal vollständig überarbeitet. Ein Kapitel über robuste Parameterschätzungen wurde hinzugefügt und der Abschnitt über die Diskriminanzanalyse fortgelassen, da dieses Thema umfassender in dem Buch des Autors „Bayesian Inference with Geodetic Applications“, Springer-Verlag, Berlin, 1990, behandelt wird. Einige Beispiele wurden aufgenommen, um die Darstellung anschaulicher zu gestalten. Kleinere Ergänzungen oder Streichungen wurden vorgenommen, um den Text zu verbessern oder auf neuere Entwicklungen hinzuweisen, und bekannt gewordene Fehler sind beseitigt worden.

Ich danke den Studentinnen und Studenten unserer Universität, die Vorschläge zur Verbesserung dieses Buches gemacht haben. Danken möchte ich auch den Mitarbeiterinnen und Mitarbeitern des Instituts für Theoretische Geodäsie, die bei der Neuauflage geholfen haben. Mein besonderer Dank gilt Herrn Dipl.-Ing. Robert Blinken, der eine Reihe von Anregungen für die Neuauflage gegeben hat, und Frau Ingrid Wahl für das Schreiben und die Gestaltung des Textes und das Zeichnen der Abbildungen dieser dritten Auflage. Schließlich möchte ich noch die gute Zusammenarbeit mit dem Verlag dankend erwähnen.

Bonn, im März 1997

Karl-Rudolf Koch

Vorwort zur zweiten Auflage

Die Notwendigkeit, dieses Buch in einer zweiten Auflage herauszugeben, wurde dazu genutzt, es vollständig zu überarbeiten. Dabei wurde erneut angestrebt, die Methoden der Parameterschätzung, der Hypothesenprüfung und der Bereichsschätzung möglichst in sich geschlossen darzustellen und das erforderliche Wissen über die lineare Algebra und die Wahrscheinlichkeitstheorie beizufügen. Der Aufbau des Buches ist also nicht geändert worden. Eine Reihe von Beispielen wurde aber hinzugefügt, um die durchweg knapp gehaltene Darstellung der Methoden aufzulockern. Wo der Text zu verbessern oder wo auf neuere Entwicklungen hinzuweisen war, sind Ergänzungen oder auch Streichungen vorgenommen worden, und bekannt gewordene Fehler wurden beseitigt.

An umfangreicheren Ergänzungen sind die direkte Berechnung der Cholesky-Faktorisierung zu nennen, die Eigenschaften einer speziellen symmetrischen reflexiven generalisierten Inversen, die nichtzentralen Wishart-Verteilungen für die Matrizen quadratischer Formen, die Interpretation der Restriktionen in einem freien Netz, die Datumstransformationen, das Gauß-Markoff-Modell für Ausreißer in den Beobachtungen, die vereinfachten iterierten Schätzungen der Varianz- und Kovarianzkomponenten, die weniger empfindlichen Hypothesentests im univariaten und multivariaten Modell der Parameterschätzung, die Berechnung der Wahrscheinlichkeit des Fehlers 2. Art für Tests im multivariaten Modell, der Konfidenzbereich für mehrere lineare Funktionen der unbekannt Parameter, die Ableitung des Ausreißertests als Spezialfall des Tests einer allgemeinen Hypothese im univariaten Modell, die Zuverlässigkeitsmaße und ihre Interpretation.

Allen Studentinnen und Studenten, die mit Verbesserungsvorschlägen zu dieser Auflage beigetragen haben, danke ich sehr. Mein Dank gilt auch den Mitarbeiterinnen und Mitarbeitern des Instituts für Theoretische Geodäsie, die an der Vorbereitung der Neuauflage mitgewirkt haben. Besonders danke ich Frau Dipl.-Ing. Stefanie Schulte, die viele Anregungen zu dieser Auflage gegeben hat, und Frau Karin Bauer, die den größten Teil der Reinschrift besorgte. Schließlich möchte ich dankend noch die gute Zusammenarbeit mit dem Verlag erwähnen.

Bonn, im Juli 1986

Karl-Rudolf Koch

Vorwort (zur ersten Auflage)

Dieses Buch soll die Methoden der Parameterschätzung, der Hypothesenprüfung und der Bereichsschätzung erläutern und begründen. Es behandelt also die statistische Schlußfolgerung oder die statistische Inferenz für Parameter. Um einen großen Kreis von Lesern unterschiedlichen Ausbildungsstandes und verschiedener Fachdisziplinen ansprechen zu können, wurde das Buch so angelegt, daß zu seinem Verständnis außer einigen Grundkenntnissen in der Analysis keinerlei Voraussetzungen benötigt werden. Die erforderlichen Kenntnisse aus der linearen Algebra und der Wahrscheinlichkeitstheorie werden in den beiden ersten Abschnitten des Buches vermittelt. Da es das Ziel ist, im Hinblick auf Anwendungen die Inferenz für Parameter darzustellen, wurde größerer Wert darauf gelegt, das Verständnis für die behandelten Verfahren zu vermitteln, als in den gegebenen Definitionen und Beweisen möglichst allgemein zu sein.

Die Schätzungen und Hypothesenprüfungen erfolgen in linearen Modellen, doch bedeutet dies keine Einschränkung der Allgemeinheit, da unter Voraussetzungen, die meistens bequem zu erfüllen sind, die nichtlinearen Modelle, wie gezeigt wird, in lineare Modelle überführt werden können. Andererseits bieten die linearen Modelle den Vorzug, daß man sich bei ihnen der Methoden der linearen Algebra bedienen kann. Die hierzu erforderlichen Definitionen und Sätze werden im Abschnitt 1 behandelt, wobei die angeführten Sätze bis auf wenige Ausnahmen bewiesen werden. Abschnitt 1 enthält auch die generalisierten Inversen, die für die Parameterschätzung in Modellen mit nicht vollem Rang benötigt werden, und die Projektionen, die zur geometrischen Interpretation der Schätzungen dienen.

Da außer der Parameterschätzung, die in univariaten und multivariaten Modellen erfolgt, auch die Bereichsschätzung, die Hypothesenprüfung und die Diskriminanzanalyse behandelt werden, befaßt sich Abschnitt 2 mit der Wahrscheinlichkeitstheorie. Hier werden die Zufallsvariablen eingeführt und die wichtigsten univariaten und multivariaten Verteilungen sowie die Testverteilungen für die univariaten und multivariaten Modelle der Parameterschätzung abgeleitet. Jede Verteilung erhält eine Methode zur numerischen Berechnung ihrer Verteilungsfunktion, so daß darauf verzichtet werden kann, Tafeln der Verteilungsfunktionen abzudrucken.

Der Abschnitt 3 behandelt in univariaten und multivariaten Modellen die Schätzung von Parametern, die als feste Größen definiert sind. Feste Parameter gemeinsam mit Zufallsparametern werden in gemischten Modellen geschätzt. Im Abschnitt 3 wird auch auf die Varianzanalyse und auf die Schätzung von Varianz- und Kovarianzkomponenten eingegangen. Der Abschnitt 4 widmet sich dann der Hypothesenprüfung, der Bereichsschätzung und dem Ausreißertest, während der Abschnitt 5 schließlich einen kurzen Überblick über die Diskriminanzanalyse gibt.

Benötigt man für Beweise Sätze, die zuvor behandelt worden sind, so wird mit Hilfe der entsprechenden Nummern auf diese Sätze verwiesen. Der Abschnitt 1 oder auch der Abschnitt 2 brauchen daher nicht vor dem Studium der übrigen Kapitel gelesen zu werden, aufgrund der Verweise läßt sich das fehlende Wissen gezielt den beiden ersten Abschnitten entnehmen. Wenn bei den als Literatur zitierten Büchern eine Seitenzahl angegeben wird, so bezeichnet sie nur die erste Seite dessen, was von Interesse ist. Auf die folgenden Seiten, die ebenfalls von Bedeutung sein können, wird nicht besonders hingewiesen.

Allen Mitarbeiterinnen und Mitarbeitern des Instituts für Theoretische Geodäsie, die zum Erscheinen dieses Buches beigetragen haben, danke ich sehr. Mein besonderer Dank gilt Herrn Dipl.-Math., Dipl.-Ing. Burkhard Schaffrin, der viele Anregungen zu diesem Buch gegeben hat. Schließlich möchte ich noch die gute Zusammenarbeit mit dem Verlag während der Entstehung des Buches dankend erwähnen.

Bonn, im September 1979

Karl-Rudolf Koch

Inhaltsverzeichnis

Einführung	1
1 Vektor- und Matrixalgebra	3
11 Mengen und Körper	3
111 Mengenbegriff	3
112 Verknüpfung von Mengen	4
113 Relationen	5
114 Körper der reellen Zahlen	6
12 Vektoralgebra	6
121 Vektordefinition und Vektorraum	6
122 Lineare Abhängigkeit und Basis eines Vektorraums	9
123 Skalarprodukt und Euklidischer Raum	12
124 Orthogonale Unterräume	13
13 Matrizen	15
131 Matrixdefinition und Matrixverknüpfungen	15
132 Rang einer Matrix	21
133 Berechnung inverser Matrizen	26
134 Matrizenidentitäten	36
135 Spaltenraum und Nullraum einer Matrix	37
136 Determinanten	39
137 Spur einer Matrix und Darstellung einer Matrix als Vektor	43
14 Quadratische Formen	45
141 Transformationen	45
142 Eigenwerte und Eigenvektoren	48
143 Definite Matrizen	50
15 Generalisierte Inversen	52
151 Rechts- und Linksinversen	52
152 Idempotente Matrizen	53
153 Generalisierte Inverse, reflexive generalisierte Inverse und Pseudoinverse	54
154 Lineare Gleichungssysteme	59
155 Generalisierte Inversen symmetrischer Matrizen	62
156 Eigenschaften der Pseudoinversen und einer speziellen symmetrischen reflexiven generalisierten Inversen	67
16 Projektionen	69
161 Allgemeine Projektionen	69

	162	Orthogonale Projektionen	70
17		Differentiation und Integration von Vektoren und Matrizen . .	72
	171	Extrema von Funktionen	72
	172	Differentialquotienten spezieller Funktionen	74
	173	Integration und Variablentransformation	78
2		Wahrscheinlichkeitstheorie	81
21		Wahrscheinlichkeit	81
	211	Einführung	81
	212	Zufällige Ereignisse	83
	213	Axiome der Wahrscheinlichkeit	85
	214	Bedingte Wahrscheinlichkeit und Bayessche Formel . . .	86
	215	Unabhängige Ereignisse	88
22		Zufallsvariable	88
	221	Definition	88
	222	Verteilungsfunktion	89
	223	Diskrete und stetige Zufallsvariable	91
	224	Binomialverteilung und Poisson-Verteilung	93
	225	Mehrdimensionale stetige Zufallsvariable	96
	226	Randverteilung	98
	227	Bedingte Verteilung	99
	228	Unabhängige Zufallsvariable	100
	229	Transformation von Verteilungen	101
23		Erwartungswerte und Momente von Zufallsvariablen	101
	231	Erwartungswert	101
	232	Multivariate Momente	104
	233	Kovarianzmatrix, Fehlerfortpflanzungsgesetz, Korrela- tionsmatrix und Gewichtsmatrix	107
	234	Momenterzeugende Funktion	115
24		Univariate Verteilungen	116
	241	Normalverteilung	116
	242	Herleitung der Normalverteilung als Verteilung von Be- obachtungsfehlern	120
	243	Gammaverteilung	121
	244	Herleitung der Gammaverteilung als Verteilung von An- kunftszeiten	124
	245	Betaverteilung	125
25		Multivariate Normalverteilung	127
	251	Definition und Herleitung	127
	252	Momenterzeugende Funktion der Normalverteilung . . .	129
	253	Randverteilung und bedingte Verteilung	130
	254	Unabhängigkeit normalverteilter Zufallsvariablen	132

255	Lineare Funktionen normalverteilter Zufallsvariablen . . .	133
256	Summe normalverteilter Zufallsvariablen	134
26	Testverteilungen für univariate Modelle der Parameterschätzung	134
261	χ^2 -Verteilung	134
262	Nichtzentrale χ^2 -Verteilung	137
263	F -Verteilung	138
264	Nichtzentrale F -Verteilung	141
265	t -Verteilung	143
27	Quadratische Formen	144
271	Erwartungswert und Kovarianz	144
272	Verteilung der quadratischen Form	146
273	Unabhängigkeit zweier quadratischer Formen	147
274	Unabhängigkeit einer linearen Form und einer quadratischen Form	147
28	Testverteilungen für multivariate Modelle der Parameterschätzung	148
281	Wishart-Verteilung	148
282	Herleitung der Wishart-Verteilung	149
283	Verteilung der Summe von Wishart-Matrizen	151
284	Verteilung der transformierten Wishart-Matrix	152
285	Verteilung der Matrizen quadratischer Formen und Unabhängigkeit der Wishart-Matrizen	152
286	Verteilung des Verhältnisses der Determinanten zweier Wishart-Matrizen	154
287	Verteilung spezieller Funktionen von Wishart-Matrizen	158
3	Parameterschätzung in linearen Modellen	161
31	Methoden der Parameterschätzung	162
311	Punktschätzung	162
312	Beste erwartungstreue Schätzung	163
313	Methode der kleinsten Quadrate	164
314	Maximum-Likelihood-Methode	165
32	Gauß-Markoff-Modell	166
321	Definition und Linearisierung	166
322	Beste lineare erwartungstreue Schätzung	169
323	Methode der kleinsten Quadrate	171
324	Maximum-Likelihood-Methode	174
325	Erwartungstreue Schätzung der Varianz der Gewichtseinheit	175
326	Numerische Berechnung der Schätzwerte und ihrer Kovarianzen	179

	327	Gauß–Markoff–Modell mit Restriktionen	184
	328	Rekursive Parameterschätzung	192
	329	Abweichungen vom Modell	193
33		Gauß–Markoff–Modell mit nicht vollem Rang	196
	331	Methode der kleinsten Quadrate und Maximum–Likelihood–Schätzung	196
	332	Schätzbare Funktionen	198
	333	Projizierte Parameter als schätzbare Funktion	200
	334	Gauß–Markoff–Modell mit nicht vollem Rang und Restriktionen	209
34		Spezielle Gauß–Markoff–Modelle	214
	341	Polynommodell	214
	342	Varianzanalyse	217
	343	Parameterschätzung für die Varianzanalyse mit einer symmetrischen reflexiven generalisierten Inversen	221
	344	Kovarianzanalyse	225
	345	Gauß–Markoff–Modell für Ausreißer in den Beobachtungen	226
35		Verallgemeinerte lineare Modelle	228
	351	Regressionsmodell	228
	352	Gemischtes Modell	231
	353	Beste lineare erwartungstreue Schätzung im gemischten Modell	233
	354	Methode der kleinsten Quadrate und Maximum–Likelihood–Methode für das gemischte Modell	235
	355	Modell der Ausgleichung nach bedingten Beobachtungen	239
	356	Prädiktion und Filterung	240
36		Schätzung von Varianz- und Kovarianzkomponenten	245
	361	Beste invariante quadratische erwartungstreue Schätzung	245
	362	Lokal beste Schätzung	249
	363	Iterierte Schätzungen	254
	364	Beste erwartungstreue Schätzung der Varianz der Gewichtseinheit	258
37		Multivariate Parameterschätzung	259
	371	Multivariates Gauß–Markoff–Modell	259
	372	Schätzung der Parametervektoren	261
	373	Schätzung der Kovarianzmatrix	263
	374	Numerische Berechnung der Schätzwerte und unvollständige multivariate Modelle	267
	375	Spezielles Modell zur Schätzung von Kovarianzmatrizen und Schätzung von Kovarianzen für stochastische Prozesse	272

	376	Multivariates Modell mit Restriktionen	275
38		Robuste Parameterschätzung	277
	381	Wahl der Zielfunktion	277
	382	Robuste M-Schätzung	278
	383	M-Schätzung nach Huber	281
	384	L_p -Norm Schätzung	284
	385	Hebelpunkte	286
	386	Modifizierte M-Schätzung nach Huber	289
	387	Verfahren von Rousseeuw	292
4		Hypothesenprüfung, Bereichsschätzung und Ausreißertest	295
41		Verteilungen aufgrund normalverteilter Beobachtungen	296
	411	Verteilungen von Funktionen der Residuen im univariaten Modell	296
	412	Verteilungen der im multivariaten Modell geschätzten Kovarianzmatrizen	299
42		Test von Hypothesen	301
	421	Methode der Hypothesenprüfung und Likelihood-Quotiententest	301
	422	Test einer allgemeinen Hypothese im univariaten Gauß-Markoff-Modell	304
	423	Spezielle Hypothesen	308
	424	Hypothesentest für die Varianz der Gewichtseinheit	312
	425	Test einer allgemeinen Hypothese im multivariaten Gauß-Markoff-Modell	314
	426	Hypothese der Identität einer Kovarianzmatrix mit einer gegebenen Matrix	321
43		Bereichsschätzung	322
	431	Konfidenzintervalle	322
	432	Konfidenzintervalle für Parameter, für lineare Funktionen der Parameter und Konfidenzhyperellipsoide	323
	433	Konfidenzintervall für die Varianz der Gewichtseinheit	328
44		Ausreißertest	329
	441	Hypothesentest für Ausreißer	329
	442	Test für einen Ausreißer	331
	443	Fraktilwerte für den Ausreißertest	333
	444	Zuverlässigkeitsmaße	335
		Literatur	339
		Sachverzeichnis	355

Einführung

Parameter müssen immer dann geschätzt werden, wenn bestimmte Abläufe, Zustände oder Ereignisse beobachtet werden, um daraus Erkenntnisse und Schlüsse, beispielsweise über künftige Entwicklungen der beobachteten Ereignisse, zu ziehen. Die Parameterschätzung benötigt man also für die mathematische Modellierung der beobachteten Phänomene. Die Beobachtungen sind Funktionen der unbekannt Parameter, wobei die Art der Funktion sich aus einer physikalischen Gesetzmäßigkeit, aus geometrischen Zusammenhängen oder einfach aus dem Experiment ergibt, das den Beobachtungen zugrunde liegt. Die funktionale Abhängigkeit zwischen den Beobachtungen und den unbekannt Parametern bildet einen Teil des Modells, in dem die Parameterschätzung erfolgt.

Die Beobachtungen für die Parameterschätzung stellen die Ergebnisse von Zufallsexperimenten dar. Die Beobachtungen, wie zum Beispiel Messungen der Lufttemperatur, sind also von zufälliger Natur und können mit einer bestimmten Wahrscheinlichkeit innerhalb gewisser Grenzen schwanken. Angaben über das Maß dieser Schwankungen mit Hilfe der Varianzen und Kovarianzen der Beobachtungen bilden den zweiten Teil des Modells für die Parameterschätzung. Um den Einfluß der Zufälligkeit der Beobachtungen auf die Parameter gering zu halten, wird davon ausgegangen, daß im allgemeinen mehr Beobachtungen in die Parameterschätzung eingeführt werden, als zur eindeutigen Bestimmung der Parameter notwendig ist.

Häufig wird zur Erfassung eines Phänomens nicht nur ein Merkmal gemessen, wie beispielsweise die Größe einer Pflanze bei einem Pflanzenversuch, sondern man beobachtet mehrere Merkmale, wie Größe, Gewicht und Aufbau einer Pflanze. Die Auswertung dieser Daten geschieht in einem multivariaten Modell, während im univariaten Modell lediglich ein Merkmal analysiert wird. Je nach Aufgabenstellung definiert man die unbekannt Parameter als feste Größen oder wie die Beobachtungen als Zufallsvariable, was eine Schätzung in unterschiedlichen Modellen bedingt.

Die Aufgabe der Parameterschätzung besteht darin, in den gegebenen Modellen beste Schätzwerte für die unbekannt Parameter zu bestimmen, wobei der Begriff der besten Schätzung als Schätzung mit minimaler Varianz definiert wird. Bei groben Fehlern in den Beobachtungen kann als weitere Eigenschaft gefordert werden, daß die Schätzungen robust gegenüber diesen

Ausreißern sind. Häufig interessieren nicht nur die Schätzwerte der Parameter, sondern es besteht auch der Wunsch, Informationen über die Parameter, die man zusätzlich zu den Beobachtungen besitzt, zu prüfen. Dies geschieht in den Hypothesentests. Weiter lassen sich mit Hilfe der Bereichsschätzungen für die unbekannt Parameter in Abhängigkeit einer vorgegebenen Wahrscheinlichkeit Intervalle angeben, in denen die Parameter liegen.

Das gesamte Aufgabengebiet, das im folgenden behandelt wird, läßt sich kurz als statistische Inferenz für Parameter charakterisieren.

1 Vektor- und Matrixalgebra

Die hier behandelte statistische Inferenz für Parameter soll in linearen Modellen erfolgen. Lineare Beziehungen lassen sich kompakt und übersichtlich durch Vektoren und Matrizen darstellen, so daß im folgenden auf die später benötigten Definitionen und Sätze der linearen Algebra eingegangen wird. Erläutert werden auch die Methoden der Vektorräume, die es erlauben, geometrische Vorstellungen dann noch zu benutzen, wenn die Räume, die benötigt werden, höhere Dimensionen als die des unserer Vorstellung geläufigen dreidimensionalen Raumes besitzen. Schließlich werden noch generalisierte Inversen behandelt, mit denen man bequem von Modellen mit vollem Rang für die Parameterschätzung auf Modelle mit nicht vollem Rang überwechseln kann.

11 Mengen und Körper

111 Mengenbegriff

Im folgenden werden häufig Objekte oder Vorkommnisse mit gleichen oder ähnlichen Eigenschaften behandelt, die auf irgendeine Weise zusammengefaßt werden müssen. Dies geschieht zweckmäßig mit dem mathematischen Begriff der *Menge*. Unter einer Menge versteht man daher die Zusammenfassung genau definierter wirklicher oder gedachter Objekte. Die zu einer Menge gehörenden Objekte sind die *Elemente* der Menge. Es sei a ein Element und M eine Menge; man schreibt

$$a \in M \quad \text{oder} \quad a \notin M ,$$

falls a ein Element oder falls a kein Element der Menge M ist. Die Menge M der Elemente a , für die die Eigenschaft B bezüglich a , also $B(a)$ erfüllt ist, bezeichnet man mit

$$M = \{a | B(a)\} ,$$

beispielsweise $M = \{a | a \text{ ist positiver Teiler von } 6\} = \{1, 2, 3, 6\}$. Gibt es in M kein Element mit der Eigenschaft $B(a)$, führt man die leere Menge $M = \emptyset$ ein.

Definition: Die Menge, die kein Element enthält, heißt *leere Menge* \emptyset . (111.1)

Wichtige Beispiele für Mengen sind die Zahlenmengen wie die Mengen natürlicher Zahlen, ganzer Zahlen oder rationaler Zahlen, also endlicher oder unendlicher, periodischer Dezimalbrüche. Hierfür haben sich Standardbezeichnungen eingebürgert; beispielsweise bezeichnet man mit \mathbb{N} die Menge der natürlichen Zahlen, also $\mathbb{N} = \{1, 2, 3, \dots\}$, und mit \mathbb{R} die Menge der reellen Zahlen, also $\mathbb{R} = \{a \mid a \text{ ist endlicher oder unendlicher Dezimalbruch mit beliebigem Vorzeichen}\}$.

Für Elemente, die nicht nur Elemente einer Menge sind, besteht die

Definition: Gehören alle Elemente einer Menge M auch einer Menge P an, so bezeichnet man M als *Teilmenge* von P und schreibt $M \subset P$. (111.2)

112 Verknüpfung von Mengen

Wie beispielsweise reelle Zahlen durch Rechenoperationen verknüpft werden können, so lassen sich auch Mengen miteinander verknüpfen.

Definition: Die *Vereinigungsmenge* $M \cup P$ (M vereinigt mit P) zweier Mengen M und P besteht aus den Elementen, die wenigstens einer der beiden Mengen angehören. (112.1)

Definition: Die *Durchschnittsmenge* $M \cap P$ (M geschnitten mit P) zweier Mengen M und P besteht aus den Elementen, die sowohl der Menge M als auch der Menge P angehören. (112.2)

Definition: Die *Differenzmenge* $M \setminus P$ (M ohne P) zweier Mengen M und P besteht aus den Elementen von M , die nicht zugleich noch P angehören. (112.3)

Eine anschauliche Darstellung von Vereinigung, Durchschnitt und Differenz zweier Mengen vermögen die sogenannten *Venn-Diagramme* zu geben, in denen die Bereiche innerhalb geschlossener Linien die Elemente einer Menge repräsentieren. In Abbildung 112-1 ist ein Venn-Diagramm der Vereinigungsmenge $M \cup P$ von M und P , in Abbildung 112-2 der Durchschnittsmenge $M \cap P$ von M und P und in Abbildung 112-3 der Differenzmenge $M \setminus P$ von M und P jeweils schraffiert dargestellt.

Gilt $M \cap P = \emptyset$, heißen M und P *disjunkt*. Ist $A \subset B$, so bezeichnet man die Menge \bar{A}

$$\bar{A} = B \setminus A \quad (112.4)$$

als *Komplementärmenge* von A in B . Für sie gilt $A \cup \bar{A} = B$ und $A \cap \bar{A} = \emptyset$.

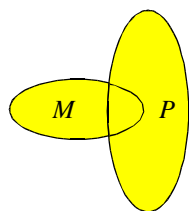


Abb. 112-1:
Vereinigungsmenge

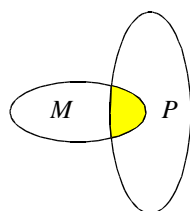


Abb. 112-2:
Durchschnittsmenge

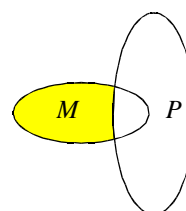


Abb. 112-3:
Differenzmenge

113 Relationen

Nachdem die Zusammenfassung von Objekten aufgrund bestimmter Eigenschaften als Menge eingeführt wurde, müssen jetzt Beziehungen zwischen den Objekten charakterisiert werden. Hierzu wird der folgende Begriff benötigt. Sind a und b irgendwelche Objekte, so nennt man den Ausdruck (a, b) das *geordnete Paar* a, b . Zu seiner Erklärung soll lediglich definiert werden, wann zwei geordnete Paare übereinstimmen: Es gilt $(a, b) = (c, d)$ genau dann, wenn $a = c$ und $b = d$ ist. Die Ausdrucksweise „genau dann, wenn“ bedeutet, daß die Folgerung nach beiden Seiten zu ziehen ist. Aus $(a, b) = (c, d)$ folgt also $a = c$ und $b = d$ und umgekehrt aus $a = c$ und $b = d$ ergibt sich $(a, b) = (c, d)$. Mit Hilfe des geordneten Paares kann jetzt die folgende Relation zwischen zwei Mengen definiert werden.

Definition: Die Menge aller geordneten Paare (a, b) mit $a \in A$ und $b \in B$ heißt *kartesisches Produkt* der Mengen A und B , und man schreibt $A \times B = \{(a, b) | a \in A, b \in B\}$. (113.1)

Geordnete Paare lassen sich als *Koordinaten* interpretieren, so daß das kartesische Produkt $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ der Menge \mathbb{R} der reellen Zahlen die Punkte in einer Ebene ergibt, deren Koordinaten mit Hilfe zweier reeller *Koordinatenachsen* definiert werden. Die Erweiterung des Paarbegriffes führt auf das *Tripel*, das sind drei geordnete Objekte, auf das *Quadrupel* mit vier Objekten und schließlich auf das *n-Tupel* mit n Objekten, beispielsweise (x_1, x_2, \dots, x_n) . Die Definition des Tripels, Quadrupels und n -Tupels erfolgt rekursiv mit Hilfe des Paarbegriffes, indem die erste Koordinate des Tripels ein geordnetes Paar ist, die erste Koordinate des Quadrupels ein Tripel und sofort. Entsprechend (113.1) ergeben sich n -Tupel als kartesisches Produkt von n Mengen. \mathbb{R}^3 definiert daher den *dreidimensionalen Raum* und \mathbb{R}^n den *n-dimensionalen Raum*.

Weitere Relationen wie Äquivalenzrelationen und Ordnungsrelationen (GROTEMEYER 1970) brauchen hier nicht definiert zu werden.

114 Körper der reellen Zahlen

Nachdem Relationen zwischen den Elementen von Mengen behandelt wurden, werden jetzt Verknüpfungen zwischen den Elementen eingeführt. Die Verknüpfungen bleiben hier auf die Addition und Multiplikation reeller Zahlen beschränkt, so daß der Körper der reellen Zahlen erhalten wird. Dieser Körper besitzt aufgrund der Verknüpfungen eine algebraische Struktur (GROTEMEYER 1970, S.37).

Definition: Die Menge \mathbb{R} heißt *Körper* der reellen Zahlen, wenn mit $x, y, z \in \mathbb{R}$ die Addition und Multiplikation das *Kommutativ-, Assoziativ- und Distributivgesetz* erfüllen

$$\begin{aligned} x + y &= y + x & \text{und} & & xy &= yx \\ x + (y + z) &= (x + y) + z & \text{und} & & x(yz) &= (xy)z \\ x(y + z) &= xy + xz, \end{aligned}$$

wenn es zwei ausgezeichnete Elemente 0 (Null) und 1 (Eins) mit $0 \neq 1$ gibt, so daß für jedes $x \in \mathbb{R}$ gilt $x + 0 = x$ und $1x = x$, wenn zu jedem $x \in \mathbb{R}$ ein additives inverses Element $y \in \mathbb{R}$ existiert, so daß $x + y = 0$ gilt, und wenn es zu jedem $x \in \mathbb{R}'$ mit $\mathbb{R}' = \mathbb{R} \setminus \{0\}$ ein multiplikatives inverses Element $z \in \mathbb{R}'$ gibt, so daß $xz = 1$ gilt. (114.1)

Im folgenden wird unter der Menge \mathbb{R} immer der Körper der reellen Zahlen verstanden.

12 Vektoralgebra

121 Vektordefinition und Vektorraum

Physikalische Größen wie Kraft und Geschwindigkeit lassen sich nicht lediglich durch eine Zahl, nämlich ihren Absolutbetrag angeben, auch ihre Richtung muß festgelegt werden. Drei Zahlen benötigt man, wenn beispielsweise die Größe, das Gewicht und das Alter von Individuen zu registrieren sind. Man bedient sich hierzu der Vektoren, die nicht nur für die Ebene \mathbb{R}^2 oder den dreidimensionalen Raum \mathbb{R}^3 , sondern auch für den n -dimensionalen Raum \mathbb{R}^n definiert werden.

Definition: Es sei $x_i \in \mathbb{R}$ mit $i \in \{1, \dots, n\}$ und $n \in \mathbb{N}$, dann bezeichnet man das n -Tupel (x_1, x_2, \dots, x_n) des n -dimensionalen Raumes \mathbb{R}^n als *Vektor* und

schreibt

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}. \quad (121.1)$$

Die x_i sind die *Komponenten* oder die *Koordinaten* von \mathbf{x} .

Der Vektor \mathbf{x} läßt sich als gerichtete Verbindung des Ursprungs eines Koordinatensystems des \mathbb{R}^n mit den Koordinaten $(0, 0, \dots, 0)$ zum Punkt mit den Koordinaten (x_1, x_2, \dots, x_n) interpretieren, wobei die Richtung von \mathbf{x} auf den Punkt weist. Eine geometrische Veranschaulichung erlauben die Vektoren mit zwei Komponenten in der Ebene \mathbb{R}^2 oder die Vektoren mit drei Komponenten im dreidimensionalen Raum \mathbb{R}^3 .

Definition: Zwei Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ werden addiert, indem ihre Komponenten addiert werden

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \dots \\ x_n + y_n \end{pmatrix}. \quad (121.2)$$

Der Vektoraddition entspricht das Parallelogramm der Kräfte, das in Abbildung 121-1 für die Ebene dargestellt ist. Die Vektoraddition ist kommutativ

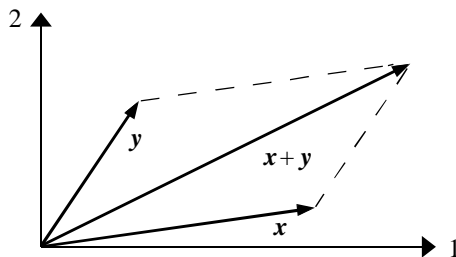


Abb. 121-1: Vektoraddition

und assoziativ, denn für $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ gilt wegen (114.1) und (121.2)

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} \quad \text{und} \quad \mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}. \quad (121.3)$$

Das n -Tupel $(0, 0, \dots, 0)$ des \mathbb{R}^n definiert den Nullvektor, und es gilt mit (114.1) und (121.2)

$$\mathbf{x} + \mathbf{0} = \mathbf{x} \quad \text{für alle} \quad \mathbf{x} \in \mathbb{R}^n. \quad (121.4)$$

Jeder Vektor \mathbf{x} besitzt die additive Inverse $-\mathbf{x}$, so daß gilt

$$\mathbf{x} + (-\mathbf{x}) = \mathbf{0} \quad \text{für alle } \mathbf{x} \in \mathbb{R}^n. \quad (121.5)$$

Definition: Ist $c \in \mathbb{R}$, dann ist die Multiplikation des Vektors $\mathbf{x} \in \mathbb{R}^n$ mit c , das man in diesem Zusammenhang auch als *Skalar* bezeichnet, bestimmt durch

$$c\mathbf{x} = \begin{pmatrix} cx_1 \\ cx_2 \\ \dots \\ cx_n \end{pmatrix}. \quad (121.6)$$

Vektoraddition und Skalarmultiplikation genügen wegen (114.1) dem Distributiv- und Assoziativgesetz. Mit $c, d \in \mathbb{R}$ und $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ gilt

$$\begin{aligned} c(\mathbf{x} + \mathbf{y}) &= c\mathbf{x} + c\mathbf{y} \\ (c + d)\mathbf{x} &= c\mathbf{x} + d\mathbf{x} \\ (cd)\mathbf{x} &= c(d\mathbf{x}) \\ 1\mathbf{x} &= \mathbf{x}. \end{aligned} \quad (121.7)$$

Die Menge der Vektoren, die die angegebenen Gesetze erfüllen, läßt sich wie folgt zusammenfassen.

Definition: Es sei V^n eine Menge von Vektoren des \mathbb{R}^n , dann heißt V^n *linearer Vektorraum* über \mathbb{R} , wenn für alle $c, d \in \mathbb{R}$ und alle Vektoren $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V^n$ die Gesetze (121.3) bis (121.5) und (121.7) erfüllt sind. (121.8)

Der Vektorraum umfaßt also die Menge der Vektoren, die durch Vektoraddition und skalare Multiplikation aus einer Menge gegebener Vektoren konstruiert werden kann, wobei die oben angegebenen Gesetze erfüllt sein müssen. Der Nullvektor $\mathbf{0}$ ist ein Element jeden Vektorraumes. Die Menge $\{\mathbf{0}\}$, die lediglich den Nullvektor enthält, ist ebenfalls ein Vektorraum. Vektoren und Vektorräume lassen sich, wie hier geschehen, nicht nur für die Körper reeller Zahlen, sondern auch für beliebige Zahlkörper definieren (BÖHME 1974, S.193; NEISS und LIERMANN 1975, S.19).

Werden nur Teilmengen von Vektoren in einem Vektorraum betrachtet, führt man den Begriff des Unterraumes ein.

Definition: Bezeichnet man mit U^n eine Teilmenge von Vektoren des Vektorraumes V^n , dann ist U^n *Unterraum* des V^n , falls U^n selbst Vektorraum ist. (121.9)

Beispiel: Die Menge der Vektoren

$$V^3 = \left\{ \mathbf{x} \mid \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, x_i \in \mathbb{R} \right\}$$

bilden einen Vektorraum und ebenso die Menge

$$U^3 = \left\{ \mathbf{u} \mid \mathbf{u} = \begin{pmatrix} u_1 \\ 0 \\ 0 \end{pmatrix}, u_1 \in \mathbb{R} \right\}.$$

Da $U^3 \subset V^3$ ist, bildet U^3 einen Unterraum von V^3 .

Extreme Beispiele von Unterräumen eines jeden Vektorraums sind die Teilmenge $\{\mathbf{0}\}$, die nur den Nullvektor enthält, und die Teilmenge, die sämtliche Vektoren enthält.

Es soll nun die Summe von Unterräumen betrachtet werden.

Definition: Es seien $V_1^n, V_2^n, \dots, V_k^n$ Unterräume des V^n , dann ist die Summe $V_i^n + V_j^n$ der Unterräume V_i^n und V_j^n definiert durch

$$V_i^n + V_j^n = \{ \boldsymbol{\alpha}_i + \boldsymbol{\alpha}_j \mid \boldsymbol{\alpha}_i \in V_i^n, \boldsymbol{\alpha}_j \in V_j^n \},$$

worin $\boldsymbol{\alpha}_i$ einen beliebigen Vektor des Unterraumes V_i^n und $\boldsymbol{\alpha}_j$ einen beliebigen Vektor des Unterraumes V_j^n bedeuten. Weiter bezeichnet man den Vektorraum V_n als *direkte* Summe seiner Unterräume $V_1^n, V_2^n, \dots, V_k^n$ und schreibt

$$V^n = V_1^n \oplus V_2^n \oplus \dots \oplus V_k^n,$$

wenn $\boldsymbol{\alpha} \in V^n$ eindeutig darstellbar ist durch $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + \dots + \boldsymbol{\alpha}_k$ mit $\boldsymbol{\alpha}_i \in V_i^n$. (121.10)

Gilt $V^n = V_1^n \oplus \dots \oplus V_k^n$, so gibt es, abgesehen vom Nullvektor, keine Vektoren, die den Unterräumen V_i^n und V_j^n für $i \neq j$ gemeinsam sind.

122 Lineare Abhängigkeit und Basis eines Vektorraums

Ein für die lineare Algebra grundlegender Begriff ist der der linearen Abhängigkeit und der linearen Unabhängigkeit.

Definition: Eine Teilmenge von Vektoren $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_k \in V^n$ bezeichnet man als *linear abhängig*, wenn Skalare $c_1, c_2, \dots, c_k \in \mathbb{R}$ existieren, die nicht alle gleich Null sind, so daß gilt

$$c_1 \boldsymbol{\alpha}_1 + c_2 \boldsymbol{\alpha}_2 + \dots + c_k \boldsymbol{\alpha}_k = \mathbf{0}.$$

Andernfalls heißen die Vektoren *linear unabhängig*. (122.1)

Die Beziehung $c_1\alpha_1 + \dots + c_k\alpha_k = \mathbf{0}$ kann man sich als ein geschlossenes Polygon vorstellen.

Die folgenden beiden Sätze werden häufiger benötigt werden.

Satz: Eine den Nullvektor enthaltende Teilmenge von Vektoren $\alpha_1, \alpha_2, \dots, \alpha_k, \mathbf{0} \in V^n$ ist stets linear abhängig. (122.2)

Beweis: Es gilt $c_1\alpha_1 + \dots + c_k\alpha_k + c_{k+1}\mathbf{0} = \mathbf{0}$, falls $c_1 = \dots = c_k = 0$ und $c_{k+1} \neq 0$, so daß die Aussage folgt.

Satz: Sind die Vektoren $\alpha_1, \alpha_2, \dots, \alpha_k \in V^n$ linear abhängig, so ist immer wenigstens einer als Linearkombination der übrigen darstellbar. (122.3)

Beweis: Bei linearer Abhängigkeit gilt $c_1\alpha_1 + \dots + c_k\alpha_k = \mathbf{0}$, wobei zumindest ein Skalar ungleich Null ist, beispielsweise $c_i \neq 0$. Somit erhält man

$$\alpha_i = - \sum_{\substack{j=1 \\ j \neq i}}^k \frac{c_j}{c_i} \alpha_j,$$

so daß die Aussage folgt.

Mit Hilfe der Linearkombination $\sum_{i=1}^k c_i\alpha_i$ von Vektoren eines Vektorraums ergibt sich jetzt die

Definition: Wenn jeder Vektor eines Vektorraums V^n durch eine Linearkombination der Vektoren $\alpha_1, \alpha_2, \dots, \alpha_k \in V^n$ erzeugt werden kann, so sagt man, daß die Vektoren $\alpha_1, \alpha_2, \dots, \alpha_k$ den Vektorraum aufspannen. (122.4)

Ist die Teilmenge der Vektoren, die den Vektorraum aufspannen, linear unabhängig, so erhält sie eine besondere Bezeichnung.

Definition: Eine *Basis* für einen Vektorraum V^n ist eine Teilmenge linear unabhängiger Vektoren, die den Vektorraum aufspannen. (122.5)

Satz: Jeder Vektorraum besitzt eine Basis. (122.6)

Dieser Satz soll lediglich erläutert werden, ein Beweis befindet sich in (GRO-TEMeyer 1970, S.192). Besteht der Vektorraum aus der Menge $\{\mathbf{0}\}$, bildet der Nullvektor die Basis. In den übrigen Fällen werden die von Null verschiedenen Vektoren $\alpha_1, \alpha_2, \dots$ des Vektorraums nacheinander ausgewählt, wobei alle Vektoren ausgeschieden werden, die linear abhängig von den bereits ausgewählten Vektoren sind. Die am Ende des Auswahlprozesses gewonnenen linear unabhängigen Vektoren bilden eine Basis des Vektorraums.

Die Bedeutung einer Basis ergibt sich aus dem

Satz: Jeder Vektor V^n wird mit Hilfe der Vektoren einer Basis eindeutig dargestellt. (122.7)

Beweis: Seien $\alpha_1, \alpha_2, \dots, \alpha_k$ die Vektoren einer Basis und stelle $\sum_{i=1}^k c_i \alpha_i$ und $\sum_{i=1}^k d_i \alpha_i$ denselben Vektor dar. Dann ist $\sum_{i=1}^k (c_i - d_i) \alpha_i = \mathbf{0}$, was nur dann möglich ist, falls $c_i - d_i = 0$ für alle i gilt, da die α_i linear unabhängig sind. Hiermit folgt die Aussage.

Die Frage nach der Anzahl der Vektoren in verschiedenen Basen eines Vektorraumes beantwortet der

Satz: Sind $\alpha_1, \dots, \alpha_k$ und β_1, \dots, β_l zwei Basen des Vektorraums V^n , dann ist $k = l$. (122.8)

Beweis: Zunächst sei $k > l$ angenommen. Die Vektoren $\alpha_1, \dots, \alpha_k$ spannen V^n auf und sind linear unabhängig, da sie eine Basis bilden. Ebenfalls spannen die Vektoren $\alpha_1, \dots, \alpha_k, \beta_1$ den Vektorraum V^n auf, sie sind jedoch linear abhängig, da $\beta_1 \in V^n$ sich nach (122.7) darstellen läßt durch $\alpha_1, \dots, \alpha_k$, folglich $\beta_1 = c_1 \alpha_1 + c_2 \alpha_2 + \dots + c_k \alpha_k$, wobei zumindest ein Skalar $c_i \neq 0$, da β_1 ein Basisvektor und somit $\beta_1 \neq \mathbf{0}$ ist. Es gelte $c_1 \neq 0$, so daß α_1 nach (122.3) durch eine Linearkombination der Vektoren $\alpha_2, \dots, \alpha_k, \beta_1$ darzustellen ist, die somit den Vektorraum aufspannen. Ebenso spannen die Vektoren $\alpha_2, \dots, \alpha_k, \beta_1, \beta_2$ den Vektorraum auf, die aber linear abhängig sind, da β_2 darstellbar ist durch $\beta_2 = d_2 \alpha_2 + \dots + d_k \alpha_k + d_1 \beta_1$. Für einen Wert d_i mit $i \in \{2, \dots, k\}$ gilt $d_i \neq 0$, da sonst β_1 und β_2 linear abhängig wären. Es gelte $d_2 \neq 0$, so daß α_2 durch $\alpha_3, \dots, \alpha_k, \beta_1, \beta_2$ darzustellen ist, so daß diese Vektoren den Vektorraum aufspannen. Die gleichen Überlegungen lassen sich bis α_k anstellen, so daß β_1, \dots, β_k den Vektorraum aufspannen und die Vektoren $\beta_{k+1}, \dots, \beta_l$ linear abhängig sind. Folglich ist $k = l$, und man erhält die Aussage.

Die minimale Anzahl der Vektoren, die einen Vektorraum aufspannen, wird also durch die Anzahl der Vektoren einer Basis des Vektorraums bestimmt.

Definition: Die *Dimension* $\dim V^n$ eines Vektorraums V^n ist die Anzahl der Vektoren einer beliebigen Basis des V^n . (122.9)

Zur Erläuterung sei der Vektorraum V^n betrachtet, der durch die Menge der Vektoren x mit den n Komponenten $x_i \in \mathbb{R}$ gegeben ist. Die Vektoren

$$e_1 = \begin{vmatrix} 1 \\ 0 \\ \dots \\ 0 \end{vmatrix}, e_2 = \begin{vmatrix} 0 \\ 1 \\ \dots \\ 0 \end{vmatrix}, \dots, e_n = \begin{vmatrix} 0 \\ 0 \\ \dots \\ 1 \end{vmatrix} \quad (122.10)$$

mit n Komponenten sind linear unabhängig und spannen V^n auf, denn jeder beliebige Vektor $\mathbf{x} \in V^n$ mit den n Komponenten x_i läßt sich darstellen durch $\mathbf{x} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \dots + x_n \mathbf{e}_n$. Die Vektoren $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ bilden also eine Basis des V^n , und es gilt $\dim V^n = n$. Geometrisch lassen sich diese Basisvektoren als Vektoren in Richtung von Koordinatenachsen deuten, die die Koordinaten der Punkte des \mathbb{R}^n definieren.

123 Skalarprodukt und Euklidischer Raum

Bis jetzt wurden Vektorräume nur unter dem Gesichtspunkt der linearen Unabhängigkeit betrachtet. Um aber auch mit Längen von Vektoren und Winkeln zwischen Vektoren arbeiten zu können, die bei geometrischen Problemen auftreten, muß das Skalarprodukt, auch inneres Produkt genannt, zweier Vektoren eingeführt werden.

Definition: Es sei $\mathbf{x}, \mathbf{y} \in V^n$, wobei \mathbf{x} die Komponenten x_i und \mathbf{y} die Komponenten y_i besitze, dann ist das *Skalarprodukt* $\mathbf{x}'\mathbf{y}$ von \mathbf{x} und \mathbf{y} gegeben durch

$$\mathbf{x}'\mathbf{y} = \sum_{i=1}^n x_i y_i . \quad (123.1)$$

Der Grund für die Schreibweise $\mathbf{x}'\mathbf{y}$, häufig findet man auch $\mathbf{x}\cdot\mathbf{y}$ oder $\langle \mathbf{x}, \mathbf{y} \rangle$, ergibt sich aus der Definition (131.6) eines Matrizenproduktes.

Satz: Für das Skalarprodukt gilt

$$\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x} , (\mathbf{x} + \mathbf{y})'\mathbf{z} = \mathbf{x}'\mathbf{z} + \mathbf{y}'\mathbf{z} , (c\mathbf{x}')\mathbf{y} = c(\mathbf{x}'\mathbf{y}) . \quad (123.2)$$

Beweis: Mit (114.1) und (123.1) folgen die Aussagen.

Die *Länge* $|\mathbf{x}|$, der *Absolutbetrag* oder die *Norm* eines Vektors \mathbf{x} ist definiert durch

$$|\mathbf{x}| = (\mathbf{x}'\mathbf{x})^{1/2} . \quad (123.3)$$

Der Winkel α zwischen zwei Vektoren \mathbf{x} und \mathbf{y} ergibt sich aus der geometrischen Definition des Skalarproduktes (z.B. STRANG 1980, S.106)

$$\mathbf{x}'\mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos \alpha \quad (123.4)$$

und somit

$$\cos \alpha = \frac{\mathbf{x}'\mathbf{y}}{((\mathbf{x}'\mathbf{x})(\mathbf{y}'\mathbf{y}))^{1/2}} . \quad (123.5)$$

Vektorräume mit Skalarprodukt erhalten eine eigene Bezeichnung.

Definition: Ein Vektorraum V^n mit $\dim V^n = n$, für den das Skalarprodukt definiert ist, bezeichnet man als n -dimensionalen *Euklidischen Raum* E^n . (123.6)

Euklidische Räume besitzen endliche Dimensionen. Bei der Erweiterung auf unendliche Dimensionen ergeben sich die Hilbertschen Räume (MESCHKOWSKI 1962), die jedoch im folgenden nicht benötigt werden.

124 Orthogonale Unterräume

Stehen zwei Vektoren aufeinander senkrecht, folgt mit $\cos \alpha = 0$ in (123.4) $\mathbf{x}'\mathbf{y} = 0$.

Definition: Zwei Vektoren $\mathbf{x}, \mathbf{y} \in V^n$ heißen genau dann zueinander *orthogonal*, wenn $\mathbf{x}'\mathbf{y} = 0$ gilt. (124.1)

Der Nullvektor $\mathbf{0}$ ist also orthogonal zu jedem anderen Vektor.

Satz: Sind die Vektoren $\alpha_1, \alpha_2, \dots, \alpha_k$ paarweise zueinander orthogonal und ungleich Null, sind sie linear unabhängig. (124.2)

Beweis: Es wird gezeigt, daß nur dann $\mathbf{0} = c_1\alpha_1 + c_2\alpha_2 + \dots + c_k\alpha_k$ sich ergibt, falls $c_i = 0$ für $i \in \{1, \dots, k\}$, was wegen (122.1) lineare Unabhängigkeit bedeutet. Skalare Multiplikation der Gleichung mit α_j' ergibt $0 = \sum_{i=1}^k c_i \alpha_j' \alpha_i = c_j \alpha_j' \alpha_j$ wegen (124.1). Da $\alpha_j \neq \mathbf{0}$, folgt $c_j = 0$ und damit die Aussage.

Es sollen jetzt zueinander orthogonale Basisvektoren behandelt werden.

Definition: Eine Basis des E^n bezeichnet man als *Orthogonalbasis*, falls die Basisvektoren paarweise zueinander orthogonal sind, und als *Orthonormalbasis*, falls außerdem die Basisvektoren die Länge Eins besitzen. (124.3)

Eine orthonormale Basis des E^n bilden die n Vektoren mit den n Komponenten

$$\mathbf{e}_1 = \begin{vmatrix} 1 \\ 0 \\ \dots \\ 0 \end{vmatrix}, \mathbf{e}_2 = \begin{vmatrix} 0 \\ 1 \\ \dots \\ 0 \end{vmatrix}, \dots, \mathbf{e}_n = \begin{vmatrix} 0 \\ 0 \\ \dots \\ 1 \end{vmatrix}.$$

Wie bereits im Zusammenhang mit (122.10) erwähnt, zeigen diese Basisvektoren in Richtung der Achsen eines Koordinatensystems für Punkte des E^n , das wegen (124.1) zudem ein *orthogonales Koordinatensystem* ist.

Satz: Zu jeder Basis $\mathbf{b}_1, \dots, \mathbf{b}_n$ des E^n existiert eine Orthonormalbasis $\mathbf{o}_1, \dots, \mathbf{o}_n$ derart, daß jeder Vektor \mathbf{o}_i eine Linearkombination von $\mathbf{b}_1, \dots, \mathbf{b}_n$ darstellt. (124.4)

Der Beweis dieses Satzes wird mit Hilfe des Schmidtschen Orthogonalisierungsverfahrens geführt (NEISS und LIERMANN 1975, S.134).

Satz: Falls $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_r$ eine orthonormale Basis eines Unterraums des E^n ist, dann läßt sie sich zur orthonormalen Basis des E^n mit den n Basisvektoren $\mathbf{o}_1, \dots, \mathbf{o}_r, \mathbf{o}_{r+1}, \dots, \mathbf{o}_n$ ergänzen. (124.5)

Der Beweis dieses Satzes ergibt sich mit (124.4) und dem Basisergänzungssatz (GROTEMEYER 1970, S.193).

Der Begriff der Orthogonalität wird auch auf Unterräume von Vektorräumen ausgedehnt.

Definition: U und W seien Unterräume des E^n . Gilt $\mathbf{x}'\mathbf{y} = 0$ für alle $\mathbf{x} \in U$ und alle $\mathbf{y} \in W$, dann bezeichnet man U und W als zueinander *orthogonale Unterräume* des E^n . (124.6)

Für die Menge der Vektoren eines Vektorraums, die orthogonal zu den Vektoren eines Unterraums stehen, gilt der

Satz: Es sei U ein Unterraum des Vektorraums E^n und U^\perp die Menge der Vektoren des E^n , die orthogonal zu jedem Vektor von U sind. U^\perp ist dann Unterraum des E^n und wird als *orthogonales Komplement* von U in E^n bezeichnet. Weiter ist jeder Vektor $\mathbf{z} \in E^n$ eindeutig darstellbar durch $\mathbf{z} = \mathbf{x} + \mathbf{y}$ mit $\mathbf{x} \in U$ und $\mathbf{y} \in U^\perp$, so daß $E^n = U \oplus U^\perp$ gilt. Ferner ist $\dim U + \dim U^\perp = \dim E^n = n$. (124.7)

Beweis: Die Vektoren $\mathbf{u}, \mathbf{v} \in E^n$ seien orthogonal zu jedem Vektor in U , so daß $\mathbf{u}, \mathbf{v} \in U^\perp$ gilt. Dann sind auch Linearkombinationen von \mathbf{u} und \mathbf{v} orthogonal zu jedem Vektor in U , so daß U^\perp nach (121.9) einen Unterraum des E^n bildet. Weiter sei $\dim U = r$ und $\alpha_1, \alpha_2, \dots, \alpha_r$ eine orthonormale Basis für U , so daß $\mathbf{x} \in U$ nach (122.7) durch $\mathbf{x} = \sum_{i=1}^r c_i \alpha_i$ mit $c_i \in \mathbb{R}$ darstellbar ist. Ferner sei $\alpha_1, \dots, \alpha_r, \alpha_{r+1}, \dots, \alpha_n$ die aufgrund von (124.5) ergänzte orthonormale Basis des E^n , so daß für $\mathbf{z} \in E^n$ die Darstellung $\mathbf{z} = \sum_{i=1}^n d_i \alpha_i$ mit $d_i \in \mathbb{R}$ erhalten wird. Die Vektoren \mathbf{z} sind nur dann orthogonal zu $\mathbf{x} \in U$, falls $d_1 = d_2 = \dots = d_r = 0$, falls sie also in dem $(n-r)$ -dimensionalen Unterraum U^\perp der Vektoren mit der Darstellung $\mathbf{y} = \sum_{i=r+1}^n d_i \alpha_i$ sich befinden. Hieraus folgen dann die Aussagen und mit (121.10) die Darstellung des E^n als direkte Summe.

Beispiel: Der Vektorraum E^3 sei definiert durch

$$E^3 = \left\{ \mathbf{x} \mid \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, x_1, x_2, x_3 \in \mathbb{R} \right\}.$$

Dann ist U

$$U = \left\{ \mathbf{x} \mid \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix}, x_1, x_2 \in \mathbb{R} \right\}$$

ein Unterraum von E^3 und U^\perp

$$U^\perp = \left\{ \mathbf{x} \mid \mathbf{x} = \begin{pmatrix} 0 \\ 0 \\ x_3 \end{pmatrix}, x_3 \in \mathbb{R} \right\}$$

das orthogonale Komplement von U in E^3 .

13 Matrizen

131 Matrixdefinition und Matrixverknüpfungen

a) Definition einer Matrix

Eine rechteckige Tabelle von Zahlen bezeichnet man als Matrix. Genauer formuliert erhält man die

Definition: Es sei K ein Körper von Zahlen und $m, n \in \mathbb{N}$. Dann heißt die rechteckige Tabelle der Elemente $a_{ij} \in K$

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

eine $m \times n$ Matrix \mathbf{A} , wobei m die Anzahl der *Zeilen* und n die der *Spalten*, also die Dimensionen der Matrix angeben. (131.1)

Im folgenden werden ausschließlich Matrizen mit Elementen aus dem Körper der reellen Zahlen, also $a_{ij} \in \mathbb{R}$ behandelt.

Mit $n = 1$ geht die Matrix \mathbf{A} in den mit (121.1) definierten $m \times 1$ Vektor über, der auch als *Spaltenvektor* bezeichnet wird. Mit $m = 1$ ergibt sich aus \mathbf{A} der $1 \times n$ *Zeilenvektor*. Ist $m = n$, heißt \mathbf{A} eine *quadratische* Matrix. Ist $a_{ij} = 0$ für alle

i, j , wird \mathbf{A} zur *Nullmatrix* $\mathbf{0}$. Die quadratische $n \times n$ Matrix \mathbf{A} mit $a_{ij} = 1$ für $i = j$ und $a_{ij} = 0$ für $i \neq j$ bezeichnet man als $n \times n$ *Einheitsmatrix* \mathbf{I} oder \mathbf{I}_n . Die Elemente a_{ii} einer quadratischen Matrix heißen *Diagonalelemente*. Gilt in einer quadratischen $n \times n$ Matrix \mathbf{A} für die Nicht-Diagonalelemente $a_{ij} = 0$ für $i \neq j$, heißt \mathbf{A} eine *Diagonalmatrix*, und man schreibt $\mathbf{A} = \text{diag}(a_{11}, \dots, a_{nn})$. Sind sämtliche Elemente einer quadratischen Matrix unterhalb der Diagonalelemente gleich Null, liegt eine *obere Dreiecksmatrix* vor. Befinden sich die Nullelemente oberhalb der Diagonalelemente, spricht man von *unterer Dreiecksmatrix*. Sind die Diagonalelemente einer Dreiecksmatrix gleich Eins, bezeichnet man sie als *Einheits-Dreiecksmatrix*.

b) Addition von Matrizen

Definition: Zwei $m \times n$ Matrizen $\mathbf{A} = (a_{ij})$ und $\mathbf{B} = (b_{ij})$ werden addiert, indem positionsgleiche Elemente addiert werden, also $\mathbf{A} + \mathbf{B} = (a_{ij} + b_{ij})$. (131.2)

Die Addition zweier Matrizen ist also nur für Matrizen gleichen Formats oder gleicher Dimensionen definiert, die gleiche Anzahl von Zeilen und Spalten besitzen.

Satz: Die Matrizenaddition ist kommutativ und assoziativ

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} \\ \mathbf{A} + (\mathbf{B} + \mathbf{C}) &= (\mathbf{A} + \mathbf{B}) + \mathbf{C} . \end{aligned} \quad (131.3)$$

Beweis: Mit (114.1) und (131.2) folgen die Aussagen.

c) Skalarmultiplikation

Definition: Eine Matrix $\mathbf{A} = (a_{ij})$ wird mit einem Skalar $c \in \mathbb{R}$ multipliziert, indem jedes Element von \mathbf{A} mit c multipliziert wird, also $c\mathbf{A} = (ca_{ij})$. (131.4)

Satz: Mit $c, d \in \mathbb{R}$ gilt

$$\begin{aligned} (c + d)\mathbf{A} &= c\mathbf{A} + d\mathbf{A} \\ c(\mathbf{A} + \mathbf{B}) &= c\mathbf{A} + c\mathbf{B} . \end{aligned} \quad (131.5)$$

Beweis: Mit (114.1) und (131.4) ergeben sich die Aussagen.

d) Matrizenmultiplikation

Definition: Das Produkt $\mathbf{AB} = (c_{ij})$ der $m \times n$ Matrix \mathbf{A} und der $n \times p$ Matrix \mathbf{B} ist definiert durch

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj} \quad \text{für } i \in \{1, \dots, m\} \quad \text{und } j \in \{1, \dots, p\} . \quad (131.6)$$

Das Produkt zweier Matrizen ist also nur dann definiert, wenn die Anzahl der Spalten von \mathbf{A} der Anzahl der Zeilen von \mathbf{B} gleich ist. Das Produkt \mathbf{AB} besitzt soviele Zeilen wie \mathbf{A} und soviele Spalten wie \mathbf{B} .

Beispiel: Die beiden unten definierten Matrizen \mathbf{A} und \mathbf{B} sollen miteinander multipliziert werden, wobei zur Rechenkontrolle als zusätzliche Spalte an die Matrix \mathbf{B} die Summe ihrer Zeilen angehängt und in die Multiplikation miteinbezogen werde, so daß eine zusätzliche Spalte in der Produktmatrix \mathbf{AB} erhalten wird, die gleich der Zeilensumme von \mathbf{AB} sein muß. Die Zeilensumme von \mathbf{B} ergibt sich nämlich aus \mathbf{Be} mit $\mathbf{e} = |1, \dots, 1|'$ und die Zeilensumme des Produktes aus $(\mathbf{AB})\mathbf{e} = \mathbf{A}(\mathbf{Be})$. Zur besseren Übersicht wird außerdem das sogenannte *Falksche Schema* benutzt.

$$\mathbf{B} = \left| \begin{array}{cccc|c} 1 & -1 & 0 & 3 & 3 \\ 0 & 1 & 1 & -4 & -2 \\ 2 & 2 & -2 & 5 & 7 \end{array} \right|$$

$$\mathbf{A} = \left| \begin{array}{ccc|cccc} -1 & 2 & 1 & 1 & 5 & 0 & -6 & 0 \\ 0 & 3 & 4 & 8 & 11 & -5 & 8 & 22 \end{array} \right|$$

$$= \mathbf{AB}$$

Satz: Die Matrizenmultiplikation ist assoziativ, distributiv, aber im allgemeinen nicht kommutativ

$$\begin{aligned} \mathbf{A}(\mathbf{BC}) &= (\mathbf{AB})\mathbf{C} \\ \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \quad \text{und} \quad (\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \\ \text{im allgemeinen} \quad \mathbf{AB} &\neq \mathbf{BA}. \end{aligned} \tag{131.7}$$

Beweis: $\mathbf{A} = (a_{ij})$ sei eine $m \times n$ Matrix, $\mathbf{B} = (b_{ij})$ eine $n \times p$ Matrix und $\mathbf{C} = (c_{ij})$ eine $p \times r$ Matrix. Ferner sei $\mathbf{D} = (d_{ij}) = \mathbf{AB}$ und $\mathbf{E} = (e_{ij}) = \mathbf{BC}$. Dann ist

$$\begin{aligned} \mathbf{A}(\mathbf{BC}) &= \left(\sum_{k=1}^n a_{ik} e_{kj} \right) = \left(\sum_{k=1}^n a_{ik} \left(\sum_{l=1}^p b_{kl} c_{lj} \right) \right) \\ &= \left(\sum_{l=1}^p \left(\sum_{k=1}^n a_{ik} b_{kl} \right) c_{lj} \right) = \left(\sum_{l=1}^p d_{il} c_{lj} \right) = (\mathbf{AB})\mathbf{C}. \end{aligned}$$

Analog läßt sich die Distributivität beweisen, während die Nicht-Kommutativität offensichtlich ist, so daß die Aussagen sich ergeben.

Die Multiplikation mit der Einheitsmatrix \mathbf{I} entsprechender Dimensionen verändert eine Matrix \mathbf{A} nicht

$$\mathbf{IA} = \mathbf{A} \quad \text{und} \quad \mathbf{AI} = \mathbf{A}. \tag{131.8}$$

e) Transponierung einer Matrix

Definition: Vertauscht man in einer $m \times n$ Matrix \mathbf{A} die Zeilen und Spalten, so entsteht die *transponierte* $n \times m$ Matrix \mathbf{A}'

$$\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{vmatrix}, \mathbf{A}' = \begin{vmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{vmatrix}. \quad (131.9)$$

Durch Transponierung geht der $m \times 1$ Spaltenvektor \mathbf{x} in den $1 \times m$ Zeilenvektor \mathbf{x}' über, so daß die Definition (123.1) des Skalarproduktes zweier $m \times 1$ Vektoren \mathbf{x} und \mathbf{y} übereinstimmt mit (131.6). Bleibt eine quadratische Matrix bei der Transponierung unverändert, also $\mathbf{A}' = \mathbf{A}$, bezeichnet man sie als *symmetrische* Matrix; ändert sie durch die Transponierung nur ihr Vorzeichen, also $\mathbf{A}' = -\mathbf{A}$, heißt sie *schiefsymmetrisch*.

Satz: Die Transponierung der Summe und des Produktes zweier Matrizen \mathbf{A} und \mathbf{B} ergibt

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}' \quad \text{und} \quad (\mathbf{AB})' = \mathbf{B}'\mathbf{A}'. \quad (131.10)$$

Beweis: Die erste Aussage folgt unmittelbar mit (131.9). Zum Beweis der zweiten seien die $m \times n$ und $n \times p$ Matrizen \mathbf{A} und \mathbf{B} gegeben, die dargestellt seien durch $\mathbf{A} = |\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m|'$ mit $\mathbf{a}'_i = |a_{i1}, \dots, a_{in}|$ und $\mathbf{B} = |\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_p|$ mit $\mathbf{b}_i = |b_{1i}, \dots, b_{ni}|'$. Dann ist $\mathbf{AB} = (\mathbf{a}'_i \mathbf{b}_j)$ und mit (131.9) $(\mathbf{AB})' = (\mathbf{a}'_j \mathbf{b}_i)$. Weiter gilt $\mathbf{B}'\mathbf{A}' = (\mathbf{b}'_i \mathbf{a}_j) = (\mathbf{a}'_j \mathbf{b}_i) = (\mathbf{AB})'$ wegen (123.2), so daß die Aussagen folgen.

Satz: $\mathbf{A}'\mathbf{A} = \mathbf{0}$ gilt genau dann, wenn $\mathbf{A} = \mathbf{0}$ ist. (131.11)

Beweis: Mit $\mathbf{A} = \mathbf{0}$ folgt $\mathbf{A}'\mathbf{A} = \mathbf{0}$. Andererseits folgt aus $\mathbf{A}'\mathbf{A} = \mathbf{0}$, daß die Summe der Quadrate der Elemente jeder Spalte von \mathbf{A} gleich Null und daher jedes Element gleich Null ist, so daß die Aussage folgt.

f) Inverse Matrix

Definition: Existiert für eine $n \times n$ Matrix \mathbf{A} eine $n \times n$ Matrix \mathbf{B} derart, daß $\mathbf{AB} = \mathbf{I}$ und $\mathbf{BA} = \mathbf{I}$ gilt, so ist \mathbf{B} die *inverse* Matrix von \mathbf{A} . Die Matrix \mathbf{A} heißt dann *regulär*, im anderen Fall *singulär*. (131.12)

Eine notwendige und hinreichende Bedingung dafür, daß eine Matrix regulär ist, wird im Satz (133.1) angegeben.

Satz: Die Inverse einer regulären Matrix ist eindeutig bestimmt und wird mit \mathbf{A}^{-1} bezeichnet. (131.13)

Beweis: Es seien \mathbf{A}_1^{-1} und \mathbf{A}_2^{-1} zwei Inversen der regulären Matrix \mathbf{A} . Aus $\mathbf{A}_2^{-1}\mathbf{A} = \mathbf{I}$ folgt $\mathbf{A}_2^{-1}\mathbf{A}\mathbf{A}_1^{-1} = \mathbf{A}_1^{-1}$ und daraus $\mathbf{A}_2^{-1} = \mathbf{A}_1^{-1}$, denn es ist $\mathbf{A}\mathbf{A}_1^{-1} = \mathbf{I}$ wegen (131.12), so daß die Aussage folgt.

Satz: Sind \mathbf{A} und \mathbf{B} reguläre Matrizen, gilt

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \quad (131.14)$$

$$(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1} \quad (131.15)$$

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}. \quad (131.16)$$

Beweis: Es sei $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, dann gilt $(\mathbf{AB})\mathbf{C} = \mathbf{I}$ und $\mathbf{C}(\mathbf{AB}) = \mathbf{I}$, so daß mit (131.12) und (131.13) die Aussage (131.14) folgt. Durch Transponierung von $\mathbf{AA}^{-1} = \mathbf{I}$ und $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ ergibt sich mit (131.10) $(\mathbf{A}^{-1})'\mathbf{A}' = \mathbf{I}$ und $\mathbf{A}'(\mathbf{A}^{-1})' = \mathbf{I}$, so daß \mathbf{A}' regulär ist. Dann gilt weiter $(\mathbf{A}')^{-1}\mathbf{A}' = \mathbf{I}$ und $\mathbf{A}'(\mathbf{A}')^{-1} = \mathbf{I}$, so daß wegen der Eindeutigkeit der Inversen die Aussage (131.15) folgt. Ebenso ergibt sich aus $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ und $\mathbf{AA}^{-1} = \mathbf{I}$, daß \mathbf{A}^{-1} regulär ist und damit $\mathbf{A}^{-1}(\mathbf{A}^{-1})^{-1} = \mathbf{I}$ sowie $(\mathbf{A}^{-1})^{-1}\mathbf{A}^{-1} = \mathbf{I}$, so daß (131.16) folgt.

Satz: Die Inverse einer symmetrischen Matrix ist wieder symmetrisch. (131.17)

Beweis: Aus (131.15) folgt mit $\mathbf{A} = \mathbf{A}'$ unmittelbar $(\mathbf{A}^{-1})' = \mathbf{A}^{-1}$.

g) Blockmatrizen

Faßt man in einer $m \times n$ Matrix \mathbf{A} die ersten r Zeilen und die ersten s Spalten zu der $r \times s$ Untermatrix \mathbf{A}_{11} zusammen, die ersten r Zeilen und die verbleibenden $n-s$ Spalten zu der $r \times (n-s)$ Untermatrix \mathbf{A}_{12} und die verbleibenden Zeilen und Spalten entsprechend, so ergibt sich \mathbf{A} als *Blockmatrix* zu

$$\mathbf{A} = \begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix}. \quad (131.18)$$

Selbstverständlich läßt sich \mathbf{A} noch in weitere Untermatrizen zerlegen, doch soll im folgenden die Unterteilung auf vier Untermatrizen beschränkt bleiben. Für die transponierte Blockmatrix \mathbf{A}' folgt mit (131.9)

$$\mathbf{A}' = \begin{vmatrix} \mathbf{A}'_{11} & \mathbf{A}'_{21} \\ \mathbf{A}'_{12} & \mathbf{A}'_{22} \end{vmatrix}. \quad (131.19)$$

Besitzt eine $m \times n$ Matrix \mathbf{B} eine entsprechende Unterteilung wie die Blockmatrix \mathbf{A} , so ergibt sich die Summe $\mathbf{A} + \mathbf{B}$ mit (131.2) zu

$$\mathbf{A} + \mathbf{B} = \begin{vmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{vmatrix}. \quad (131.20)$$

Ist eine $n \times u$ Matrix \mathbf{C} aufgeteilt in s und $n - s$ Zeilen und in eine beliebige Zerlegung der Spalten, lassen sich die Blockmatrix \mathbf{A} aus (131.18) und \mathbf{C} miteinander multiplizieren, und man erhält mit (131.6)

$$\begin{aligned} \mathbf{AC} &= \begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} \begin{vmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{vmatrix} \\ &= \begin{vmatrix} \mathbf{A}_{11}\mathbf{C}_{11} + \mathbf{A}_{12}\mathbf{C}_{21} & \mathbf{A}_{11}\mathbf{C}_{12} + \mathbf{A}_{12}\mathbf{C}_{22} \\ \mathbf{A}_{21}\mathbf{C}_{11} + \mathbf{A}_{22}\mathbf{C}_{21} & \mathbf{A}_{21}\mathbf{C}_{12} + \mathbf{A}_{22}\mathbf{C}_{22} \end{vmatrix}. \end{aligned} \quad (131.21)$$

h) Kronecker-Produkt

Definition: Es sei $\mathbf{A} = (a_{ij})$ eine $m \times n$ Matrix und $\mathbf{B} = (b_{ij})$ eine $p \times q$ Matrix, dann ist das *Kronecker-Produkt* $\mathbf{A} \otimes \mathbf{B}$ von \mathbf{A} und \mathbf{B} definiert als die $mp \times nq$ Matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{vmatrix} a_{11}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ \dots & \dots & \dots \\ a_{m1}\mathbf{B} & \dots & a_{mn}\mathbf{B} \end{vmatrix}. \quad (131.22)$$

Das Kronecker-Produkt wird für die multivariaten Modelle der Parameterschätzung benötigt. Folgende Rechenregeln sind zu beachten.

Satz: Es gilt $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$. (131.23)

Beweis: Transponiert man die Matrix $\mathbf{A} \otimes \mathbf{B}$, ergibt sich

$$(\mathbf{A} \otimes \mathbf{B})' = \begin{vmatrix} a_{11}\mathbf{B}' & \dots & a_{m1}\mathbf{B}' \\ \dots & \dots & \dots \\ a_{1n}\mathbf{B}' & \dots & a_{mn}\mathbf{B}' \end{vmatrix}$$

und mit (131.22) die Aussage.

Satz: Sind \mathbf{A} und \mathbf{B} sowie \mathbf{E} und \mathbf{F} jeweils $m \times n$ Matrizen, dann gilt

$$\begin{aligned} (\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} &= (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C}) \\ \mathbf{D} \otimes (\mathbf{E} + \mathbf{F}) &= (\mathbf{D} \otimes \mathbf{E}) + (\mathbf{D} \otimes \mathbf{F}). \end{aligned} \quad (131.24)$$

Beweis: Mit (131.2) und (131.22) erhält man

$$\begin{aligned} (\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} &= \begin{vmatrix} (a_{11} + b_{11})\mathbf{C} & \dots & (a_{1n} + b_{1n})\mathbf{C} \\ \dots & \dots & \dots \\ (a_{m1} + b_{m1})\mathbf{C} & \dots & (a_{mn} + b_{mn})\mathbf{C} \end{vmatrix} \\ &= (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C}) \end{aligned}$$

und die zweite Aussage entsprechend.

Satz: Mit $c \in \mathbb{R}$ gilt

$$c(\mathbf{A} \otimes \mathbf{B}) = (c\mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (c\mathbf{B}). \quad (131.25)$$

Beweis: Mit (131.4) und (131.22) folgen die Aussagen.

Satz: Ist \mathbf{A} eine $l \times m$, \mathbf{C} eine $m \times n$ Matrix, \mathbf{B} eine $p \times q$ und \mathbf{D} eine $q \times r$ Matrix, dann gilt

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}. \quad (131.26)$$

Beweis: Definiert man $\mathbf{AC} = (f_{ij})$, ergibt sich mit (131.6) und (131.22)

$$\begin{aligned} (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) &= \left| \begin{array}{ccc} a_{11}\mathbf{B} & \dots & a_{1m}\mathbf{B} \\ \dots & \dots & \dots \\ a_{l1}\mathbf{B} & \dots & a_{lm}\mathbf{B} \end{array} \right| \left| \begin{array}{ccc} c_{11}\mathbf{D} & \dots & c_{1n}\mathbf{D} \\ \dots & \dots & \dots \\ c_{m1}\mathbf{D} & \dots & c_{mn}\mathbf{D} \end{array} \right| \\ &= \left| \begin{array}{ccc} \sum_{i=1}^m a_{1i}c_{i1}\mathbf{BD} & \dots & \sum_{i=1}^m a_{1i}c_{in}\mathbf{BD} \\ \dots & \dots & \dots \\ \sum_{i=1}^m a_{li}c_{i1}\mathbf{BD} & \dots & \sum_{i=1}^m a_{li}c_{in}\mathbf{BD} \end{array} \right| \\ &= \left| \begin{array}{ccc} f_{11}\mathbf{BD} & \dots & f_{1n}\mathbf{BD} \\ \dots & \dots & \dots \\ f_{l1}\mathbf{BD} & \dots & f_{ln}\mathbf{BD} \end{array} \right| = \mathbf{AC} \otimes \mathbf{BD}. \end{aligned}$$

Satz: Sind \mathbf{A} und \mathbf{B} zwei reguläre $m \times m$ beziehungsweise $n \times n$ Matrizen, dann gilt

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}. \quad (131.27)$$

Beweis: Es sei $\mathbf{C} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$, dann erhält man mit (131.26) $(\mathbf{A} \otimes \mathbf{B})\mathbf{C} = \mathbf{I}$ und $\mathbf{C}(\mathbf{A} \otimes \mathbf{B}) = \mathbf{I}$, so daß mit (131.12) und (131.13) die Aussage folgt.

132 Rang einer Matrix

Wesentliche Eigenschaften einer Matrix sind mit der Anzahl ihrer linear unabhängigen Zeilen und Spalten verbunden.

Definition: Die maximale Anzahl der linear unabhängigen Zeilen einer Matrix heißt der *Rang* einer Matrix, und man schreibt $r = \text{rg}\mathbf{A}$, wenn r den Rang und \mathbf{A} die Matrix bezeichnet. (132.1)

Wie aus dem folgenden Satz sich ergibt, kann in dieser Definition das Wort Zeilen auch durch das Wort Spalten ersetzt werden.

Satz: Die maximale Anzahl r der linear unabhängigen Zeilen einer $m \times n$ Matrix \mathbf{A} gleicht der maximalen Anzahl der linear unabhängigen Spalten, und es gilt $r = \operatorname{rg} \mathbf{A} = \operatorname{rg} \mathbf{A}' \leq \min(m, n)$. (132.2)

Beweis: Die $m \times n$ Matrix \mathbf{A} sei mit $\mathbf{A} = |\mathbf{a}_1, \dots, \mathbf{a}_n|$ durch ihre n Spaltenvektoren \mathbf{a}_k dargestellt. Unter diesen n Spaltenvektoren sollen sich $r_s \leq n$ linear unabhängige Vektoren befinden, die zu der $m \times r_s$ Matrix \mathbf{B} mit $\mathbf{B} = |\mathbf{b}_1, \dots, \mathbf{b}_{r_s}|$ zusammengefaßt seien. Alle Spaltenvektoren \mathbf{a}_k von \mathbf{A} lassen sich aus den Spaltenvektoren von \mathbf{B} linear kombinieren

$$\mathbf{a}_k = c_{1k}\mathbf{b}_1 + c_{2k}\mathbf{b}_2 + \dots + c_{r_s k}\mathbf{b}_{r_s} \quad \text{mit } k \in \{1, \dots, n\}$$

oder

$$\mathbf{A} = \mathbf{BC} \quad \text{mit } \mathbf{C} = (c_{ik}), \quad i \in \{1, \dots, r_s\}, \quad k \in \{1, \dots, n\}.$$

Werden die Matrizen \mathbf{A} und \mathbf{C} durch ihre Zeilenvektoren dargestellt

$$\mathbf{A} = \begin{vmatrix} \alpha'_1 \\ \dots \\ \alpha'_m \end{vmatrix} = (a_{ij}), \quad \mathbf{C} = \begin{vmatrix} \gamma'_1 \\ \dots \\ \gamma'_{r_s} \end{vmatrix} = (c_{ik}),$$

ergibt sich mit $\mathbf{B} = (b_{ij})$

$$\alpha'_l = b_{l1}\gamma'_1 + b_{l2}\gamma'_2 + \dots + b_{lr_s}\gamma'_{r_s}, \quad l \in \{1, \dots, m\}.$$

Hieraus folgt nämlich das Element a_{lp} der Matrix \mathbf{A} mit

$$a_{lp} = b_{l1}c_{1p} + b_{l2}c_{2p} + \dots + b_{lr_s}c_{r_s p}$$

in Übereinstimmung mit dem Element a_{lp} aus dem Matrizenprodukt \mathbf{BC} . Alle Zeilenvektoren α'_l der Matrix \mathbf{A} können also als Linearkombinationen der r_s Zeilenvektoren γ'_i der Matrix \mathbf{C} erhalten werden. Die Anzahl r_z der linear unabhängigen Zeilen der Matrix \mathbf{A} ist höchstens gleich der Anzahl r_s der Zeilen von \mathbf{C} und damit höchstens gleich der Anzahl r_s der linear unabhängigen Spalten der Matrix \mathbf{A} , also $r_z \leq r_s$. Entsprechende Überlegungen lassen sich auch für die transponierte Matrix \mathbf{A}' vornehmen. Dabei ergibt sich, daß die Anzahl der linear unabhängigen Spalten von \mathbf{A} höchstens gleich der Anzahl der linear unabhängigen Zeilen sein kann, also $r_s \leq r_z$. Aus beiden Beziehungen folgt dann $r_z = r_s = r = \operatorname{rg} \mathbf{A} = \operatorname{rg} \mathbf{A}' \leq \min(m, n)$.

Besitzt eine $m \times n$ Matrix \mathbf{A} den Rang $\operatorname{rg} \mathbf{A} = m$, bezeichnet man sie als Matrix mit *vollem Zeilenrang* und im Fall von $\operatorname{rg} \mathbf{A} = n$ als Matrix mit *vollem Spaltenrang*. Gilt $\operatorname{rg} \mathbf{A} < \min(m, n)$, weist \mathbf{A} einen *Rangdefekt* auf.

Weiter sollen zwei Sätze über den Rang von Matrizenprodukten angegeben werden.

Satz: $\text{rg}(\mathbf{AB}) \leq \min(\text{rg}\mathbf{A}, \text{rg}\mathbf{B})$. (132.3)

Beweis: Es sei $\mathbf{A} = (a_{ij})$ eine $m \times n$ Matrix und $\mathbf{B} = (b_{ij})$ eine $n \times p$ Matrix. Dann besteht die Spalte j des Produktes \mathbf{AB} mit $\sum_{k=1}^n a_{ik}b_{kj}$ für $i \in \{1, \dots, m\}$ aus Linearkombinationen der Spalten von \mathbf{A} , so daß die Anzahl der linear unabhängigen Spalten in \mathbf{AB} nicht die in \mathbf{A} überschreiten kann, also $\text{rg}\mathbf{AB} \leq \text{rg}\mathbf{A}$. Weiter sind die Zeilen von \mathbf{AB} Linearkombinationen der Zeilen von \mathbf{B} , so daß $\text{rg}\mathbf{AB} \leq \text{rg}\mathbf{B}$ gilt. Beide Aussagen ergeben dann den Satz.

Satz: Für eine beliebige $m \times n$ Matrix \mathbf{A} und zwei beliebige reguläre $m \times m$ und $n \times n$ Matrizen \mathbf{B} und \mathbf{C} gilt $\text{rg}(\mathbf{BAC}) = \text{rg}\mathbf{A}$. (132.4)

Beweis: Mit (132.3) folgt $\text{rg}\mathbf{A} \geq \text{rg}(\mathbf{AC}) \geq \text{rg}(\mathbf{ACC}^{-1}) = \text{rg}\mathbf{A}$, so daß $\text{rg}\mathbf{A} = \text{rg}(\mathbf{AC})$ folgt. Weiter gilt $\text{rg}(\mathbf{AC}) \geq \text{rg}(\mathbf{BAC}) \geq \text{rg}(\mathbf{B}^{-1}\mathbf{BAC}) = \text{rg}(\mathbf{AC})$, so daß schließlich $\text{rg}\mathbf{A} = \text{rg}(\mathbf{AC}) = \text{rg}(\mathbf{BAC})$ und damit die Aussage folgt.

Die praktische Rangbestimmung wird mit Hilfe *elementarer Umformungen* vorgenommen, die, wie gezeigt wird, den Rang einer Matrix nicht ändern. Sie bestehen 1. in der Vertauschung von Zeilen (Spalten), 2. in der Multiplikation einer Zeile (Spalte) mit einem Skalar $c \neq 0$ und 3. in der Addition einer mit $c \neq 0$ multiplizierten Zeile (Spalte) zu einer anderen Zeile (Spalte). Die elementaren Umformungen werden ebenfalls bei der Berechnung inverser Matrizen oder bei der Lösung linearer Gleichungssysteme angewendet, was im folgenden Kapitel behandelt wird.

Die elementaren Umformungen von Zeilen ergeben sich aus der linksseitigen Multiplikation der $m \times n$ Matrix \mathbf{A} mit speziellen $m \times m$ Matrizen \mathbf{E} , den *elementaren Matrizen*. Gilt

$$\mathbf{E}_1 = \begin{vmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix}, \quad \mathbf{E}_2 = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & c & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix},$$

$$\mathbf{E}_3 = \begin{vmatrix} 1 & 0 & 0 & \dots & 0 \\ c & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{vmatrix}, \quad (132.5)$$

bewirkt $\mathbf{E}_1\mathbf{A}$ den Austausch der ersten und zweiten Zeile von \mathbf{A} , $\mathbf{E}_2\mathbf{A}$ die Multiplikation der zweiten Zeile von \mathbf{A} mit dem Skalar c und $\mathbf{E}_3\mathbf{A}$ die Addi-

tion der mit c multiplizierten Elemente der ersten Zeile von \mathbf{A} zu den entsprechenden Elementen der zweiten Zeile.

Für jede elementare Matrix \mathbf{E} existiert die inverse Matrix \mathbf{E}^{-1} , die die elementare Umformung rückgängig macht. Es gilt $\mathbf{E}_1^{-1} = \mathbf{E}_1$ und

$$\mathbf{E}_2^{-1} = \left| \begin{array}{cccccc} 1 & 0 & 0 & \dots & 0 & \\ 0 & 1/c & 0 & \dots & 0 & \\ 0 & 0 & 1 & \dots & 0 & \\ \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & \dots & 1 & \end{array} \right|, \quad \mathbf{E}_3^{-1} = \left| \begin{array}{cccccc} 1 & 0 & 0 & \dots & 0 & \\ -c & 1 & 0 & \dots & 0 & \\ 0 & 0 & 1 & \dots & 0 & \\ \dots & \dots & \dots & \dots & \dots & \\ 0 & 0 & 0 & \dots & 1 & \end{array} \right|, \quad (132.6)$$

denn mit $\mathbf{E}_1 \mathbf{E}_1^{-1} = \mathbf{I}$ und $\mathbf{E}_1^{-1} \mathbf{E}_1 = \mathbf{I}$ sowie den entsprechenden Gleichungen für \mathbf{E}_2 und \mathbf{E}_3 ist (131.12) erfüllt.

Rechtsseitige Multiplikationen einer Matrix mit den transponierten elementaren Matrizen bewirken die Spaltenumformungen. So wird durch $\mathbf{A} \mathbf{E}'_1$ die erste und zweite Spalte von \mathbf{A} ausgetauscht, und durch $\mathbf{A} \mathbf{E}'_2$ und $\mathbf{A} \mathbf{E}'_3$ ergeben sich entsprechende Spaltenoperationen, wobei $\mathbf{E}'_1, \mathbf{E}'_2$ und \mathbf{E}'_3 jetzt $n \times n$ Matrizen bedeuten.

Wie bereits erwähnt, gilt der

Satz: Elementare Umformungen ändern den Rang einer Matrix nicht. (132.7)

Beweis: Da die elementaren Matrizen regulär sind, folgt aus (132.4), daß elementare Umformungen den Rang einer Matrix nicht ändern.

Bei einer praktischen Rangbestimmung bringt man die $m \times n$ Matrix \mathbf{A} durch elementare Zeilenumformungen, deren Anzahl k betrage, zunächst auf die Gestalt

$$\mathbf{E}^{(k)} \dots \mathbf{E}^{(2)} \mathbf{E}^{(1)} \mathbf{A} = \mathbf{P} \mathbf{A} = \mathbf{B} \quad (132.8)$$

mit

$$\mathbf{B} = \left| \begin{array}{cccccc} 1 & b_{12} & b_{13} & \dots & b_{1r} & \dots & b_{1n} \\ 0 & 1 & b_{23} & \dots & b_{2r} & \dots & b_{2n} \\ 0 & 0 & 1 & \dots & b_{3r} & \dots & b_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & \dots & b_{rn} \\ 0 & 0 & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & \dots & 0 \end{array} \right|,$$

worin \mathbf{P} die Dimensionen $m \times m$ besitzt. Man geht so vor, daß sämtliche Elemente der ersten Zeile durch das erste Element dividiert werden, nötigenfalls

nach Zeilenvertauschungen, um das erste Element der ersten Zeile von Null verschieden zu erhalten. Anschließend wird die erste Zeile mit entsprechenden Skalaren multipliziert und zu den folgenden Zeilen addiert, um Nullelemente in der ersten Spalte der zweiten Zeile und der folgenden zu erzeugen. Dann werden die Elemente der zweiten Zeile durch das in dieser Zeile an zweiter Stelle stehende Element dividiert, nachdem, falls erforderlich, zuvor diese Zeile mit einer der folgenden vertauscht wurde. Anschließend werden mit Hilfe der zweiten Zeile Nullelemente in der zweiten Spalte der dritten Zeile und der folgenden Zeilen erzeugt. Entsprechend werden sämtliche Zeilen bearbeitet, bis die Matrix \mathbf{B} in (132.8) erhalten wird.

Durch rechtsseitige Multiplikationen von \mathbf{B} in (132.8) mit den transponierten elementaren Matrizen \mathbf{E}' folgen dann die Spaltenumformungen. Durch Multiplikationen der ersten Spalte mit entsprechenden Skalaren und durch Additionen zu den folgenden Spalten lassen sich Nullen für die ersten Elemente der folgenden Spalten erzeugen. Wiederholt man die entsprechenden Umformungen für die zweite bis zur r -ten Spalte, ergibt sich, falls l Umformungen benötigt werden

$$\mathbf{B}\mathbf{E}'^{(k+1)}\mathbf{E}'^{(k+2)}\dots\mathbf{E}'^{(k+l)} = \mathbf{B}\mathbf{Q} = \mathbf{P}\mathbf{A}\mathbf{Q} = \left| \begin{array}{cc} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right|. \quad (132.9)$$

\mathbf{I}_r ist die $r \times r$ Einheitsmatrix, und \mathbf{Q} besitzt die Dimensionen $n \times n$.

Mit (122.2) ist offensichtlich, daß die aus \mathbf{A} durch elementare Umformungen erhaltene Matrix (132.9) r linear unabhängige Zeilen und Spalten besitzt, so daß $\text{rg}\mathbf{A} = r$ folgt.

Wie bereits erwähnt, existieren für elementare Matrizen \mathbf{E} die inversen Matrizen \mathbf{E}^{-1} . Durch linksseitige Multiplikation von (132.8) mit $\mathbf{P}^{-1} = (\mathbf{E}^{(1)})^{-1}(\mathbf{E}^{(2)})^{-1}\dots(\mathbf{E}^{(k)})^{-1}$ wegen (131.14) und durch rechtsseitige Multiplikation von (132.9) mit $\mathbf{Q}^{-1} = (\mathbf{E}'^{(k+1)})^{-1}\dots(\mathbf{E}'^{(k+2)})^{-1}(\mathbf{E}'^{(k+l)})^{-1}$ kann aus (132.9) \mathbf{A} zurückerhalten werden. Da $\mathbf{P}\mathbf{P}^{-1} = \mathbf{P}^{-1}\mathbf{P} = \mathbf{I}$ und $\mathbf{Q}\mathbf{Q}^{-1} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{I}$ gelten, sind \mathbf{P} und \mathbf{Q} nach (131.12) reguläre Matrizen. Folglich ergibt sich der

Satz: Jede $m \times n$ Matrix \mathbf{A} mit $\text{rg}\mathbf{A} = r$ läßt sich mit den regulären $m \times m$ und $n \times n$ Matrizen \mathbf{P} und \mathbf{Q} zerlegen in

$$\mathbf{P}\mathbf{A}\mathbf{Q} = \left| \begin{array}{cc} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right|. \quad (132.10)$$

Löst man (132.10) nach \mathbf{A} auf, ergibt sich

$$\mathbf{A} = \mathbf{P}^{-1} \left| \begin{array}{cc} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right| \mathbf{Q}^{-1} \quad \text{oder} \quad \mathbf{A} = \mathbf{R}\mathbf{S}, \quad (132.11)$$

worin für die $m \times r$ Matrix \mathbf{R} gilt $\text{rg}\mathbf{R} = r$ und für die $r \times n$ Matrix \mathbf{S} entsprechend $\text{rg}\mathbf{S} = r$, denn aus den regulären Matrizen \mathbf{P}^{-1} und \mathbf{Q}^{-1} wurden $m - r$ Spalten beziehungsweise $n - r$ Zeilen gestrichen, um \mathbf{R} und \mathbf{S} zu erhalten. Die Zerlegung (132.11) bezeichnet man als *Rangfaktorisierung* von \mathbf{A} . Da die Matrizen \mathbf{P} und \mathbf{Q} wegen der verschiedenen Möglichkeiten der elementaren Transformationen nicht eindeutig sind, so daß die von Null und Eins verschiedenen Elemente von \mathbf{B} in (132.8) sich unterschiedlich ergeben können, ist auch die Rangfaktorisierung nicht eindeutig.

133 Berechnung inverser Matrizen

Nach der Definition (131.12) der Inversen einer regulären Matrix soll jetzt die Bedingung für die Regularität einer Matrix angegeben werden.

Satz: Die $n \times n$ Matrix \mathbf{A} ist genau dann regulär, wenn \mathbf{A} vollen Rang besitzt, wenn also $\text{rg}\mathbf{A} = n$ gilt. (133.1)

Beweis: Wie aus dem Beweis von (132.10) sich ergibt und wie bei der folgenden Behandlung des Gaußschen Algorithmus noch einmal gezeigt wird, läßt sich im Falle von $\text{rg}\mathbf{A} = n$ die Matrix \mathbf{A} durch elementare Zeilenumformungen mit Hilfe der $n \times n$ Matrix \mathbf{R} in eine Einheitsmatrix überführen, also $\mathbf{R}\mathbf{A} = \mathbf{I}$. Entsprechend gilt für die Spaltenumformungen mit der $n \times n$ Matrix \mathbf{S} die Beziehung $\mathbf{A}\mathbf{S} = \mathbf{I}$. Dann ist aber $\mathbf{R}\mathbf{A}\mathbf{S} = \mathbf{S}$ und $\mathbf{R}\mathbf{A}\mathbf{S} = \mathbf{R}$, so daß $\mathbf{R} = \mathbf{S}$ und wegen (131.12) $\mathbf{R} = \mathbf{S} = \mathbf{A}^{-1}$ folgt. Ist andererseits \mathbf{A} regulär, dann existiert die reguläre Inverse \mathbf{A}^{-1} , so daß mit (132.4) $\text{rg}(\mathbf{A}^{-1}\mathbf{A}) = \text{rg}\mathbf{A} = \text{rg}\mathbf{I} = n$ und damit die Aussage sich ergibt.

Im folgenden sollen Methoden zur Inversion regulärer Matrizen und zur Lösung linearer Gleichungssysteme behandelt werden.

a) Gaußscher Algorithmus

Beim *Gaußschen Algorithmus* wird wie in (132.8) die reguläre $n \times n$ Matrix \mathbf{A} zunächst durch $p = (n - 1) + (n - 2) + \dots + 1$ linksseitige Multiplikationen mit elementaren Matrizen vom Typ \mathbf{E}_3 in (132.5) auf die Form einer oberen Dreiecksmatrix gebracht, was man als *Gaußsche Elimination* bezeichnet,

$$\mathbf{B} = (\mathbf{E}_3^{(p)} \mathbf{E}_3^{(p-1)} \dots \mathbf{E}_3^{(1)}) \mathbf{A} = \mathbf{C}^{-1} \mathbf{A}, \quad (133.2)$$

und zwar werden im ersten Eliminationsschritt durch Multiplikation der ersten Zeile von \mathbf{A} mit entsprechenden Skalaren und durch Addition zu den folgenden Zeilen Nullelemente in der ersten Spalte von \mathbf{A} unterhalb des Diagonalelementes erzeugt. Die $n - 1$ elementaren Umformungen des 1. Eliminationsschrittes

lassen sich mit $\mathbf{A} = (a_{ij})$ in der folgenden Matrix zusammenfassen

$$\mathbf{E}_3^{(n-1)} \mathbf{E}_3^{(n-2)} \dots \mathbf{E}_3^{(1)} = \left| \begin{array}{cccccc} 1 & 0 & 0 & \dots & 0 \\ -a_{21}/a_{11} & 1 & 0 & \dots & 0 \\ -a_{31}/a_{11} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -a_{n1}/a_{11} & 0 & 0 & \dots & 1 \end{array} \right|. \quad (133.3)$$

Im zweiten Eliminationsschritt werden die Nullelemente unterhalb des Diagonalelementes der zweiten Spalte erzeugt, indem die zweite Zeile mit entsprechenden Skalaren multipliziert zu den folgenden Zeilen addiert wird. Im i -ten Eliminationsschritt besitzt die Matrix (133.3) das folgende Aussehen, falls $a_{ij}^{(i)}$ die Elemente der mit $i - 1$ Eliminationsschritten umgeformten Matrix \mathbf{A} bedeuten

$$\left| \begin{array}{cccccc} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & -a_{i+1,i}^{(i)}/a_{ii}^{(i)} & 1 & \dots & 0 \\ 0 & 0 & \dots & -a_{i+2,i}^{(i)}/a_{ii}^{(i)} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -a_{ni}^{(i)}/a_{ii}^{(i)} & 0 & \dots & 1 \end{array} \right|. \quad (133.4)$$

Mit dem $(n - 1)$ -ten Eliminationsschritt ergibt sich dann aus \mathbf{A} die Matrix \mathbf{B} in (133.2). Allgemein erhält man die Elemente von \mathbf{B} aus \mathbf{A} durch

$$a_{jk}^{(i+1)} = a_{jk}^{(i)} - \frac{a_{ji}^{(i)} a_{ik}^{(i)}}{a_{ii}^{(i)}} \quad (133.5)$$

für $i \in \{1, \dots, n - 1\}$, $j \in \{i + 1, \dots, n\}$ und $k \in \{i, \dots, n\}$.

Durch n linksseitige Multiplikationen mit elementaren Matrizen vom Typ \mathbf{E}_2 wird anschließend die Matrix \mathbf{B} in die Einheits-Dreiecksmatrix \mathbf{F} überführt

$$\mathbf{F} = (\mathbf{E}_2^{(n)} \mathbf{E}_2^{(n-1)} \dots \mathbf{E}_2^{(1)}) \mathbf{B} = \mathbf{D}^{-1} \mathbf{B} = \mathbf{D}^{-1} \mathbf{C}^{-1} \mathbf{A} \quad (133.6)$$

mit $\mathbf{D}^{-1} = \text{diag}(1/b_{11}, 1/b_{22}, \dots, 1/b_{nn})$, falls $\mathbf{B} = (b_{ij})$ gilt.

Die obere Einheits-Dreiecksmatrix \mathbf{F} läßt sich nun durch die sogenannte *Rückrechnung* mit p weiteren elementaren Umformungen der Zeilen durch die elementaren Matrizen vom Typ \mathbf{E}_3 in eine Einheitsmatrix überführen

$$(\mathbf{E}_3^{(2p)} \mathbf{E}_3^{(2p-1)} \dots \mathbf{E}_3^{(p+1)}) \mathbf{F} = \mathbf{F}^{-1} \mathbf{F} = \mathbf{I}, \quad (133.7)$$

und zwar wird im ersten Schritt der Rückrechnung, die $n - 1$ elementare Umformungen enthält, die letzte Zeile von \mathbf{F} , in der nur das letzte Element von Null verschieden ist, mit entsprechenden Skalaren multipliziert zu den darüber liegenden Zeilen addiert, um Nullelemente oberhalb des Diagonalelementes in der letzten Spalte zu erzeugen. Mit $\mathbf{F} = (f_{ij})$ erhält man

$$\mathbf{E}_3^{(p+n-1)} \dots \mathbf{E}_3^{(p+1)} \left| \begin{array}{cccc|c} 1 & 0 & \dots & 0 & -f_{1n} \\ 0 & 1 & \dots & 0 & -f_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -f_{n-1,n} \\ 0 & 0 & \dots & 0 & 1 \end{array} \right|. \quad (133.8)$$

Mit den übrigen Spalten wird dann entsprechend verfahren, bis die Einheitsmatrix in (133.7) erhalten wird, wobei die Matrix des i -ten Schrittes der Rückrechnung den folgenden Aufbau besitzt

$$\left| \begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & -f_{1,n-i+1} & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & -f_{2,n-i+1} & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & -f_{n-i,n-i+1} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 \end{array} \right|. \quad (133.9)$$

Vereinigt man die Schritte (133.2), (133.6) und (133.7), ergibt sich

$$(\mathbf{F}^{-1} \mathbf{D}^{-1} \mathbf{C}^{-1}) \mathbf{A} = \mathbf{A}^{-1} \mathbf{A} = \mathbf{I}, \quad (133.10)$$

so daß die elementaren Umformungen die inverse Matrix \mathbf{A}^{-1} bestimmen. Praktisch kann man so vorgehen, daß die $n \times 2n$ Blockmatrix $|\mathbf{A}, \mathbf{I}|$ mit (133.6) in $|\mathbf{F}, \mathbf{D}^{-1} \mathbf{C}^{-1}|$ umgeformt wird, woraus durch die Rückrechnung (133.10) die Inverse \mathbf{A}^{-1} folgt, also

$$|\mathbf{I}, \mathbf{F}^{-1} \mathbf{D}^{-1} \mathbf{C}^{-1}| = |\mathbf{I}, \mathbf{A}^{-1}|. \quad (133.11)$$

Das zeilenweise Vorgehen beim Aufbau der oberen Dreiecksmatrix \mathbf{B} in (133.2) führt nur dann zum Ziel, falls im Verlauf der Elimination keine Nullelemente auf der Diagonalen auftreten. Nullelemente erhält man immer dann, wenn führende Hauptuntermatrizen von \mathbf{A} singulär sind. Führende Hauptuntermatrizen sind quadratische Blockmatrizen auf der Diagonalen, die das erste Element der ersten Zeile und ersten Spalte enthalten. Als Beispiel sei auf die Matrix (155.7) verwiesen. In einem solchen Fall sind zur Beseitigung der singulären Untermatrizen Zeilenvertauschungen und bei symmetrischen Matrizen zur Wahrung der Symmetrie Zeilen- und Spaltenvertauschungen vorzunehmen, oder man arbeitet mit einer *Pivotstrategie*, indem bei jedem Eliminationsschritt durch Zeilen- und Spaltenvertauschungen das absolut größte

Element als Diagonalelement, durch das zu dividieren ist, benutzt wird (RUTISHAUSER 1976, Bd.1, S.28; STIEFEL 1970, S.21).

Beispiel: Mit dem Gaußschen Algorithmus werde die folgende 3×3 *Vandermonde*-Matrix \mathbf{V} (GREGORY und KARNEY 1969, S.27) invertiert, wobei zur Rechenkontrolle eine Spalte der Zeilensummen mitgeführt werden soll. An den Zeilenenden sind jeweils die Faktoren angegeben, mit denen die Zeilen in den einzelnen Eliminations- und Rückrechnungsschritten zu multiplizieren sind.

$$|\mathbf{V}, \mathbf{I}, \text{Summe}| = \left| \begin{array}{ccc|ccc|c} 1 & 1 & 1 & 1 & 0 & 0 & 4 \\ 1 & 2 & 3 & 0 & 1 & 0 & 7 \\ 1 & 4 & 9 & 0 & 0 & 1 & 15 \end{array} \right| \begin{array}{l} (-1), (-1) \\ \\ \end{array} \quad (133.12)$$

Ende des 1. Eliminationsschrittes

$$\left| \begin{array}{ccc|ccc|c} 1 & 1 & 1 & 1 & 0 & 0 & 4 \\ 0 & 1 & 2 & -1 & 1 & 0 & 3 \\ 0 & 3 & 8 & -1 & 0 & 1 & 11 \end{array} \right| \begin{array}{l} \\ (-3) \\ \end{array}$$

Ende des 2. Eliminationsschrittes

$$\left| \begin{array}{ccc|ccc|c} 1 & 1 & 1 & 1 & 0 & 0 & 4 \\ 0 & 1 & 2 & -1 & 1 & 0 & 3 \\ 0 & 0 & 2 & 2 & -3 & 1 & 2 \end{array} \right| \begin{array}{l} (+1) \\ (+1) \\ (+1/2) \end{array}$$

Erzeugung der Einheits-Dreiecksmatrix

$$\left| \begin{array}{ccc|ccc|c} 1 & 1 & 1 & 1 & 0 & 0 & 4 \\ 0 & 1 & 2 & -1 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 & -3/2 & 1/2 & 1 \end{array} \right| \begin{array}{l} \\ \\ (-2), (-1) \end{array}$$

Ende des ersten Rückrechnungsschrittes

$$\left| \begin{array}{ccc|ccc|c} 1 & 1 & 0 & 0 & 3/2 & -1/2 & 3 \\ 0 & 1 & 0 & -3 & 4 & -1 & 1 \\ 0 & 0 & 1 & 1 & -3/2 & 1/2 & 1 \end{array} \right| \begin{array}{l} \\ (-1) \\ \end{array}$$

Ende des zweiten Rückrechnungsschrittes

$$|\mathbf{I}, \mathbf{V}^{-1}, \text{Summe}| = \left| \begin{array}{ccc|ccc|c} 1 & 0 & 0 & 3 & -5/2 & 1/2 & 2 \\ 0 & 1 & 0 & -3 & 4 & -1 & 1 \\ 0 & 0 & 1 & 1 & -3/2 & 1/2 & 1 \end{array} \right|$$

Die Faktorisierung mit Hilfe des Gaußschen Algorithmus kann man zusammenfassen in dem

Satz: Jede Matrix \mathbf{A} mit regulären führenden Hauptuntermatrizen läßt sich eindeutig zerlegen in das Produkt dreier regulärer Matrizen und zwar einer

unteren Einheits-Dreiecksmatrix \mathbf{C} , einer Diagonalmatrix \mathbf{D} und einer oberen Einheits-Dreiecksmatrix \mathbf{F} , also $\mathbf{A} = \mathbf{CDF}$. (133.13)

Beweis: Linksseitige Multiplikation von (133.6) mit \mathbf{D} ergibt $\mathbf{C}^{-1}\mathbf{A} = \mathbf{DF}$. Wie aus (133.2) bis (133.4) ersichtlich, ist \mathbf{C}^{-1} eine untere Einheits-Dreiecksmatrix, während \mathbf{F} und \mathbf{D} in (133.6) eine obere Einheits-Dreiecksmatrix beziehungsweise eine Diagonalmatrix darstellen. Nach (131.16) ist \mathbf{C} die Inverse von \mathbf{C}^{-1} und ebenfalls untere Einheits-Dreiecksmatrix, so wie \mathbf{F} und \mathbf{F}^{-1} mit (133.7) bis (133.9) obere Einheits-Dreiecksmatrizen sind. Somit folgt $\mathbf{A} = \mathbf{CDF}$, wobei \mathbf{C} , \mathbf{D} und \mathbf{F} regulär sind, da sie aus elementaren Matrizen entstanden sind. Um die Eindeutigkeit der Faktorisierung zu beweisen, soll $\mathbf{A} = \mathbf{C}_1\mathbf{D}_1\mathbf{F}_1 = \mathbf{C}_2\mathbf{D}_2\mathbf{F}_2$ angenommen werden, woraus $\mathbf{C}_2^{-1}\mathbf{C}_1\mathbf{D}_1 = \mathbf{D}_2\mathbf{F}_2\mathbf{F}_1^{-1}$ folgt. Die Produkte $\mathbf{C}_2^{-1}\mathbf{C}_1$ und $\mathbf{F}_2\mathbf{F}_1^{-1}$ stellen untere beziehungsweise obere Einheits-Dreiecksmatrizen dar, so daß sich Identität nur mit $\mathbf{C}_2^{-1}\mathbf{C}_1 = \mathbf{I}$, $\mathbf{F}_2\mathbf{F}_1^{-1} = \mathbf{I}$ und $\mathbf{D}_1 = \mathbf{D}_2$ einstellen kann. Da die Matrizen regulär sind, folgt mit $\mathbf{C}_2\mathbf{C}_2^{-1}\mathbf{C}_1 = \mathbf{C}_1$ die Identität $\mathbf{C}_1 = \mathbf{C}_2$ und entsprechend $\mathbf{F}_1 = \mathbf{F}_2$, so daß die Aussage sich ergibt.

b) Lösung linearer Gleichungssysteme

Die Begriffe Elimination und Rückrechnung entstammen dem Verfahren bei der Lösung linearer Gleichungssysteme. Die $n \times n$ Matrix \mathbf{A} und der $n \times 1$ Vektor \mathbf{l} seien gegeben. Gesucht wird der $n \times 1$ Vektor $\boldsymbol{\beta}$, für den

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{l} \quad (133.14)$$

oder

$$\begin{aligned} a_{11}\beta_1 + a_{12}\beta_2 + \dots + a_{1n}\beta_n &= l_1 \\ a_{21}\beta_1 + a_{22}\beta_2 + \dots + a_{2n}\beta_n &= l_2 \\ \dots & \\ a_{n1}\beta_1 + a_{n2}\beta_2 + \dots + a_{nn}\beta_n &= l_n \end{aligned}$$

gilt. Man bezeichnet (133.14) als *lineares Gleichungssystem* mit der *Koeffizientenmatrix* \mathbf{A} , den *unbekannten Parametern* $\boldsymbol{\beta}$ und den *Absolutgliedern* \mathbf{l} . Ist $\text{rg}\mathbf{A} = n$, sind die Parameter $\boldsymbol{\beta}$ des Gleichungssystems mit

$$\boldsymbol{\beta} = \mathbf{A}^{-1}\mathbf{l} \quad (133.15)$$

eindeutig bestimmt. Lineare Gleichungssysteme für $m \times n$ Koeffizientenmatrizen mit beliebigem Rang werden im Kapitel 154 behandelt.

Für die Berechnung von $\boldsymbol{\beta}$ ist es nicht notwendig, \mathbf{A}^{-1} zu bestimmen. Mit $\mathbf{F} = \mathbf{D}^{-1}\mathbf{C}^{-1}\mathbf{A}$ aus (133.6) und $\mathbf{g} = \mathbf{D}^{-1}\mathbf{C}^{-1}\mathbf{l}$ wird (133.14) in die Dreiecksform

$$\mathbf{F}\boldsymbol{\beta} = \mathbf{g} \quad (133.16)$$

überführt, was einer schrittweisen Elimination der Parameter β_i entspricht. Hieraus ergeben sich dann die β_i mit $\mathbf{F} = (f_{ij})$ und $\mathbf{g} = (g_i)$ durch Rückrechnung

$$\beta_n = g_n, \beta_{n-1} = g_{n-1} - f_{n-1,n}\beta_n, \dots$$

also

$$\beta_i = g_i - \sum_{j=i+1}^n f_{ij}\beta_j \quad \text{für } i \in \{n-1, \dots, 1\} \quad (133.17)$$

oder allgemein

$$\boldsymbol{\beta} = \mathbf{F}^{-1}\mathbf{g}. \quad (133.18)$$

Die Identität von (133.17) und (133.18) folgt aus der Multiplikation der Matrizen (133.9) der einzelnen Schritte der Rückrechnung.

c) Gauß–Jordan–Methode

Eine kompaktere Form der Inversion einer regulären Matrix erhält man mit der *Gauß–Jordan–Methode*. Hierbei werden mit jedem Eliminationsschritt nicht nur wie bei der Gaußschen Elimination Nullelemente in der jeweiligen Spalte unterhalb, sondern auch oberhalb des Diagonalelementes und außerdem eine Eins auf der Diagonalen erzeugt, so daß in n Eliminations- beziehungsweise Reduktionsschritten die Einheitsmatrix erhalten wird

$$\mathbf{T}^{(n)}\mathbf{T}^{(n-1)} \dots \mathbf{T}^{(1)}\mathbf{A} = \mathbf{I} \quad \text{mit} \quad \mathbf{T}^{(n)} \dots \mathbf{T}^{(1)} = \mathbf{A}^{-1}, \quad (133.19)$$

falls $\mathbf{T}^{(i)}$ die Matrix der elementaren Umformungen des i -ten Reduktionsschrittes bedeutet. $\mathbf{T}^{(1)}$ ist identisch mit (133.3), falls dort das erste Element durch $1/a_{11}$ ersetzt wird. Für $\mathbf{T}^{(i)}$ gilt

$$\mathbf{T}^{(i)} = \begin{pmatrix} 1 & 0 & \dots & 0 & -a_{1i}^{(i)}/a_{ii}^{(i)} & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & -a_{2i}^{(i)}/a_{ii}^{(i)} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1/a_{ii}^{(i)} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & -a_{i+1,i}^{(i)}/a_{ii}^{(i)} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & -a_{ni}^{(i)}/a_{ii}^{(i)} & 0 & \dots & 1 \end{pmatrix}, \quad (133.20)$$

falls $\mathbf{T}^{(i-1)} \dots \mathbf{T}^{(1)}\mathbf{A} = (a_{ij}^{(i)})$ bedeutet. Die Inverse \mathbf{A}^{-1} wird unmittelbar in der Matrix \mathbf{A} aufgebaut, indem sowohl die Produkte $\mathbf{T}^{(i-1)} \dots \mathbf{T}^{(1)}\mathbf{A}$ zur

Erzeugung der Einheitsmatrix als auch die Produkte $\mathbf{T}^{(i-1)} \dots \mathbf{T}^{(1)}$ zur Erzeugung der Inversen sukzessiv gebildet werden. Dies erreicht man dadurch, daß die Spalte, in der die Nullelemente und die Eins als Diagonalelement erzeugt werden, die Reduktionsfaktoren aus (133.20) aufnimmt, mit denen alle übrigen Spalten durchreduziert werden. Es gilt daher für $i, j, k \in \{1, \dots, n\}$

$$\begin{aligned} a_{jk}^{(i+1)} &= a_{jk}^{(i)} - \frac{a_{ji}^{(i)} a_{ik}^{(i)}}{a_{ii}^{(i)}} \quad \text{für } j \neq i \quad \text{und } k \neq i \\ a_{ik}^{(i+1)} &= a_{ik}^{(i)} / a_{ii}^{(i)} \quad \text{für } k \neq i \\ a_{ki}^{(i+1)} &= -a_{ki}^{(i)} / a_{ii}^{(i)} \quad \text{für } k \neq i \\ a_{ii}^{(i+1)} &= 1 / a_{ii}^{(i)}. \end{aligned} \tag{133.21}$$

Die beiden letzten Gleichungen entsprechen den Reduktionsfaktoren in (133.20).

Wie beim Gaußschen Algorithmus kann zeilenweise nur dann vorgegangen werden, falls im Verlauf der Reduktion keine Nullelemente auf der Diagonalen auftreten. Ist das der Fall, müssen Zeilen- oder Zeilen- und Spaltenvertauschungen vorgenommen werden, oder man muß mit einer Pivotisierung arbeiten, wozu sich das Gauß–Jordan–Verfahren gut eignet.

Beispiel: Als Beispiel soll wieder die 3×3 Vandermonde–Matrix \mathbf{V} in (133.12) invertiert werden. Am Ende des ersten, zweiten und letzten Reduktionsschrittes erhält man die Matrizen

$$\left| \begin{array}{ccc} 1 & 1 & 1 \\ -1 & 1 & 2 \\ -1 & 3 & 8 \end{array} \right|, \left| \begin{array}{ccc} 2 & -1 & -1 \\ -1 & 1 & 2 \\ 2 & -3 & 2 \end{array} \right|, \left| \begin{array}{ccc} 3 & -5/2 & 1/2 \\ -3 & 4 & -1 \\ 1 & -3/2 & 1/2 \end{array} \right| = \mathbf{V}^{-1}.$$

d) Symmetrische Matrizen und Cholesky–Verfahren

Die Inversen symmetrischer Matrizen, die nach (131.17) ebenfalls symmetrisch sind, interessieren besonders im Hinblick auf die im Abschnitt 3 zu behandelnden symmetrischen Normalgleichungen für die Parameterschätzung. Bei der Berechnung mit Hilfe von elektronischen Datenverarbeitungsanlagen kann Speicherplatz gespart werden, da die Elemente der zu invertierenden Matrix und ihrer Inversen unterhalb der Diagonalelemente nicht benötigt werden. Für symmetrische Matrizen sind in (133.5) $a_{ji}^{(i)}$ durch $a_{ij}^{(i)}$ und $k \in \{i, \dots, n\}$ durch $k \in \{j, \dots, n\}$ zu ersetzen und entsprechende Substitutionen in (133.21) durchzuführen.

Mit (133.13) erhält man für symmetrische Matrizen $\mathbf{A} = \mathbf{A}' = \mathbf{C}\mathbf{D}\mathbf{F} = \mathbf{F}'\mathbf{D}\mathbf{C}'$, und da die Faktorisierung eindeutig ist, $\mathbf{C} = \mathbf{F}'$ und $\mathbf{F} = \mathbf{C}'$ sowie

$$\mathbf{A} = \mathbf{C}\mathbf{D}\mathbf{C}' . \quad (133.22)$$

Ist \mathbf{A} eine positiv definite Matrix, dann sind, wie aus (143.1) und (143.3) folgen wird, alle Untermatrizen auf der Diagonalen regulär und nach (143.4) die Diagonalelemente von \mathbf{D} positiv. Mit $\mathbf{D} = \text{diag}(d_{11}, \dots, d_{nn})$ läßt sich daher die Matrix $\mathbf{D}^{1/2} = \text{diag}(d_{11}^{1/2}, \dots, d_{nn}^{1/2})$ definieren, so daß gilt $\mathbf{A} = (\mathbf{C}\mathbf{D}^{1/2})(\mathbf{D}^{1/2}\mathbf{C}') = \mathbf{G}\mathbf{G}'$. Weiter ist $\mathbf{D}^{1/2}\mathbf{C}' = \mathbf{D}^{-1/2}\mathbf{C}^{-1}\mathbf{A}$ oder mit (133.2) $\mathbf{D}^{1/2}\mathbf{C}' = \mathbf{D}^{-1/2}\mathbf{B}$, folglich

$$\mathbf{A} = \mathbf{G}\mathbf{G}' \quad \text{mit} \quad \mathbf{G}' = \mathbf{D}^{-1/2}\mathbf{B} , \quad (133.23)$$

wobei \mathbf{G} eine untere Dreiecksmatrix bedeutet. Die Zerlegung (133.23) bezeichnet man als *Cholesky-Faktorisierung*. Sie ist wie (133.22) eindeutig.

Eine Zerlegung nach (133.23) zur Inversion einer Matrix oder zur Lösung eines Gleichungssystems bezeichnet man als Cholesky-Verfahren. Dabei ist genau wie beim Gaußschen Algorithmus vorzugehen, nur wird die Matrix \mathbf{A} nicht in die obere Einheits-Dreiecksmatrix $\mathbf{D}^{-1}\mathbf{B}$, sondern in die obere Dreiecksmatrix $\mathbf{G}' = \mathbf{D}^{-1/2}\mathbf{B}$ überführt, woran sich die Rückrechnung anschließt.

Die Cholesky-Zerlegung läßt sich auch direkt ohne Zuhilfenahme des Gaußschen Algorithmus berechnen. Mit $\mathbf{G}' = (g_{ij})$, $\mathbf{A} = (a_{ij})$ und $\mathbf{G}\mathbf{G}' = \mathbf{A}$ erhält man zum Beispiel für eine 4×4 Matrix \mathbf{A}

$$\begin{array}{c} \left| \begin{array}{cccc} g_{11} & 0 & 0 & 0 \\ g_{12} & g_{22} & 0 & 0 \\ g_{13} & g_{23} & g_{33} & 0 \\ g_{14} & g_{24} & g_{34} & g_{44} \end{array} \right| \left| \begin{array}{cccc} g_{11} & g_{12} & g_{13} & g_{14} \\ 0 & g_{22} & g_{23} & g_{24} \\ 0 & 0 & g_{33} & g_{34} \\ 0 & 0 & 0 & g_{44} \end{array} \right| \\ = \left| \begin{array}{cccc} a_{11} & a_{12} & a_{13} & a_{14} \\ & a_{22} & a_{23} & a_{24} \\ & & a_{33} & a_{34} \\ & & & a_{44} \end{array} \right| . \end{array}$$

Durch Multiplikation folgt für die Diagonalelemente von \mathbf{A}

$$\begin{array}{ll} g_{11}^2 = a_{11} & g_{13}^2 + g_{23}^2 + g_{33}^2 = a_{33} \\ g_{12}^2 + g_{22}^2 = a_{22} & g_{14}^2 + g_{24}^2 + g_{34}^2 + g_{44}^2 = a_{44} \end{array}$$

oder allgemein für eine $n \times n$ Matrix \mathbf{A}

$$g_{ii} = (a_{ii} - \sum_{k=1}^{i-1} g_{ki}^2)^{1/2} \quad \text{für} \quad i \in \{1, \dots, n\} \quad (133.24)$$

und für die Elemente oberhalb der Diagonalen von \mathbf{A}

$$\begin{aligned} g_{11}g_{12} &= a_{12} & g_{12}g_{13} + g_{22}g_{23} &= a_{23} \\ g_{11}g_{13} &= a_{13} & g_{12}g_{14} + g_{22}g_{24} &= a_{24} \\ g_{11}g_{14} &= a_{14} & g_{13}g_{14} + g_{23}g_{24} + g_{33}g_{34} &= a_{34} \end{aligned}$$

oder allgemein

$$g_{ij} = (a_{ij} - \sum_{k=1}^{i-1} g_{ki}g_{kj})/g_{ii} \quad \text{für } j \in \{i+1, \dots, n\}, i \in \{1, \dots, n\}. \quad (133.25)$$

Man kann also \mathbf{G}' zeilenweise aus der Multiplikation von Spalten berechnen, und es gilt $g_{ij} = 0$ für alle Elemente oberhalb des letzten von Null verschiedenen Elementes a_{ij} einer Spalte j . Das Profil der Matrix \mathbf{A} , das durch die Abstände der Diagonalelemente jeder Spalte von dem letzten von Null verschiedenen Element der jeweiligen Spalte definiert wird, bestimmt also den Rechenaufwand bei der Ermittlung von \mathbf{G}' .

Für die Lösung des Gleichungssystems $\mathbf{A}\boldsymbol{\beta} = \mathbf{G}\mathbf{G}'\boldsymbol{\beta} = \mathbf{l}$ durch die Cholesky-Zerlegung erhält man

$$\mathbf{G}\mathbf{s} = \mathbf{l} \quad \text{mit} \quad \mathbf{G}'\boldsymbol{\beta} = \mathbf{s},$$

so daß man sich den Absolutgliedvektor \mathbf{l} als zusätzliche Spalte von \mathbf{A} vorstellen kann und für \mathbf{s} erhält

$$s_i = (l_i - \sum_{k=1}^{i-1} g_{ki}s_k)/g_{ii} \quad \text{für } i \in \{1, \dots, n\}. \quad (133.26)$$

Der Vektor $\boldsymbol{\beta}$ ergibt sich dann durch Rückrechnung wie in (133.17) mit $\beta_n = s_n/g_{nn}$ zu

$$\beta_i = (s_i - \sum_{k=i+1}^n g_{ik}\beta_k)/g_{ii} \quad \text{für } i \in \{n-1, \dots, 1\}. \quad (133.27)$$

Die Berechnung der Inversen $\mathbf{A}^{-1} = (\mathbf{G}')^{-1}\mathbf{G}^{-1}$ folgt aus $\mathbf{G}'\mathbf{A}^{-1} = \mathbf{G}^{-1}$, wobei \mathbf{A}^{-1} symmetrisch und \mathbf{G}^{-1} eine untere Dreiecksmatrix ist, deren Diagonalelemente aus den inversen Diagonalelementen von \mathbf{G} sich ergeben. Für die Diagonalelemente von \mathbf{G}^{-1} gilt dann mit $\mathbf{A}^{-1} = (\bar{a}_{ij})$

$$\frac{1}{g_{ii}} = \sum_{k=i}^n g_{ik}\bar{a}_{ki} = g_{ii}\bar{a}_{ii} + \sum_{k=i+1}^n g_{ik}\bar{a}_{ki}.$$

Nach \bar{a}_{ii} aufgelöst ergibt sich

$$\bar{a}_{ii} = \frac{1}{g_{ii}} \left(\frac{1}{g_{ii}} - \sum_{k=i+1}^n g_{ik} \bar{a}_{ki} \right). \quad (133.28)$$

Für die Elemente von \mathbf{G}^{-1} oberhalb der Diagonalen erhält man

$$0 = \sum_{k=i}^n g_{ik} \bar{a}_{kj} = g_{ii} \bar{a}_{ij} + \sum_{k=i+1}^n g_{ik} \bar{a}_{kj} \quad \text{für } i < j$$

und nach \bar{a}_{ij} aufgelöst

$$\bar{a}_{ij} = -\frac{1}{g_{ii}} \sum_{k=i+1}^n g_{ik} \bar{a}_{kj} \quad \text{für } i < j. \quad (133.29)$$

Aus der Produktschrittsumme in (133.28) ist ersichtlich, daß zur Berechnung der Diagonalelemente \bar{a}_{ii} von \mathbf{A}^{-1} lediglich die Elemente von \mathbf{A}^{-1} innerhalb des Profils benötigt werden.

e) Inversion großer und schwach besetzter großer Matrizen

Zur Inversion großer Matrizen und zur Lösung großer linearer Gleichungssysteme werden elektronische Datenverarbeitungsanlagen eingesetzt. Rechenprogramme für die verschiedenen Inversions- und Auflösungsmethoden wurden in vielen Varianten aufgestellt und veröffentlicht (z.B. EHLERT 1977; GEORGE und LIU 1981; LAWSON und HANSON 1974; PODER und TSCHERNING 1973; RUTISHAUSER 1976; SCHWARZ et al. 1972; SPÄTH 1974). Die Rechenverfahren beschränken sich nicht nur auf direkte Inversions- und Lösungsmethoden, auch iterative Verfahren werden angewendet (FADDEEV und FADDEEVA 1963; HOUSEHOLDER 1964; SCHWARZ et al. 1972).

Matrizen großer Dimensionen besitzen häufig die Eigenschaft, daß nur ein kleiner Prozentsatz von Elementen von Null verschieden ist. Man bezeichnet sie als *schwach besetzte* oder *Sparse*-Matrizen. Besondere Techniken wurden für die Inversion und Lösung entwickelt, um Vorteil aus den vielen Nullelementen zu ziehen (z.B. ACKERMANN et al. 1970; BARKER 1977; GEORGE und LIU 1981; GREPEL 1987; JENNINGS 1977, S.145; SCHEK et al. 1977; SCHENDEL 1977; SCHWARZ 1978; SNAY 1976; TEWARSON 1973). Durch Umordnen kann man zum Beispiel das Profil schwach besetzter Matrizen reduzieren, so daß sich die von Null verschiedenen Elemente in der Nähe der Diagonalen konzentrieren, was die Inversion oder die Lösung stark vereinfacht, wie aus (133.24) bis (133.29) ersichtlich. Besitzt die Koeffizientenmatrix eines großen linearen Gleichungssystems *Block-Diagonal-Struktur*, bei der die Blöcke nur wenige gemeinsame Parameter besitzen, kann eine wirkungsvolle Bearbeitung

des Gleichungssystems in der Zerlegung in Teilsysteme und in der getrennten Elimination der nicht gemeinsamen Parameter bestehen (WOLF 1968, S.75; WOLF 1978B). Die reduzierten Systeme für die gemeinsamen Parameter sind dann abschließend zu addieren und zu lösen.

Schließlich erhebt sich bei der Inversion großer Matrizen oder der Lösung großer linearer Gleichungssysteme die Frage nach der Genauigkeit der numerischen Rechnung. Abschätzungen hierüber erlauben die Kondition einer Matrix (LAWSON und HANSON 1974, S.49; SCHWARZ et al. 1972, S.21; WERNER 1975, S.155; WROBEL 1974) oder stochastische Rundungsfehler-Modelle (MEISSL 1980).

134 Matrizenidentitäten

Die Inverse einer regulären quadratischen Blockmatrix M von der Form (131.18) soll nun abgeleitet und daraus einige Matrizenidentitäten entwickelt werden. Es sei

$$M = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix}, \quad (134.1)$$

worin \mathbf{A} und \mathbf{D} reguläre quadratische Untermatrizen seien. Mit

$$MM^{-1} = \begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} \begin{vmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{vmatrix} = \begin{vmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{vmatrix} \quad (134.2)$$

folgen mit (131.21) die Bestimmungsgleichungen für die Untermatrizen der Inversen M^{-1}

1. $\mathbf{AE} + \mathbf{BG} = \mathbf{I}$
2. $\mathbf{AF} + \mathbf{BH} = \mathbf{0}$
3. $\mathbf{CE} + \mathbf{DG} = \mathbf{0}$
4. $\mathbf{CF} + \mathbf{DH} = \mathbf{I}$.

Aus 1. ergibt sich $\mathbf{E} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{BG}$ und damit aus 3. $\mathbf{CA}^{-1} - \mathbf{CA}^{-1}\mathbf{BG} + \mathbf{DG} = \mathbf{0}$. Somit ist $\mathbf{G} = -(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1}$, denn die Matrix $\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}$ ist regulär, da sie durch elementare Umformungen von M entsteht, wie in (134.8) gezeigt wird. Damit folgt $\mathbf{E} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1}$. Aus 2. erhält man $\mathbf{F} = -\mathbf{A}^{-1}\mathbf{BH}$ und damit aus 4. $-\mathbf{CA}^{-1}\mathbf{BH} + \mathbf{DH} = \mathbf{I}$ oder $\mathbf{H} = (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}$. Folglich

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix}^{-1} = \begin{vmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} & (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \end{vmatrix}.$$

$$(134.3)$$

Eine weitere Möglichkeit die Gleichungen 1. und 3. aufzulösen besteht darin, aus 3. die Matrix $\mathbf{G} = -\mathbf{D}^{-1}\mathbf{C}\mathbf{E}$ zu ermitteln, so daß mit 1. $\mathbf{A}\mathbf{E} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\mathbf{E} = \mathbf{I}$ und $\mathbf{E} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}$ folgt. Somit ergibt sich $\mathbf{G} = -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}$. Aus dem Vergleich mit der ersten Spalte auf der rechten Seite von (134.3) folgen dann die beiden Identitäten

$$(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \quad (134.4)$$

und

$$\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (134.5)$$

Ersetzt man in (134.4) \mathbf{A} durch \mathbf{A}^{-1} und \mathbf{B} durch $-\mathbf{B}$ sowie in (134.5) \mathbf{D}^{-1} durch \mathbf{D} und \mathbf{B} durch $-\mathbf{B}$, erhält man

$$(\mathbf{A}^{-1} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{A} - \mathbf{A}\mathbf{B}(\mathbf{D} + \mathbf{C}\mathbf{A}\mathbf{B})^{-1}\mathbf{C}\mathbf{A} \quad (134.6)$$

und

$$\mathbf{D}\mathbf{C}(\mathbf{A} + \mathbf{B}\mathbf{D}\mathbf{C})^{-1} = (\mathbf{D}^{-1} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (134.7)$$

Die Gaußsche Elimination (133.2) zur Erzeugung einer Dreiecksmatrix läßt sich auch für die Blockmatrix \mathbf{M} in (134.1) durchführen. Man erhält

$$\left| \begin{array}{cc|cc} \mathbf{I} & \mathbf{0} & \mathbf{A} & \mathbf{B} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I} & \mathbf{C} & \mathbf{D} \end{array} \right| = \left| \begin{array}{cc|cc} \mathbf{A} & \mathbf{B} & \mathbf{0} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{array} \right|. \quad (134.8)$$

Hieraus ist ersichtlich, daß die Berechnung einer inversen Matrix nach (133.11) und die Lösung eines linearen Gleichungssystems nach (133.15) auch ohne Rückrechnung nach (133.7) und (133.18) ausschließlich durch eine Elimination nach (133.2) erfolgen kann, falls die zu invertierende Matrix \mathbf{A} durch Einheits- und Nullmatrizen sowie durch den Absolutgliedvektor \mathbf{l} in (133.14) erweitert wird, denn faßt man die einzelnen Eliminationsschritte entsprechend (134.8) in einer Blockmatrix zusammen, ergibt sich

$$\left| \begin{array}{cc|ccc} \mathbf{I} & \mathbf{0} & \mathbf{A} & \mathbf{I} & \mathbf{l} \\ -\mathbf{A}^{-1} & \mathbf{I} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{array} \right| = \left| \begin{array}{ccc|ccc} \mathbf{A} & \mathbf{I} & \mathbf{l} & \mathbf{0} & -\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{l} \end{array} \right|. \quad (134.9)$$

135 Spaltenraum und Nullraum einer Matrix

Die Spalten einer Matrix lassen sich als Vektoren auffassen, die einen Vektorraum aufspannen.

Definition: Der *Spalten-* oder *Rangraum* $R(\mathbf{A})$ einer $m \times n$ Matrix \mathbf{A} wird durch die Menge der Vektoren $\mathbf{y} = \mathbf{A}\mathbf{x}$ mit $\mathbf{x} \in E^n$ definiert, $R(\mathbf{A}) = \{\mathbf{y} | \mathbf{y} = \mathbf{A}\mathbf{x}, \mathbf{x} \in E^n\}$. (135.1)

Stellt man \mathbf{A} durch seine n Spaltenvektoren \mathbf{a}_i mit $\mathbf{A} = |\mathbf{a}_1, \dots, \mathbf{a}_n|$ dar, dann ist $\mathbf{y} = \sum_{i=1}^n x_i \mathbf{a}_i$ mit $x_i \in \mathbb{R}$, und es ist offensichtlich, daß die Vektoren \mathbf{a}_i den Spaltenraum $R(\mathbf{A})$ aufspannen. Für ihn gilt der

Satz: Der Spaltenraum $R(\mathbf{A})$ einer $m \times n$ Matrix \mathbf{A} ist Unterraum des E^m .
(135.2)

Beweis: Mit $\mathbf{y} = \mathbf{A}\mathbf{x}$ ist $\mathbf{y} \in E^m$, so daß mit (121.9) die Aussage folgt.

Ein weiterer Vektorraum einer Matrix ist gegeben durch die

Definition: Der Nullraum $N(\mathbf{A})$ einer $m \times n$ Matrix \mathbf{A} wird durch die Menge der Vektoren \mathbf{x} definiert, für die $\mathbf{A}\mathbf{x} = \mathbf{0}$ mit $\mathbf{x} \in E^n$ gilt, $N(\mathbf{A}) = \{\mathbf{x} | \mathbf{A}\mathbf{x} = \mathbf{0}, \mathbf{x} \in E^n\}$.
(135.3)

Die Beziehung zwischen Null- und Spaltenraum einer Matrix beschreibt der

Satz: Für eine $m \times n$ Matrix \mathbf{A} sind der Nullraum von \mathbf{A}' und das orthogonale Komplement des Spaltenraums von \mathbf{A} identisch, also $N(\mathbf{A}') = R(\mathbf{A})^\perp$, und entsprechend $N(\mathbf{A}) = R(\mathbf{A}')^\perp$.
(135.4)

Beweis: Stellt man \mathbf{A} durch seine n Spalten \mathbf{a}_i mit $\mathbf{A} = |\mathbf{a}_1, \dots, \mathbf{a}_n|$ dar, ist $\mathbf{x} \in N(\mathbf{A}')$, falls $\mathbf{a}'_i \mathbf{x} = 0$ für $i \in \{1, \dots, n\}$ gilt. Der Vektor \mathbf{x} ist dann orthogonal zu jeder Spalte von \mathbf{A} und daher orthogonal zum Spaltenraum $R(\mathbf{A})$. Nach (124.7) ist folglich \mathbf{x} Element des orthogonalen Komplements von $R(\mathbf{A})$, also $\mathbf{x} \in R(\mathbf{A})^\perp$, und es gilt $N(\mathbf{A}') = R(\mathbf{A})^\perp$. Durch entsprechende Überlegungen folgt die zweite Aussage.

Die Dimension des Spalten- und Nullraums einer Matrix erhält man mit dem

Satz: Es sei \mathbf{A} eine $m \times n$ Matrix mit $\text{rg}\mathbf{A} = r$. Dann gilt $\text{rg}\mathbf{A} = \dim R(\mathbf{A}) = \text{rg}\mathbf{A}' = \dim R(\mathbf{A}') = r$ und $\dim N(\mathbf{A}) = n - r$ sowie $\dim N(\mathbf{A}') = m - r$.
(135.5)

Beweis: Eine Basis für den Spaltenraum $R(\mathbf{A})$ beziehungsweise $R(\mathbf{A}')$ bilden die r linear unabhängigen Spalten von \mathbf{A} beziehungsweise von \mathbf{A}' , so daß mit (122.9) und (132.2) die erste Aussage folgt. Mit (135.4) erhält man $\dim N(\mathbf{A}) = \dim R(\mathbf{A}')^\perp$. Da $R(\mathbf{A}')$ nach (135.2) Unterraum des E^n ist, gilt mit (124.7) $\dim R(\mathbf{A}') + \dim R(\mathbf{A}')^\perp = n$ und daher $\dim R(\mathbf{A}') + \dim N(\mathbf{A}) = n$ und entsprechend $\dim R(\mathbf{A}) + \dim N(\mathbf{A}') = m$, woraus mit $\dim R(\mathbf{A}) = \dim R(\mathbf{A}') = r$ die zweite Aussage folgt.

Der folgende Satz ist für die später zu behandelnden Normalgleichungen von Interesse.

Satz: Es sei \mathbf{A} eine $m \times n$ Matrix mit $\operatorname{rg} \mathbf{A} = r$. Dann gilt $R(\mathbf{A}') = R(\mathbf{A}'\mathbf{A})$ und daher $\operatorname{rg}(\mathbf{A}'\mathbf{A}) = r$ und entsprechend $R(\mathbf{A}) = R(\mathbf{A}\mathbf{A}')$ sowie $\operatorname{rg}(\mathbf{A}\mathbf{A}') = r$. (135.6)

Beweis: Aus $\mathbf{A}\mathbf{x} = \mathbf{0}$ folgt $\mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{0}$. Aus $\mathbf{A}'\mathbf{A}\mathbf{x} = \mathbf{0}$ andererseits folgt $\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x} = 0$ und daraus mit $\mathbf{y} = \mathbf{A}\mathbf{x}$ weiter $\mathbf{y}'\mathbf{y} = 0$ und daher $\mathbf{y} = \mathbf{0}$ sowie $\mathbf{A}\mathbf{x} = \mathbf{0}$. Die beiden Nullräume von \mathbf{A} und $\mathbf{A}'\mathbf{A}$, die beide Unterräume des E^n sind, sind also identisch, $N(\mathbf{A}) = N(\mathbf{A}'\mathbf{A})$. Mit (135.4) folgt dann $R(\mathbf{A}')^\perp = R(\mathbf{A}'\mathbf{A})^\perp$ und damit aus (124.7) $R(\mathbf{A}') = R(\mathbf{A}'\mathbf{A})$. Dann ist $\dim R(\mathbf{A}') = \dim R(\mathbf{A}'\mathbf{A})$, so daß mit (135.5) $r = \operatorname{rg}(\mathbf{A}'\mathbf{A})$ sich ergibt. Die beiden restlichen Aussagen erhält man durch entsprechende Überlegungen.

136 Determinanten

Skalare Größen, die für quadratische Matrizen definiert sind und die sich für die Charakterisierung dieser Matrizen eignen, sind die Determinanten. Zu ihrer Definition benötigt man den Begriff der Permutation, der zusammen mit dem Begriff der Kombination auch im Abschnitt 2 verwendet wird, so daß zunächst Permutationen und Kombinationen behandelt werden.

a) Permutation und Kombination

Als eine *Permutation* von Elementen einer Menge bezeichnet man jede Zusammenstellung, die dadurch entsteht, daß die Elemente in irgendeiner Reihenfolge nebeneinander gesetzt werden. Will man die Anzahl der Permutationen beispielsweise der drei Buchstaben a, b, c ermitteln, so kann jeder der drei Buchstaben an die erste Position gerückt werden. Jeder der zwei verbleibenden Buchstaben läßt sich an die zweite Stelle setzen, während die dritte Position von dem unbenutzten Buchstaben eingenommen wird, somit

$$\begin{array}{l} a b c \quad b a c \quad c a b \\ a c b \quad b c a \quad c b a . \end{array}$$

Das Besetzen der ersten Position kann auf drei Arten geschehen, das der zweiten auf zwei und das der dritten auf eine, so daß die Anzahl der Permutationen $3 \cdot 2 \cdot 1 = 6$ ergibt. Allgemein gilt daher der

Satz: Die Anzahl der Permutationen n verschiedener Elemente ist gleich $1 \cdot 2 \cdot 3 \cdot \dots \cdot n = n!$. (136.1)

Es soll jetzt die Anzahl der Permutationen bestimmt werden, die man mit n Elementen erhält, wenn nur k Elemente in den Permutationen benutzt werden. Man bezeichnet dies als *Kombination* k -ter Ordnung. Stellt man die gleichen Überlegungen wie für (136.1) an, so kann die erste Position auf n Arten besetzt werden, die zweite auf $n - 1$ Arten und die k -te Position auf

$n - (k - 1)$ Arten, insgesamt also auf $n(n - 1) \dots (n - k + 1)$ Arten. Nimmt man auf die Anordnung der Elemente in den Kombinationen keine Rücksicht, so sind die Kombinationen identisch, die die gleichen Elemente in verschiedenen Anordnungen enthalten, z.B. $a b c$ und $c a b$. Falls r die Anzahl der Kombinationen ohne Berücksichtigung der Anordnung ist, so ist $rk!$ die Anzahl mit Berücksichtigung der Anordnung, da k Elemente sich nach (136.1) $k!$ mal permutieren lassen. Es gilt daher mit $0! = 1$ der

Satz: Für n verschiedene Elemente beträgt die Anzahl der Kombinationen k -ter Ordnung ohne Berücksichtigung der Anordnung

$$\frac{n(n-1) \dots (n-k+1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}. \quad (136.2)$$

b) Definition der Determinante

Definition: Die *Determinante* $\det \mathbf{A}$ einer $n \times n$ Matrix $\mathbf{A} = (a_{ij})$ ist gegeben durch

$$\det \mathbf{A} = \sum \pm a_{1\alpha} a_{2\beta} \dots a_{n\epsilon},$$

wobei über die $n!$ Permutationen $\alpha, \beta, \dots, \epsilon$ der natürlichen Zahlen $1, 2, \dots, n$ zu summieren ist. Das positive Vorzeichen gilt, falls die Anzahl der Zahlenpaare in der Permutation, die nicht in der natürlichen Ordnung sich befinden, gerade ist, und das negative Vorzeichen für eine ungerade Anzahl. (136.3)

Beispiel: Aufgrund von (136.3) gilt

$$\det \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

c) Sätze für Determinanten

Für Determinanten gelten die beiden im folgenden benötigten Sätze (KOWALSKY 1977, S.87 und 94; NEISS und LIERMANN 1975, S.107 und 111), die nicht bewiesen werden.

Satz: Werden zwei Spalten oder zwei Zeilen von \mathbf{A} vertauscht, ändert $\det \mathbf{A}$ das Vorzeichen. (136.4)

Satz (Laplacescher Entwicklungssatz): Bezeichnet man die Determinante der Untermatrix von \mathbf{A} , die durch das Streichen der i -ten Zeile und j -ten Spalte entsteht, mit $\det \mathbf{A}_{ij}$, dann gilt

$$\det \mathbf{A} = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det \mathbf{A}_{ij} \quad \text{für } j \in \{1, \dots, n\} \quad (136.5)$$

und

$$\det \mathbf{A} = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det \mathbf{A}_{ij} \quad \text{für } i \in \{1, \dots, n\}. \quad (136.6)$$

Mit Hilfe dieser Sätze lassen sich weitere ableiten.

Satz: Besitzt \mathbf{A} zwei identische Zeilen oder Spalten, ist $\det \mathbf{A} = 0$. (136.7)

Beweis: Es gelte $\det \mathbf{A}$ für \mathbf{A} mit zwei identischen Zeilen oder Spalten. Mit (136.4) folgt durch Vertauschen dieser Zeilen oder Spalten $\det \mathbf{A} = -\det \mathbf{A}$, was aber nur für $\det \mathbf{A} = 0$ erfüllt sein kann.

Aus (136.5) und (136.6) ergibt sich

$$\det \mathbf{A} = \det \mathbf{A}' \quad (136.8)$$

und mit $\mathbf{D} = \text{diag}(d_{11}, d_{22}, \dots, d_{nn})$

$$\det \mathbf{D} = d_{11} d_{22} \dots d_{nn}, \quad (136.9)$$

denn entwickelt man $\det \mathbf{D}$ mit (136.5), erhält man $\det \mathbf{D} = d_{11} \det \mathbf{D}_{11} = d_{11} d_{22} \det(\mathbf{D}_{11})_{22}$ und so fort. Aus den gleichen Überlegungen folgt

$$\det \mathbf{I} = 1 \quad \text{und} \quad \det \mathbf{G} = g_{11} g_{22} \dots g_{nn}, \quad (136.10)$$

falls $\mathbf{G} = (g_{ij})$ eine $n \times n$ obere oder untere Dreiecksmatrix bedeutet.

Mit (136.6) läßt sich zeigen, daß elementare Zeilenumformungen mit Hilfe der Matrizen vom Typ \mathbf{E}_3 in (132.5) den Wert einer Determinante nicht ändern. Addiert man nämlich die mit dem Skalar c multiplizierten Elemente der Zeile k zur Zeile i , erhält man

$$\begin{aligned} \det \mathbf{A} &= \sum_{j=1}^n (-1)^{i+j} (a_{ij} + ca_{kj}) \det \mathbf{A}_{ij} \\ &= \sum_{j=1}^n (-1)^{i+j} a_{ij} \det \mathbf{A}_{ij} + c \sum_{j=1}^n (-1)^{i+j} a_{kj} \det \mathbf{A}_{ij}. \end{aligned}$$

Der zweite Summand auf der rechten Seite verschwindet aber wegen (136.7), da er die Anwendung des Entwicklungssatzes auf eine Matrix mit zwei identischen Zeilen bedeutet. Gleiches gilt auch für die entsprechenden Spaltenumformungen.

Praktisch läßt sich daher $\det \mathbf{A}$ mit (133.2) und

$$\det \mathbf{A} = \det(\mathbf{C}^{-1} \mathbf{A}) = \det \mathbf{B} = b_{11} b_{22} \dots b_{nn} \quad (136.11)$$

aus den Diagonalelementen der bei der Gaußschen Elimination entstehenden oberen Dreiecksmatrix \mathbf{B} berechnen. Weiter gilt für die Determinante einer Blockmatrix

$$\det \begin{vmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{vmatrix} = \det(\mathbf{B} - \mathbf{C}\mathbf{E}^{-1}\mathbf{D}) \det \mathbf{E} = \det \mathbf{B} \det(\mathbf{E} - \mathbf{D}\mathbf{B}^{-1}\mathbf{C}), \quad (136.12)$$

sofern \mathbf{E} beziehungsweise \mathbf{B} regulär sind. Die Determinante der Blockmatrix ändert nämlich durch die folgende Reduktion auf eine Diagonalität in Blöcken mit elementaren Umformungen mittels der Matrizen vom Typ \mathbf{E}_3 ihren Wert nicht

$$\begin{vmatrix} \mathbf{I} & -\mathbf{C}\mathbf{E}^{-1} \\ \mathbf{0} & \mathbf{I} \end{vmatrix} \begin{vmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{vmatrix} = \begin{vmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{E}^{-1}\mathbf{D} & \mathbf{I} \end{vmatrix} = \begin{vmatrix} \mathbf{B} - \mathbf{C}\mathbf{E}^{-1}\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{E} \end{vmatrix}.$$

Eine weitere Reduktion der rechten Seite auf Dreiecksmatrizen wie in (136.11) ergibt dann den ersten Ausdruck in (136.12). Der zweite folgt durch entsprechende Umformungen.

Die Determinante des Produktes zweier Matrizen erhält man mit dem

Satz: Sind \mathbf{A} und \mathbf{B} zwei quadratische $n \times n$ Matrizen, dann ist $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$. (136.13)

Beweis: In der Identität

$$\begin{vmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{vmatrix} \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = \begin{vmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix}$$

beinhaltet die erste Matrix auf der linken Seite elementare Umformungen mit Hilfe der Matrizen vom Typ \mathbf{E}_3 . Für die Determinanten beider Seiten erhält man daher, falls noch eine n -fache Vertauschung der Spalten der Matrix der rechten Seite vorgenommen wird, da \mathbf{I} eine $n \times n$ Einheitsmatrix ist

$$\det \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = \det \begin{vmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = (-1)^n \det \begin{vmatrix} \mathbf{AB} & \mathbf{0} \\ \mathbf{B} & -\mathbf{I} \end{vmatrix}.$$

Ist \mathbf{B} regulär, läßt sich (136.12) anwenden, und es ergibt sich $\det \mathbf{A} \det \mathbf{B} = (-1)^n \det(-\mathbf{I}) \det \mathbf{AB}$, woraus $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$ folgt. Sind \mathbf{A} und \mathbf{B} singulär, ist \mathbf{AB} wegen (132.3) ebenfalls singulär, so daß $\det \mathbf{A} = \det \mathbf{B} = \det \mathbf{AB} = 0$ gilt, wie in (136.17) gezeigt wird, und somit $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$ auch für diesen Fall gilt.

Aus $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ erhält man mit (136.10) und (136.13)

$$\det \mathbf{A}^{-1} = (\det \mathbf{A})^{-1}. \quad (136.14)$$

Man bezeichnet

$$a_{ij}^* = (-1)^{i+j} \det \mathbf{A}_{ij} \quad (136.15)$$

als die zum Element a_{ij} der Matrix \mathbf{A} gehörende *Adjunkte* und die Matrix $\bar{\mathbf{A}} = (a_{ji}^*)$ die zu \mathbf{A} *adjungierte* Matrix. Wegen (136.5) bis (136.7) gilt

$$\begin{aligned} \mathbf{A}\bar{\mathbf{A}} &= \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \begin{vmatrix} a_{11}^* & a_{21}^* & \dots & a_{n1}^* \\ a_{12}^* & a_{22}^* & \dots & a_{n2}^* \\ \dots & \dots & \dots & \dots \\ a_{1n}^* & a_{2n}^* & \dots & a_{nn}^* \end{vmatrix} \\ &= \begin{vmatrix} \det \mathbf{A} & 0 & \dots & 0 \\ 0 & \det \mathbf{A} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \det \mathbf{A} \end{vmatrix} = \bar{\mathbf{A}}\mathbf{A} \end{aligned}$$

oder

$$\mathbf{A}\bar{\mathbf{A}} = \bar{\mathbf{A}}\mathbf{A} = (\det \mathbf{A})\mathbf{I}.$$

Falls $\det \mathbf{A} \neq 0$, ergibt sich daher mit (131.12)

$$\mathbf{A}^{-1} = \frac{\bar{\mathbf{A}}}{\det \mathbf{A}}. \quad (136.16)$$

Damit folgt der

Satz: Eine quadratische $n \times n$ Matrix \mathbf{A} ist genau dann regulär, wenn $\det \mathbf{A} \neq 0$ ist. (136.17)

Beweis: Ist $\det \mathbf{A} \neq 0$ folgt aus (136.16), daß \mathbf{A} regulär ist. Ist andererseits \mathbf{A} regulär, ergibt sich aus $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ mit (136.13) und (136.14) $\det \mathbf{A}(\det \mathbf{A})^{-1} = \det \mathbf{I} = 1$, so daß $\det \mathbf{A} \neq 0$ sein muß und damit die Aussage folgt.

Eine notwendige und hinreichende Bedingung dafür, daß die $n \times n$ Matrix \mathbf{A} regulär ist, war in (133.1) mit Hilfe des Ranges der Matrix \mathbf{A} formuliert worden. Aus (133.1) und (136.17) folgt daher der

Satz: Es gilt $\text{rg} \mathbf{A} = n$ genau dann, wenn $\det \mathbf{A} \neq 0$ ist. (136.18)

Damit ist die Beziehung zwischen dem Rang einer quadratischen Matrix und ihrer Determinante hergestellt worden.

137 Spur einer Matrix und Darstellung einer Matrix als Vektor

Mit Hilfe der Spur einer Matrix werden häufig Gütekriterien für die Parameterschätzung angegeben.

Definition: Die *Spur* $\text{sp}\mathbf{A}$ einer quadratischen $n \times n$ Matrix \mathbf{A} mit $\mathbf{A} = (a_{ij})$ ist durch $\text{sp}\mathbf{A} = \sum_{i=1}^n a_{ii}$ gegeben. (137.1)

Satz: Es seien \mathbf{A} und \mathbf{B} zwei $n \times n$ Matrizen, dann gilt

$$\text{sp}(\mathbf{A} + \mathbf{B}) = \text{sp}\mathbf{A} + \text{sp}\mathbf{B}. \quad (137.2)$$

Weiter sei \mathbf{A} eine $n \times r$ und \mathbf{B} eine $r \times n$ Matrix, dann gilt

$$\text{sp}(\mathbf{AB}) = \text{sp}(\mathbf{BA}). \quad (137.3)$$

Beweis: (137.2) folgt unmittelbar aus der Definition. Mit $\mathbf{B} = (b_{ij})$ erhält man weiter $\text{sp}(\mathbf{AB}) = \sum_{i=1}^n \sum_{j=1}^r a_{ij}b_{ji}$ und $\text{sp}(\mathbf{BA}) = \sum_{k=1}^r \sum_{l=1}^n b_{kl}a_{lk} = \sum_{l=1}^n \sum_{k=1}^r a_{lk}b_{kl}$ und damit (137.3).

Ist die Matrix \mathbf{A} ein Skalar, also $\mathbf{A} = a$, dann ist

$$\text{sp}\mathbf{A} = a. \quad (137.4)$$

Eine Matrix läßt sich auch als Vektor darstellen.

Definition: Es sei $\mathbf{A} = (a_{ij})$ eine $m \times n$ Matrix, dann bezeichnet $\text{vec}\mathbf{A}$ den $mn \times 1$ Vektor, der durch das Untereinanderschreiben der Spalten von \mathbf{A} entsteht, folglich

$$\text{vec}\mathbf{A} = |a_{11}, \dots, a_{m1}, a_{12}, \dots, a_{mn}|'. \quad (137.5)$$

Zur Umwandlung von Matrixgleichungen der Form $\mathbf{ABC} = \mathbf{L}$, in der die Spalten von \mathbf{B} die unbekannt Parametervektoren enthalten, in ein übliches lineares Gleichungssystem dient der

Satz: Es sei \mathbf{A} eine $m \times n$, \mathbf{B} eine $n \times p$ und \mathbf{C} eine $p \times s$ Matrix, dann gilt

$$\text{vec}\mathbf{ABC} = (\mathbf{C}' \otimes \mathbf{A})\text{vec}\mathbf{B}. \quad (137.6)$$

Beweis: Mit $\mathbf{B} = (\mathbf{b}_i)$ und $\mathbf{C} = (c_{ij})$ erhält man

$$\begin{aligned} \text{vec}\mathbf{ABC} &= \begin{vmatrix} \sum_{i=1}^p c_{i1}\mathbf{A}\mathbf{b}_i \\ \dots\dots\dots \\ \sum_{i=1}^p c_{is}\mathbf{A}\mathbf{b}_i \end{vmatrix} = \begin{vmatrix} c_{11}\mathbf{A} & \dots & c_{p1}\mathbf{A} \\ \dots\dots\dots \\ c_{1s}\mathbf{A} & \dots & c_{ps}\mathbf{A} \end{vmatrix} \begin{vmatrix} \mathbf{b}_1 \\ \dots \\ \mathbf{b}_p \end{vmatrix} \\ &= (\mathbf{C}' \otimes \mathbf{A})\text{vec}\mathbf{B}. \end{aligned}$$

Es folgen noch zwei Sätze mit ähnlichen Aussagen.

Satz: Es seien \mathbf{A} und \mathbf{B} zwei $n \times n$ Matrizen, dann gilt

$$(\text{vec}\mathbf{A})'\text{vec}\mathbf{B} = (\text{vec}\mathbf{B})'\text{vec}\mathbf{A} = \text{sp}(\mathbf{A}\mathbf{B}') . \quad (137.7)$$

Beweis: Mit (123.2) folgt die erste Aussage. Weiter gilt mit $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$ und $\mathbf{B}' = (b'_{ij})$ sowie (131.6)

$$(\text{vec}\mathbf{A})'\text{vec}\mathbf{B} = \sum_{i=1}^n \sum_{j=1}^n a_{ij}b_{ij} = \sum_{i=1}^n \sum_{j=1}^n a_{ij}b'_{ji} = \text{sp}(\mathbf{A}\mathbf{B}')$$

wegen (137.1), so daß die zweite Aussage folgt.

Satz: Es sei \mathbf{A} eine $m \times n$, \mathbf{B} eine $n \times n$ und \mathbf{C} eine $m \times m$ Matrix, dann gilt

$$(\text{vec}\mathbf{A})'(\mathbf{B} \otimes \mathbf{C})\text{vec}\mathbf{A} = \text{sp}(\mathbf{A}\mathbf{B}\mathbf{A}'\mathbf{C}') = \text{sp}(\mathbf{A}\mathbf{B}'\mathbf{A}'\mathbf{C}) . \quad (137.8)$$

Beweis: Mit (137.6) und (137.7) erhält man

$$(\text{vec}\mathbf{A})'(\mathbf{B} \otimes \mathbf{C})\text{vec}\mathbf{A} = (\text{vec}\mathbf{A})'\text{vec}\mathbf{C}\mathbf{A}\mathbf{B}' = \text{sp}(\mathbf{A}\mathbf{B}\mathbf{A}'\mathbf{C}') .$$

Da die Spur einer transponierten Matrix der Spur der ursprünglichen Matrix gleicht, folgt mit (137.3) die zweite Aussage.

14 Quadratische Formen

141 Transformationen

a) Affine Transformationen

Die lineare Transformation

$$\mathbf{y} = \mathbf{B}\mathbf{x} , \quad (141.1)$$

die den Vektor \mathbf{x} mit Hilfe der Matrix \mathbf{B} in den Vektor \mathbf{y} transformiert, bezeichnet man als *affine Transformation*. Tritt noch mit $\mathbf{z} = \mathbf{B}\mathbf{x} + \mathbf{c}$ der *Translationsvektor* \mathbf{c} hinzu, läßt er sich durch $\mathbf{y} = \mathbf{z} - \mathbf{c} = \mathbf{B}\mathbf{x}$ eliminieren, so daß es genügt, die Transformation (141.1) zu behandeln. Ist \mathbf{B} eine $m \times n$ Matrix, kann (141.1) als Abbildung des Vektors $\mathbf{x} \in E^n$ in den Vektor $\mathbf{y} \in E^m$ angesehen werden.

Ist \mathbf{B} eine reguläre $n \times n$ Matrix, läßt sich (141.1) als Transformation eines Vektors $\mathbf{x} \in E^n$ bezüglich einer Basis des E^n in einen anderen Vektor $\mathbf{y} \in E^n$ bezüglich derselben Basis interpretieren. Man kann mit (141.1) aber auch die Vorstellung der Koordinatentransformation verbinden, indem ein und derselbe Vektor von der Darstellung bezüglich einer Basis oder eines Koordinatensystems, dessen Achsen in Richtung der Basisvektoren zeigen, in die Darstellung

bezüglich einer anderen Basis oder eines anderen Koordinatensystems übergeht. Bei der ersten Interpretation wird der Vektor transformiert, und die Basis bleibt fest, bei der zweiten wird die Basis oder das Koordinatensystem transformiert, und der Vektor bleibt unverändert.

Die affine Transformation überführt als Vektortransformation Strecken in Strecken, ändert aber die Längen der Strecken, denn im allgemeinen gilt $|\mathbf{y}|^2 = \mathbf{y}'\mathbf{y} = \mathbf{x}'\mathbf{B}'\mathbf{B}\mathbf{x} \neq |\mathbf{x}|^2$. Folglich werden auch Winkel geändert, da sie sich durch die Seiten eines Dreiecks ausdrücken lassen. Eine affine Koordinatentransformation rotiert also die einzelnen Koordinatenachsen um beliebige Winkel und nimmt Längenänderungen vor.

b) Orthogonale Transformationen

Gilt für eine $n \times n$ Matrix \mathbf{B} in (141.1) $\mathbf{B}'\mathbf{B} = \mathbf{I}$, bezeichnet man \mathbf{B} als *orthogonale* Matrix und die Transformation mit einer orthogonalen Matrix als *orthogonale Transformation*. Es gilt der

Satz: Es sei \mathbf{C} eine orthogonale $n \times n$ Matrix, das heißt $\mathbf{C}'\mathbf{C} = \mathbf{I}$. Dann ist \mathbf{C} regulär und daher $\mathbf{C}^{-1} = \mathbf{C}'$. (141.2)

Beweis: Aus $\mathbf{C}'\mathbf{C} = \mathbf{I}$ folgt, daß die Spalten von \mathbf{C} paarweise zueinander orthogonal sind, so daß nach (124.2) und (132.2) $\text{rg}\mathbf{C} = n$ und damit nach (133.1) die Aussage folgt.

Orthogonale Transformationen ändern Streckenlängen und damit auch Winkel nicht, denn bezeichnet man mit $\mathbf{y} = \mathbf{C}\mathbf{x}$ und $\mathbf{z} = \mathbf{C}\mathbf{u}$ die Endpunkte einer transformierten Strecke $\mathbf{x} - \mathbf{u}$, erhält man mit (123.3) und (141.2)

$$|\mathbf{y} - \mathbf{z}|^2 = (\mathbf{C}\mathbf{x} - \mathbf{C}\mathbf{u})'(\mathbf{C}\mathbf{x} - \mathbf{C}\mathbf{u}) = (\mathbf{x} - \mathbf{u})'\mathbf{C}'\mathbf{C}(\mathbf{x} - \mathbf{u}) = |\mathbf{x} - \mathbf{u}|^2.$$

Die Transformation

$$\mathbf{x}^* = \mathbf{A}\mathbf{x} \quad (141.3)$$

sei nun durch die Transformation von n orthonormalen Basisvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$ des E^n in die n orthonormalen Basisvektoren $\mathbf{e}_1^*, \dots, \mathbf{e}_n^*$ des E^n hervorgehoben. Mit $\mathbf{E} = |\mathbf{e}_1, \dots, \mathbf{e}_n|$ enthält dann \mathbf{x} die Komponenten eines Vektors $\mathbf{E}\mathbf{x}$, dargestellt durch die n Basisvektoren $\mathbf{e}_1, \dots, \mathbf{e}_n$, und \mathbf{x}^* die Komponenten eines Vektors $\mathbf{E}^*\mathbf{x}^*$, dargestellt mit $\mathbf{E}^* = |\mathbf{e}_1^*, \dots, \mathbf{e}_n^*|$ durch die n Basisvektoren $\mathbf{e}_1^*, \dots, \mathbf{e}_n^*$, und es gilt $\mathbf{E}\mathbf{x} = \mathbf{E}^*\mathbf{x}^*$, da die Transformation den Vektor unverändert läßt, sowie $\mathbf{E}'\mathbf{E} = \mathbf{I}$ und $\mathbf{E}^{*'}\mathbf{E}^* = \mathbf{I}$, da orthonormale Basisvektoren vorliegen. Mit (141.3) folgt $\mathbf{E}\mathbf{x} = \mathbf{E}^*\mathbf{x}^* = \mathbf{E}^*\mathbf{A}\mathbf{x}$ für alle Vektoren $\mathbf{x} \in E^n$, so daß der Transformation (141.3) die Basistransformation $\mathbf{E} = \mathbf{E}^*\mathbf{A}$ entspricht. Weiter ergibt sich $\mathbf{x}^* = \mathbf{E}^{*'}\mathbf{E}\mathbf{x} = \mathbf{A}\mathbf{x}$ für alle \mathbf{x} , so daß man $\mathbf{A} = \mathbf{E}^{*'}\mathbf{E}$ und mit $\mathbf{A} = (a_{ij})$ sowie (123.4) $a_{ij} = \mathbf{e}_i^{*'}\mathbf{e}_j = \cos(\mathbf{e}_i^*, \mathbf{e}_j)$

erhält, falls $(\mathbf{e}_i^*, \mathbf{e}_j)$ den Winkel zwischen den Basisvektoren \mathbf{e}_i^* und \mathbf{e}_j bezeichnet. Den Kosinus dieses Winkels nennt man *Richtungskosinus*. Ferner ergibt sich $\mathbf{x} = \mathbf{E}'\mathbf{E}^*\mathbf{x}^* = \mathbf{A}'\mathbf{x}^*$ wegen $\mathbf{A}' = \mathbf{E}'\mathbf{E}^*$ und mit (141.3) $\mathbf{x} = \mathbf{A}'\mathbf{A}\mathbf{x}$ für alle \mathbf{x} , so daß $\mathbf{A}'\mathbf{A} = \mathbf{I}$ und \mathbf{A} als orthogonale Matrix folgt. Sind andererseits \mathbf{A} und \mathbf{E} orthogonale Matrizen, erhält man $\mathbf{E}^* = \mathbf{E}\mathbf{A}'$ aus $\mathbf{E} = \mathbf{E}^*\mathbf{A}$, so daß wegen $\mathbf{E}^{*'}\mathbf{E}^* = \mathbf{A}\mathbf{E}'\mathbf{E}\mathbf{A}' = \mathbf{I}$ die Basisvektoren $\mathbf{e}_1^*, \dots, \mathbf{e}_n^*$ orthonormal sind. Es folgt damit der

Satz: Eine Transformation ist genau dann orthogonal, wenn sie eine orthonormale Basis in eine andere orthonormale Basis überführt. Die Transformationsmatrix enthält dann die Richtungskosinus zwischen den ursprünglichen und den transformierten Basisvektoren. (141.4)

Der orthogonalen Transformation orthonormaler Basisvektoren entspricht die orthogonale Transformation orthogonaler Koordinatensysteme.

Beispiel: Orthogonale Transformationen dreidimensionaler, orthogonaler Koordinatensysteme bewirken die *Drehmatrizen*

$$\mathbf{R}_1(\theta) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{vmatrix}, \quad \mathbf{R}_2(\theta) = \begin{vmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{vmatrix},$$

$$\mathbf{R}_3(\theta) = \begin{vmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{vmatrix}, \quad (141.5)$$

die jeweils Koordinatentransformationen in der x_2, x_3 -Ebene, in der x_1, x_3 -Ebene und in der x_1, x_2 -Ebene vornehmen, wobei der Winkel θ zwischen den Achsen des ursprünglichen und des transformierten Systems im Gegenuhrzeigersinn positiv gezählt wird. Mit Hilfe der Drehmatrizen kann durch das Aneinandersetzen von ebenen Transformationen jede beliebige dreidimensionale Transformation ausgeführt werden, da orthogonale Transformationen nacheinander ausgeführt wieder orthogonale Transformationen ergeben, denn sind \mathbf{A} und \mathbf{B} orthogonale Matrizen, gilt $(\mathbf{A}\mathbf{B})'(\mathbf{A}\mathbf{B}) = \mathbf{B}'\mathbf{A}'\mathbf{A}\mathbf{B} = \mathbf{I}$.

Differentielle Drehungen in den Koordinatenebenen um den infinitesimal kleinen Winkel $d\theta$ ergeben sich mit $\cos d\theta = 1$ und $\sin d\theta = d\theta$ aus (141.5) zu

$$\mathbf{R}_1(d\theta) = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & d\theta \\ 0 & -d\theta & 1 \end{vmatrix}, \quad \mathbf{R}_2(d\theta) = \begin{vmatrix} 1 & 0 & -d\theta \\ 0 & 1 & 0 \\ d\theta & 0 & 1 \end{vmatrix},$$

$$\mathbf{R}_3(d\theta) = \begin{vmatrix} 1 & d\theta & 0 \\ -d\theta & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}. \quad (141.6)$$

Drei differentielle Drehungen um die Winkel $d\alpha, d\beta, d\gamma$ aneinandergesetzt betragen, falls $d\alpha d\beta = d\alpha d\gamma = d\beta d\gamma = 0$ gesetzt wird,

$$\mathbf{R}_1(d\alpha)\mathbf{R}_2(d\beta)\mathbf{R}_3(d\gamma) = \begin{vmatrix} 1 & d\gamma & -d\beta \\ -d\gamma & 1 & d\alpha \\ d\beta & -d\alpha & 1 \end{vmatrix}. \quad (141.7)$$

c) Quadratische und bilineare Formen

Ist die Matrix \mathbf{B} der affinen Transformation (141.1) quadratisch und regulär, existiert die inverse Transformation $\mathbf{x} = \mathbf{B}^{-1}\mathbf{y}$, und die quadratische Länge des Vektors \mathbf{x} läßt sich mit (123.3) durch die transformierten Koordinaten ausdrücken

$$\mathbf{x}'\mathbf{x} = \mathbf{y}'(\mathbf{B}^{-1})'\mathbf{B}^{-1}\mathbf{y} = \mathbf{y}'\mathbf{A}\mathbf{y}. \quad (141.8)$$

$\mathbf{y}'\mathbf{A}\mathbf{y}$ heißt *quadratische Form* und \mathbf{A} die Matrix der quadratischen Form. Sie ist symmetrisch, denn $\mathbf{A} = (\mathbf{B}^{-1})'\mathbf{B}^{-1} = \mathbf{A}'$. Drückt man (123.5) entsprechend auch Winkel im transformierten System aus, treten *bilineare Formen* $\mathbf{y}'\mathbf{A}\mathbf{z}$ auf. Die bilineare Form stellt eine Verallgemeinerung des Skalarproduktes (123.1) dar.

142 Eigenwerte und Eigenvektoren

Die Extremwerte der quadratischen Form $\mathbf{x}'\mathbf{A}\mathbf{x}$, in der \mathbf{x} ein $n \times 1$ Vektor und \mathbf{A} eine symmetrische $n \times n$ Matrix bedeuten, soll durch Variation von \mathbf{x} bestimmt werden. Da $\mathbf{x}'\mathbf{A}\mathbf{x}$ bei beliebigem \mathbf{x} beliebig groß oder klein gemacht werden kann, wird als Nebenbedingung

$$\mathbf{x}'\mathbf{x} = 1 \quad (142.1)$$

eingeführt. Das Extremum wird, wie in (171.6) gezeigt wird, mit Hilfe der Lagrangeschen Funktion L bestimmt, $L = \mathbf{x}'\mathbf{A}\mathbf{x} - \lambda(\mathbf{x}'\mathbf{x} - 1)$, in der λ der Lagrangesche Multiplikator bedeutet. Die Differentialquotienten $\partial L/\partial \mathbf{x}$ gleich Null gesetzt ergeben die Werte für das Extremum, die mit \mathbf{x}_i und λ_i bezeichnet seien. Da $\partial L/\partial \mathbf{x} = 2\mathbf{A}\mathbf{x} - 2\lambda\mathbf{x}$ gilt, wie in (172.2) abgeleitet wird, erhält man

$$(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{x}_i = \mathbf{0}. \quad (142.2)$$

Man nennt λ_i *Eigenwerte* und \mathbf{x}_i *Eigenvektoren* der Matrix \mathbf{A} , falls sie (142.2) erfüllen, wobei \mathbf{A} nicht, wie hier vorausgesetzt, symmetrisch zu sein braucht.

Damit Vektoren \mathbf{x}_i existieren, die nicht Nullvektoren sind, müssen nach (122.1) die Spalten der Matrix $\mathbf{A} - \lambda_i\mathbf{I}$ linear abhängig sein. Nach (136.17) gilt dann

$$\det(\mathbf{A} - \lambda_i\mathbf{I}) = 0. \quad (142.3)$$

Die Entwicklung der Determinante nach (136.5) ergibt, geordnet nach Potenzen von λ_i , die *charakteristische Gleichung* für \mathbf{A}

$$\lambda_i^r + K_1 \lambda_i^{r-1} + \dots + K_{r-1} \lambda_i + K_r = 0, \quad (142.4)$$

in der die Koeffizienten K_i Funktionen der Elemente von \mathbf{A} sind. Die Ordnung r der Potenzen ergibt sich mit $r = \text{rg} \mathbf{A}$, denn die Ordnung der größten von Null verschiedenen Unterdeterminante von \mathbf{A} kann nach (136.18) $\text{rg} \mathbf{A}$ nicht überschreiten. Man erhält r von Null verschiedene Lösungen für λ_i aus (142.4), die reell sind, falls \mathbf{A} symmetrisch ist (STIEFEL 1970, S.108).

Die zu den verschiedenen Eigenwerten λ_i gehörenden Eigenvektoren \mathbf{x}_i sind zueinander orthogonal, denn mit (142.2) und der entsprechenden Gleichung $(\mathbf{A} - \lambda_j \mathbf{I}) \mathbf{x}_j = \mathbf{0}$ für λ_j und \mathbf{x}_j folgen $\mathbf{x}_j' \mathbf{A} \mathbf{x}_i = \lambda_i \mathbf{x}_j' \mathbf{x}_i$ und $\mathbf{x}_i' \mathbf{A} \mathbf{x}_j = \lambda_j \mathbf{x}_i' \mathbf{x}_j$. Es gilt aber $\mathbf{x}_j' \mathbf{A} \mathbf{x}_i = \mathbf{x}_i' \mathbf{A} \mathbf{x}_j$ und $\mathbf{x}_j' \mathbf{x}_i = \mathbf{x}_i' \mathbf{x}_j$, so daß folgt $\lambda_i \mathbf{x}_i' \mathbf{x}_j = \lambda_j \mathbf{x}_i' \mathbf{x}_j$ und weiter $(\lambda_i - \lambda_j) \mathbf{x}_i' \mathbf{x}_j = 0$ und somit, falls $\lambda_i \neq \lambda_j$

$$\mathbf{x}_i' \mathbf{x}_j = 0. \quad (142.5)$$

Bei mehrfachen Eigenwerten, beispielsweise $\lambda_i = \lambda_j$, müssen die Eigenvektoren aus dem Lösungsraum, der in (154.6) definiert wird, der homogenen Gleichungen (142.2) derart ausgewählt werden, daß sie zu den übrigen Eigenvektoren jeweils paarweise orthogonal sind.

Nimmt man mit $n = \text{rg} \mathbf{A}$ vollen Rang für \mathbf{A} an und faßt mit $\mathbf{X} = |\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n|$ die Eigenvektoren \mathbf{x}_i zur $n \times n$ Matrix \mathbf{X} zusammen, die mit (142.1) und (142.5) orthogonal ist, also $\mathbf{X}' \mathbf{X} = \mathbf{I}$ und führt mit $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ die $n \times n$ Diagonalmatrix der Eigenwerte λ_i ein, so ergibt sich mit $\mathbf{A} \mathbf{x}_i = \lambda_i \mathbf{x}_i$ aus (142.2) die Beziehung $\mathbf{X}' \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$. Ist $\text{rg} \mathbf{A} = r < n$, sind $n - r$ Eigenwerte gleich Null. Die entsprechenden Eigenvektoren sind dann beliebig festzulegen, sie müssen lediglich (142.1), (142.2) mit $\lambda_i = 0$ und (142.5) erfüllen. Damit gilt der

Satz: Jede symmetrische $n \times n$ Matrix \mathbf{A} mit $r = \text{rg} \mathbf{A}$ läßt sich mit Hilfe einer orthogonalen $n \times n$ Matrix \mathbf{X} der Eigenvektoren von \mathbf{A} derart zerlegen, daß $\mathbf{X}' \mathbf{A} \mathbf{X} = \mathbf{\Lambda}$ gilt, worin $\mathbf{\Lambda}$ eine $n \times n$ Diagonalmatrix bedeutet, deren Diagonalelemente die reellen Eigenwerte von \mathbf{A} enthält, unter denen r Werte von Null verschieden sind. (142.6)

Bei der praktischen Berechnung der Eigenwerte und Eigenvektoren greift man gewöhnlich nicht auf (142.2) und (142.3) zurück, sondern benutzt iterative Methoden (FADDEEV und FADDEEVA 1963; HOUSEHOLDER 1964; RUTISHAUSER 1976, Bd.2; SCHWARZ et al. 1972).

Im Hinblick auf multivariate Hypothesentests ist die Invarianzeigenschaft der

Eigenwerte bedeutsam.

Satz: Die Eigenwerte einer Matrix sind invariant gegenüber orthogonalen Transformationen. (142.7)

Beweis: Es sei C mit $C'C = I$ eine orthogonale Matrix, die den Vektor x transformiere in $y = Cx$, so daß mit $x = C'y$ wegen (141.2) anstelle von $x'Ax$ die quadratische Form $y'CAC'y$ erhalten wird und anstelle von (142.3) mit (136.13) $\det(CAC' - \lambda_i I) = \det(C(A - \lambda_i I)C') = \det(C'C) \det(A - \lambda_i I) = \det(A - \lambda_i I)$, woraus die Aussage folgt.

143 Definite Matrizen

Mit (141.8) war die quadratische Form als Länge eines transformierten Vektors eingeführt worden. Damit sie als Maß einer Länge dienen kann, darf sie nicht negativ werden. Dies führt für die Matrix A der quadratischen Form auf die

Definition: Eine symmetrische $n \times n$ Matrix A bezeichnet man als *positiv definit*, wenn

$$x'Ax > 0 \quad \text{für alle } x \neq 0$$

und als *positiv semidefinit*, wenn

$$x'Ax \geq 0 \quad \text{für alle } x \neq 0. \quad (143.1)$$

Die folgenden Sätze geben Kriterien dafür an, daß eine Matrix positiv definit oder positiv semidefinit ist.

Satz: Eine symmetrische Matrix ist genau dann positiv definit, wenn ihre Eigenwerte positiv sind, und positiv semidefinit, wenn ihre Eigenwerte nicht negativ sind. (143.2)

Beweis: Für eine symmetrische Matrix A gilt nach (142.6) $X'AX = \Lambda$. Setzt man $y = X'x$, so daß wegen (141.2) $x = Xy$ folgt, ergibt sich $x'Ax = y'X'AXy = y'\Lambda y = \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 = Q$ für alle x . Da X vollen Rang besitzt, gilt wegen (122.1) $X'x = y = 0$ nur für $x = 0$. Man erhält daher $\lambda_i > 0$ für $Q > 0$. Gilt umgekehrt $\lambda_i > 0$, folgt $Q > 0$, so daß A positiv definit ist. Weiter ergibt sich $\lambda_i \geq 0$ für $Q \geq 0$ und umgekehrt aus $\lambda_i \geq 0$, daß A positiv semidefinit ist.

Satz: Eine positiv definite Matrix ist regulär. (143.3)

Beweis: Nach (143.2) besitzt eine positiv definite $n \times n$ Matrix A positive Eigenwerte, womit aus (142.6) $\text{rg}A = n$ und damit aus (133.1) die Aussage folgt.

Satz: Eine symmetrische Matrix ist genau dann positiv definit, wenn die Diagonalelemente der bei der Gaußschen Faktorisierung entstehenden Diagonalmatrix positiv sind. (143.4)

Beweis: Nach (143.3) ist eine positiv definite Matrix \mathbf{A} regulär, so daß wegen (133.22) die Gaußsche Faktorisierung $\mathbf{C}^{-1}\mathbf{A}(\mathbf{C}')^{-1} = \mathbf{D}$ gilt, in der \mathbf{C}^{-1} regulär ist. Setzt man $\mathbf{x} = (\mathbf{C}')^{-1}\mathbf{y}$, folgt der Rest des Beweises wie der für (143.2).

Satz: Eine symmetrische Matrix ist genau dann positiv definit, wenn die Cholesky-Faktorisierung $\mathbf{A} = \mathbf{G}\mathbf{G}'$ gilt, in der \mathbf{G} eine reguläre untere Dreiecksmatrix bedeutet. (143.5)

Beweis: Für eine positiv definite Matrix gilt wegen (143.4) die Cholesky-Zerlegung (133.23). Gilt umgekehrt die Cholesky-Faktorisierung, folgt mit (143.4), daß die Matrix positiv definit ist.

Satz: Ist \mathbf{A} eine positiv definite oder positiv semidefinite Matrix, gilt $\text{sp}\mathbf{A} > 0$ für $\mathbf{A} \neq \mathbf{0}$. (143.6)

Beweis: Ist \mathbf{A} positiv definit oder positiv semidefinit, ergibt sich mit (137.3) aus (142.6) $\text{sp}\mathbf{A} = \text{sp}(\mathbf{X}\mathbf{\Lambda}\mathbf{X}') = \text{sp}(\mathbf{X}'\mathbf{X}\mathbf{\Lambda}) = \text{sp}\mathbf{\Lambda} = \sum_{i=1}^n \lambda_i$ mit $\lambda_i \geq 0$ aus (143.2), so daß $\text{sp}\mathbf{A} > 0$ gilt und $\text{sp}\mathbf{A} = 0$ lediglich mit $\lambda_i = 0$ für $i \in \{1, \dots, n\}$. Im letzteren Fall gilt mit (135.5) und (142.6) $\dim N(\mathbf{A}) = n$, so daß $\mathbf{A}\mathbf{x} = \mathbf{0}$ für alle $\mathbf{x} \in E^n$ sich ergibt und damit $\mathbf{A} = \mathbf{0}$ folgt.

Satz: Ist \mathbf{A} eine positiv definite $n \times n$ Matrix, dann ist $\mathbf{B}'\mathbf{A}\mathbf{B}$ positiv definit, falls die $n \times m$ Matrix \mathbf{B} vollen Spaltenrang $m = \text{rg}\mathbf{B}$ besitzt. Bei beliebigem Rang von \mathbf{B} ist $\mathbf{B}'\mathbf{A}\mathbf{B}$ positiv definit oder positiv semidefinit. Ist \mathbf{A} positiv semidefinit, ist auch $\mathbf{B}'\mathbf{A}\mathbf{B}$ unabhängig vom Rang von \mathbf{B} positiv semidefinit. (143.7)

Beweis: \mathbf{A} sei positiv definit. Mit $\mathbf{B}\mathbf{y} = \mathbf{x}$ folgt dann $\mathbf{y}'\mathbf{B}'\mathbf{A}\mathbf{B}\mathbf{y} = \mathbf{x}'\mathbf{A}\mathbf{x} > 0$ für alle \mathbf{y} , da $\mathbf{B}\mathbf{y} = \mathbf{x} = \mathbf{0}$ wegen des vollen Spaltenrangs für \mathbf{B} nur für $\mathbf{y} = \mathbf{0}$ gilt. Bei beliebigem Rang für \mathbf{B} kann $\mathbf{x} = \mathbf{0}$ auch für $\mathbf{y} \neq \mathbf{0}$ sich ergeben, so daß $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ folgt. Ist \mathbf{A} positiv semidefinit, folgt mit $\mathbf{B}\mathbf{y} = \mathbf{x}$ unabhängig von dem Rang von \mathbf{B} schließlich $\mathbf{y}'\mathbf{B}'\mathbf{A}\mathbf{B}\mathbf{y} = \mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$.

Satz: Besitzt die $m \times n$ Matrix \mathbf{B} vollen Spaltenrang $n = \text{rg}\mathbf{B}$, dann ist $\mathbf{B}'\mathbf{B}$ positiv definit. Bei beliebigem Rang für \mathbf{B} ist $\mathbf{B}'\mathbf{B}$ positiv definit oder positiv semidefinit. (143.8)

Beweis: Ersetzt man in (143.7) \mathbf{A} durch die positiv definite Einheitsmatrix \mathbf{I} , folgen die Aussagen.

Satz: Mit \mathbf{A} ist auch \mathbf{A}^{-1} positiv definit. (143.9)

Beweis: Ersetzt man in (143.7) \mathbf{B} durch die wegen (131.16) reguläre Matrix \mathbf{A}^{-1} , folgt mit (131.17) $\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}$ und damit die Aussage.

Satz: Eine positiv semidefinite $n \times n$ Matrix \mathbf{A} mit $\text{rg}\mathbf{A} = r$ läßt sich zerlegen in $\mathbf{A} = \mathbf{H}\mathbf{H}'$, worin \mathbf{H} eine $n \times r$ Matrix mit $\text{rg}\mathbf{H} = r$ bedeutet. (143.10)

Beweis: Nach (142.6) folgt mit der orthogonalen $n \times n$ Matrix \mathbf{X}

$$\mathbf{X}'\mathbf{A}\mathbf{X} = \begin{vmatrix} \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} = \begin{vmatrix} \mathbf{D} \\ \mathbf{0} \end{vmatrix} \begin{vmatrix} \mathbf{D} & \mathbf{0} \end{vmatrix},$$

worin \mathbf{D}^2 die $r \times r$ Diagonalmatrix der von Null verschiedenen Eigenwerte von \mathbf{A} bedeutet, die wegen (143.2) positiv sind, so daß \mathbf{D} eine reelle Diagonalmatrix darstellt. Mit $\mathbf{H}' = |\mathbf{D}, \mathbf{0}| \mathbf{X}'$, wobei \mathbf{H}' eine $r \times n$ Matrix mit vollem Zeilenrang bedeutet, ergibt sich $\mathbf{A} = \mathbf{H}\mathbf{H}' = \mathbf{X}(\mathbf{X}'\mathbf{A}\mathbf{X})\mathbf{X}'$, womit die Aussage folgt.

15 Generalisierte Inversen

151 Rechts- und Linksinversen

Nach (133.1) existiert für die $n \times n$ Matrix \mathbf{A} mit $\text{rg}\mathbf{A} = n$ die inverse Matrix \mathbf{A}^{-1} derart, daß $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$ gilt. Entsprechend sollen jetzt Inversen für Rechteckmatrizen mit vollem Zeilen- oder Spaltenrang eingeführt werden.

Es sei \mathbf{A} eine $m \times n$ Matrix mit vollem Zeilenrang m . Die $m \times m$ Matrix $\mathbf{A}\mathbf{A}'$ besitzt nach (135.6) ebenfalls den Rang m , so daß $(\mathbf{A}\mathbf{A}')^{-1}$ existiert. Es gilt dann

$$\mathbf{I}_m = \mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1} = \mathbf{A}(\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}) = \mathbf{A}\mathbf{B}, \quad (151.1)$$

so daß eine $n \times m$ Matrix \mathbf{B} angegeben werden kann, mit der \mathbf{A} von rechts multipliziert die $m \times m$ Einheitsmatrix \mathbf{I}_m ergibt. \mathbf{B} bezeichnet man als *Rechtsinverse* von \mathbf{A} . Die Rechtsinverse ist nicht eindeutig, denn falls $\text{rg}(\mathbf{A}\mathbf{C}\mathbf{A}') = m$ gilt, ist beispielsweise $\mathbf{C}\mathbf{A}'(\mathbf{A}\mathbf{C}\mathbf{A}')^{-1}$ ebenfalls eine Rechtsinverse von \mathbf{A} .

Die $m \times n$ Matrix \mathbf{A} besitze nun vollen Spaltenrang n . Dann gilt wegen (135.6) $\text{rg}(\mathbf{A}'\mathbf{A}) = n$, so daß folgt

$$\mathbf{I}_n = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A} = ((\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}')\mathbf{A} = \mathbf{B}\mathbf{A}. \quad (151.2)$$

\mathbf{B} bezeichnet man als *Linksinverse* von \mathbf{A} , sie ist wie die Rechtsinverse nicht eindeutig.

152 Idempotente Matrizen

Im Zusammenhang mit den im folgenden Kapitel zu behandelnden generalisierten Inversen und später wieder bei den Projektionen treten idempotente Matrizen auf.

Definition: Eine quadratische Matrix heißt *idempotent*, wenn sie die Bedingung erfüllt

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}. \quad (152.1)$$

Folgende Eigenschaften idempotenter Matrizen sind von Interesse.

Satz: Die Eigenwerte einer idempotenten Matrix sind entweder Null oder Eins. (152.2)

Beweis: Es sei λ ein Eigenwert von \mathbf{A} und \mathbf{x} ein zugehöriger Eigenvektor. Aus (142.2) folgt dann $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ und weiter $\mathbf{A}\mathbf{A}\mathbf{x} = \lambda\mathbf{A}\mathbf{x} = \lambda^2\mathbf{x}$. Wegen $\mathbf{A}^2 = \mathbf{A}$ ist aber $\mathbf{A}^2\mathbf{x} = \lambda\mathbf{x}$ und daher $(\lambda^2 - \lambda)\mathbf{x} = \mathbf{0}$. Da $\mathbf{x} \neq \mathbf{0}$, folgt $\lambda(\lambda - 1) = 0$ und daher $\lambda = 0$ oder $\lambda = 1$.

Satz: Ist \mathbf{A} idempotent, dann gilt $\text{rg}\mathbf{A} = \text{sp}\mathbf{A}$. (152.3)

Beweis: Es sei $\text{rg}\mathbf{A} = r$. Aus der Rangfaktorisierung (132.11) für \mathbf{A} folgt $\mathbf{A} = \mathbf{R}\mathbf{S}$ und mit (152.1) $\mathbf{R}\mathbf{S}\mathbf{R}\mathbf{S} = \mathbf{R}\mathbf{S}$. Nach (151.1) und (151.2) besitzt \mathbf{R} eine Linksinverse \mathbf{L} und \mathbf{S} eine Rechtsinverse \mathbf{T} , so daß $\mathbf{L}\mathbf{R}\mathbf{S}\mathbf{R}\mathbf{S}\mathbf{T} = \mathbf{L}\mathbf{R}\mathbf{S}\mathbf{T}$ und damit $\mathbf{S}\mathbf{R} = \mathbf{I}_r$ sich ergibt. Schließlich gilt mit (137.3) $\text{sp}\mathbf{A} = \text{sp}(\mathbf{S}\mathbf{R}) = \text{sp}\mathbf{I}_r = r$, so daß $r = \text{rg}\mathbf{A} = \text{sp}\mathbf{A}$ folgt.

Satz: Ist die $n \times n$ Matrix \mathbf{A} mit $\text{rg}\mathbf{A} = r$ idempotent, dann ist auch $\mathbf{I} - \mathbf{A}$ idempotent mit $\text{rg}(\mathbf{I} - \mathbf{A}) = n - r$. (152.4)

Beweis: Es gilt $(\mathbf{I} - \mathbf{A})^2 = \mathbf{I} - 2\mathbf{A} + \mathbf{A}^2 = \mathbf{I} - \mathbf{A}$. Mit (152.3) erhält man weiter $\text{rg}(\mathbf{I} - \mathbf{A}) = \text{sp}(\mathbf{I} - \mathbf{A}) = n - \text{sp}(\mathbf{A}) = n - \text{rg}\mathbf{A} = n - r$.

Satz: Ist \mathbf{A} idempotent und regulär, ist $\mathbf{A} = \mathbf{I}$. (152.5)

Beweis: Multipliziert man $\mathbf{A}\mathbf{A} = \mathbf{A}$ von links oder rechts mit \mathbf{A}^{-1} , erhält man $\mathbf{A} = \mathbf{I}$.

Im folgenden sollen noch drei Sätze bewiesen werden, die für symmetrische idempotente Matrizen gelten.

Satz: Ist die Matrix \mathbf{A} mit $\text{rg}\mathbf{A} = r$ idempotent und symmetrisch, gibt es eine orthogonale Matrix \mathbf{X} , so daß gilt

$$\mathbf{X}'\mathbf{A}\mathbf{X} = \begin{vmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix}. \quad (152.6)$$

Beweis: Mit (142.6) folgt für die symmetrische Matrix \mathbf{A} die Zerlegung $\mathbf{X}'\mathbf{A}\mathbf{X} = \mathbf{\Lambda}$, in der die Diagonalmatrix $\mathbf{\Lambda}$ der Eigenwerte wegen (152.2) r Werte Eins besitzt, so daß die Aussage folgt.

Satz: Es gilt $\mathbf{A} = \mathbf{p}_1\mathbf{p}'_1 + \dots + \mathbf{p}_r\mathbf{p}'_r$, wobei $\mathbf{p}_1, \dots, \mathbf{p}_r$ orthonormale Vektoren bedeuten, genau dann, wenn \mathbf{A} mit $\text{rg}\mathbf{A} = r$ idempotent und symmetrisch ist. (152.7)

Beweis: Ist \mathbf{A} mit $\text{rg}\mathbf{A} = r$ idempotent und symmetrisch, ergibt sich mit einer orthogonalen Matrix \mathbf{X} aus (152.6)

$$\begin{aligned} \mathbf{A} &= \mathbf{X} \begin{vmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} \mathbf{X}' = \mathbf{P}\mathbf{P}' = |\mathbf{p}_1, \dots, \mathbf{p}_r| \begin{vmatrix} \mathbf{p}'_1 \\ \dots \\ \mathbf{p}'_r \end{vmatrix} \\ &= \mathbf{p}_1\mathbf{p}'_1 + \dots + \mathbf{p}_r\mathbf{p}'_r, \end{aligned}$$

denn die Matrix \mathbf{P} besitzt r Spalten \mathbf{p}_i , die orthonormale Vektoren sind. Wird andererseits \mathbf{A} mit Hilfe r orthonormaler Vektoren \mathbf{p}_i dargestellt und faßt man sie in der Matrix $\mathbf{P} = |\mathbf{p}_1, \dots, \mathbf{p}_r|$ zusammen, besitzt \mathbf{P} wegen (124.2) und (132.2) den Rang r . Mit (135.6) folgt dann $r = \text{rg}(\mathbf{P}\mathbf{P}') = \text{rg}(\sum_{i=1}^r \mathbf{p}_i\mathbf{p}'_i) = \text{rg}\mathbf{A}$. Weiter ist \mathbf{A} mit $\mathbf{A} = \mathbf{P}\mathbf{P}'$ symmetrisch und idempotent, denn mit (131.7) folgt $\mathbf{A}^2 = (\mathbf{p}_1\mathbf{p}'_1 + \dots + \mathbf{p}_r\mathbf{p}'_r)^2 = \sum_{i=1}^r \mathbf{p}_i(\mathbf{p}'_i\mathbf{p}_i)\mathbf{p}'_i + \sum_{i \neq j} \mathbf{p}_i(\mathbf{p}'_i\mathbf{p}_j)\mathbf{p}'_j = \sum_{i=1}^r \mathbf{p}_i\mathbf{p}'_i = \mathbf{A}$ wegen $\mathbf{p}'_i\mathbf{p}_i = 1$ und $\mathbf{p}'_i\mathbf{p}_j = 0$, da die Vektoren \mathbf{p}_i orthonormal sind, so daß die Aussage folgt.

Satz: Ist eine Matrix, die nicht die Einheitsmatrix ist, idempotent und symmetrisch, so ist sie positiv semidefinit. (152.8)

Beweis: Gilt $\mathbf{A}\mathbf{A} = \mathbf{A}$ und $\mathbf{A} = \mathbf{A}'$, folgt mit $\mathbf{A}'\mathbf{A} = \mathbf{A}$ aus (143.8) und (152.5) die Aussage.

153 Generalisierte Inverse, reflexive generalisierte Inverse und Pseudoinverse

a) Generalisierte Inverse

Die Definition einer inversen Matrix, die sich bislang auf reguläre quadratische Matrizen und auf Rechteckmatrizen mit vollem Zeilen- oder Spaltenrang beschränkte, soll nun auf Rechteckmatrizen von beliebigem Rang ausgedehnt werden.

Definition: Eine $n \times m$ Matrix \mathbf{A}^- bezeichnet man als *generalisierte Inverse* der $m \times n$ Matrix \mathbf{A} , falls

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}. \quad (153.1)$$

Daß diese Definition im Hinblick auf die Lösung linearer Gleichungssysteme sinnvoll ist, wird im nächsten Kapitel gezeigt. Es sollen nun Eigenschaften generalisierter Inversen betrachtet werden.

Satz: Eine generalisierte Inverse \mathbf{A}^- der $m \times n$ Matrix \mathbf{A} mit $m \geq n$ und $r = \text{rg}\mathbf{A}$ existiert für jedes fest vorgegebene $r \leq k \leq n$ mit $\text{rg}\mathbf{A}^- = k$. (153.2)

Beweis: Aus (132.10) folgt für \mathbf{A} mit den regulären $m \times m$ und $n \times n$ Matrizen \mathbf{P} und \mathbf{Q}

$$\mathbf{PAQ} = \begin{vmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} \quad \text{und} \quad \mathbf{A} = \mathbf{P}^{-1} \begin{vmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} \mathbf{Q}^{-1}.$$

Sind $\mathbf{R}, \mathbf{S}, \mathbf{T}$ beliebige Matrizen zutreffender Dimensionen, ist eine generalisierte Inverse \mathbf{A}^- gegeben mit

$$\mathbf{A}^- = \mathbf{Q} \begin{vmatrix} \mathbf{I}_r & \mathbf{R} \\ \mathbf{S} & \mathbf{T} \end{vmatrix} \mathbf{P}, \quad (153.3)$$

denn

$$\mathbf{AA}^- \mathbf{A} = \mathbf{P}^{-1} \begin{vmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} \begin{vmatrix} \mathbf{I}_r & \mathbf{R} \\ \mathbf{S} & \mathbf{T} \end{vmatrix} \begin{vmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} \mathbf{Q}^{-1} = \mathbf{A}.$$

Setzt man beispielsweise $\mathbf{R} = \mathbf{0}, \mathbf{S} = \mathbf{0}$ und $\mathbf{T} = \begin{vmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix}$ mit $0 \leq p \leq n - r$ in (153.3), ergibt sich wegen (132.4) $\text{rg}\mathbf{A}^- = r + p$. Dies gilt für alle generalisierten Inversen \mathbf{A}^- , denn aus $\mathbf{AA}^- \mathbf{A} = \mathbf{A}$ folgt mit (132.3) $\text{rg}\mathbf{A}^- \geq \text{rg}(\mathbf{AA}^-) \geq \text{rg}(\mathbf{AA}^- \mathbf{A}) = \text{rg}\mathbf{A}$ und somit $\text{rg}\mathbf{A} \leq \text{rg}\mathbf{A}^- \leq n$.

Satz: Das Produkt $\mathbf{A}^- \mathbf{A}$ ist idempotent und $\text{rg}(\mathbf{A}^- \mathbf{A}) = \text{rg}\mathbf{A}$. (153.4)

Beweis: Es gilt $(\mathbf{A}^- \mathbf{A})^2 = \mathbf{A}^- \mathbf{AA}^- \mathbf{A} = \mathbf{A}^- \mathbf{A}$ wegen (153.1). Mit (132.3) folgt $\text{rg}\mathbf{A} \geq \text{rg}(\mathbf{A}^- \mathbf{A}) \geq \text{rg}(\mathbf{AA}^- \mathbf{A}) = \text{rg}\mathbf{A}$ und daher $\text{rg}\mathbf{A} = \text{rg}(\mathbf{A}^- \mathbf{A})$.

Satz: Es gilt

$$\mathbf{A}(\mathbf{A}'\mathbf{A})^- \mathbf{A}'\mathbf{A} = \mathbf{A}, \quad \mathbf{A}'\mathbf{A}(\mathbf{A}'\mathbf{A})^- \mathbf{A}' = \mathbf{A}'$$

sowie

$$\mathbf{A}'(\mathbf{AA}')^- \mathbf{AA}' = \mathbf{A}', \quad \mathbf{AA}'(\mathbf{AA}')^- \mathbf{A} = \mathbf{A} \quad (153.5)$$

und falls \mathbf{V} eine positiv definite Matrix ist

$$\mathbf{A}(\mathbf{A}'\mathbf{V}\mathbf{A})^- \mathbf{A}'\mathbf{V}\mathbf{A} = \mathbf{A}, \quad \mathbf{A}'\mathbf{V}\mathbf{A}(\mathbf{A}'\mathbf{V}\mathbf{A})^- \mathbf{A}' = \mathbf{A}'. \quad (153.6)$$

Beweis: Setzt man $E = A(A'A)^- A'A - A$ folgt $E'E = (A(A'A)^- A'A - A)'A((A'A)^- A'A - I) = \mathbf{0}$ wegen (131.10) und (153.1) und daher $E = \mathbf{0}$ nach (131.11). Setzt man $E = A'A(A'A)^- A' - A'$ folgt aus $EE' = \mathbf{0}$ die zweite Gleichung von (153.5) und entsprechend die folgenden beiden Beziehungen. Setzt man weiter $E = A(A'VA)^- A'VA - A$, folgt $E'VE = (A(A'VA)^- A'VA - A)'VA((A'VA)^- A'VA - I) = \mathbf{0}$ und mit (131.11) sowie $V = GG'$ aus (143.5) $G'E = \mathbf{0}$ oder $GG'E = \mathbf{0}$ und somit $E = \mathbf{0}$. Die zweite Gleichung von (153.6) ergibt sich entsprechend.

Satz: Ist G generalisierte Inverse von $A'A$ und F generalisierte Inverse von AA' , dann sind

- a) G' beziehungsweise F' generalisierte Inversen von $A'A$ beziehungsweise AA' , (153.7)
- b) AGA' beziehungsweise $A'FA$ invariant gegenüber der Wahl von G beziehungsweise F , (153.8)
- c) AGA' beziehungsweise $A'FA$ symmetrisch unabhängig davon, ob G beziehungsweise F symmetrisch ist, (153.9)
- d) $A(A'VA)^- A'$ invariant gegenüber der Wahl von $(A'VA)^-$ und immer symmetrisch, falls V eine positiv definite Matrix ist. (153.10)

Beweis: (153.7) folgt durch Transponieren von $A'AGA'A = A'A$ und der entsprechenden Gleichung für F . Es seien G und \bar{G} zwei generalisierte Inversen von $A'A$. Dann erhält man mit (153.5) $AGA'A = A\bar{G}A'A$ und weiter $AGA'AGA' = A\bar{G}A'AGA'$, so daß mit (153.5) $AGA' = A\bar{G}A'$ und damit (153.8) folgt, da entsprechende Gleichungen für F gelten. Für symmetrische Matrizen existieren auch symmetrische generalisierte Inversen, was aus (153.3) folgt, da für symmetrische Matrizen A in (132.10) $Q = P'$ gilt. Ist G eine symmetrische generalisierte Inverse von $A'A$, ist auch AGA' symmetrisch. Da aber AGA' unabhängig von der Wahl von G ist, muß AGA' und entsprechend $A'FA$ immer symmetrisch sein. Aus (153.6) folgt genau, wie für AGA' gezeigt wurde, daß $A(A'VA)^- A'$ invariant gegenüber der Wahl von $(A'VA)^-$ und daher symmetrisch ist, so daß (153.10) folgt.

In der Menge der generalisierten Inversen, die (153.1) erfüllen, lassen sich durch Zusatzbedingungen Teilmengen spezieller generalisierter Inversen definieren (BEN-ISRAEL und GREVILLE 1974; BJERHAMMAR 1973; BOULLION und ODELL 1971; CASPARY und WICHMANN 1994; RAO und MITRA 1971). Im folgenden sollen lediglich die reflexive generalisierte Inverse und die Pseudoinverse eingeführt werden, die für die Parameterschätzung in Modellen mit nicht vollem Rang benötigt werden.

b) Reflexive generalisierte Inverse

Definition: Eine $n \times m$ Matrix \mathbf{A}_r^- bezeichnet man als *reflexive generalisierte Inverse* der $m \times n$ Matrix \mathbf{A} , falls

$$\mathbf{A}\mathbf{A}_r^-\mathbf{A} = \mathbf{A} \quad \text{und} \quad \mathbf{A}_r^-\mathbf{A}\mathbf{A}_r^- = \mathbf{A}_r^- . \quad (153.11)$$

Satz: Eine reflexive generalisierte Inverse ist durch $\mathbf{A}_r^- = \mathbf{A}^-\mathbf{A}\mathbf{A}^-$ gegeben. (153.12)

Beweis: Mit $\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ und $\mathbf{A}^-\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-\mathbf{A}\mathbf{A}^-$ ist (153.11) erfüllt.

Satz: \mathbf{A}_r^- ist genau dann reflexive generalisierte Inverse von \mathbf{A} , wenn \mathbf{A}_r^- generalisierte Inverse von \mathbf{A} und $\text{rg}\mathbf{A}_r^- = \text{rg}\mathbf{A}$ gilt. (153.13)

Beweis: Ist \mathbf{A}_r^- reflexive generalisierte Inverse von \mathbf{A} , ist sie auch generalisierte Inverse von \mathbf{A} und mit (132.3) und (153.11) folgt $\text{rg}\mathbf{A} \geq \text{rg}(\mathbf{A}_r^-\mathbf{A}) \geq \text{rg}(\mathbf{A}_r^-\mathbf{A}\mathbf{A}_r^-) = \text{rg}\mathbf{A}_r^- \geq \text{rg}(\mathbf{A}_r^-\mathbf{A}) \geq \text{rg}(\mathbf{A}\mathbf{A}_r^-\mathbf{A}) = \text{rg}\mathbf{A}$ und damit $\text{rg}\mathbf{A}_r^- = \text{rg}\mathbf{A}$. Gilt andererseits $\text{rg}\mathbf{A}_r^- = \text{rg}\mathbf{A}$ und $\mathbf{A}\mathbf{A}_r^-\mathbf{A} = \mathbf{A}$ folgt mit (153.4) $\text{rg}(\mathbf{A}_r^-\mathbf{A}) = \text{rg}\mathbf{A}_r^-$, so daß für den Fall $n \leq m$ mit einer beliebigen $n \times m$ Matrix \mathbf{Q} mit vollem Zeilenrang $\mathbf{A}_r^- = \mathbf{A}_r^-\mathbf{A}\mathbf{Q}$ gesetzt werden kann, da $\text{rg}\mathbf{A}_r^-$ wegen $\text{rg}(\mathbf{A}_r^-\mathbf{A}) \geq \text{rg}(\mathbf{A}_r^-\mathbf{A}\mathbf{Q}) \geq \text{rg}(\mathbf{A}_r^-\mathbf{A}\mathbf{Q}\mathbf{Q}'(\mathbf{Q}\mathbf{Q}')^{-1}) = \text{rg}(\mathbf{A}_r^-\mathbf{A})$ unverändert bleibt. Durch linksseitige Multiplikation mit \mathbf{A} erhält man $\mathbf{A}\mathbf{A}_r^- = \mathbf{A}\mathbf{A}_r^-\mathbf{A}\mathbf{Q} = \mathbf{A}\mathbf{Q}$ und schließlich $\mathbf{A}_r^-\mathbf{A}\mathbf{A}_r^- = \mathbf{A}_r^-\mathbf{A}\mathbf{Q} = \mathbf{A}_r^-$. Für den Fall $n > m$ setzt man $\mathbf{A}_r^- = \mathbf{Q}\mathbf{A}\mathbf{A}_r^-$, so daß $\mathbf{A}_r^-\mathbf{A} = \mathbf{Q}\mathbf{A}\mathbf{A}_r^-\mathbf{A} = \mathbf{Q}\mathbf{A}$ und $\mathbf{A}_r^-\mathbf{A}\mathbf{A}_r^- = \mathbf{Q}\mathbf{A}\mathbf{A}_r^- = \mathbf{A}_r^-$ und damit die Aussage folgt.

Satz: Eine symmetrische reflexive generalisierte Inverse $(\mathbf{A}'\mathbf{A})_{rs}^-$ von $\mathbf{A}'\mathbf{A}$ ist positiv semidefinit. (153.14)

Beweis: Die Existenz einer symmetrischen reflexiven generalisierten Inversen $(\mathbf{A}'\mathbf{A})_{rs}^-$ von $\mathbf{A}'\mathbf{A}$ ergibt sich mit (153.13) aus (153.3), da für symmetrische Matrizen $\mathbf{A}'\mathbf{A}$ in (132.10) $\mathbf{Q} = \mathbf{P}'$ gilt. Da $\mathbf{A}'\mathbf{A}$ nach (143.8) bei beliebigem Rang von \mathbf{A} positiv semidefinit ist, ergibt sich nach (143.7), daß auch $(\mathbf{A}'\mathbf{A})_{rs}^-\mathbf{A}'\mathbf{A}(\mathbf{A}'\mathbf{A})_{rs}^- = (\mathbf{A}'\mathbf{A})_{rs}^-$ positiv semidefinit ist, so daß die Aussage folgt.

c) Pseudoinverse

Für die Pseudoinverse, auch *Moore–Penrose–Inverse* genannt, gilt die

Definition: Die $n \times m$ Matrix \mathbf{A}^+ ist *Pseudoinverse* der $m \times n$ Matrix \mathbf{A} , falls

$$\begin{aligned} \mathbf{A}\mathbf{A}^+\mathbf{A} &= \mathbf{A} , \quad \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+ , \\ (\mathbf{A}\mathbf{A}^+)' &= \mathbf{A}\mathbf{A}^+ , \quad (\mathbf{A}^+\mathbf{A})' = \mathbf{A}^+\mathbf{A} . \end{aligned} \quad (153.15)$$

Satz: Es gilt $\mathbf{A}^+ = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$. (153.16)

Beweis: Substituiert man die Gleichung für \mathbf{A}^+ in (153.15), folgt $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A} = \mathbf{A}$ wegen (153.5). Entsprechend ergibt sich durch Substitution $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$. Weiter sind die Matrizen $\mathbf{A}\mathbf{A}^+ = \mathbf{A}\mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ und $\mathbf{A}^+\mathbf{A} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}\mathbf{A}$ wegen (153.9) symmetrisch, so daß die Aussage folgt.

Satz: \mathbf{A}^+ ist eindeutig und $\text{rg}\mathbf{A}^+ = \text{rg}\mathbf{A}$. (153.17)

Beweis: \mathbf{G} und \mathbf{F} seien zwei Pseudoinversen von \mathbf{A} . Dann gilt wegen (153.15) $\mathbf{G} = \mathbf{G}\mathbf{G}'\mathbf{A}' = \mathbf{G}\mathbf{G}'\mathbf{A}'\mathbf{A}\mathbf{F} = \mathbf{G}\mathbf{A}\mathbf{F} = \mathbf{G}\mathbf{A}\mathbf{A}'\mathbf{F}'\mathbf{F} = \mathbf{A}'\mathbf{F}'\mathbf{F} = \mathbf{F}$, so daß $\mathbf{G} = \mathbf{F} = \mathbf{A}^+$ folgt. Aus (153.13) ergibt sich $\text{rg}\mathbf{A}^+ = \text{rg}\mathbf{A}$.

Satz: Es gilt $(\mathbf{A}')^+ = (\mathbf{A}^+)'$, so daß für $\mathbf{A}' = \mathbf{A}$ folgt $\mathbf{A}^+ = (\mathbf{A}^+)'$. (153.18)

Beweis: Aus (153.15) ergibt sich $(\mathbf{A}')^+$ als Pseudoinverse von \mathbf{A}' . Transponiert man (153.15), so erhält man $(\mathbf{A}^+)'$ als Pseudoinverse von \mathbf{A}' . Da die Pseudoinverse eindeutig ist, folgt $(\mathbf{A}')^+ = (\mathbf{A}^+)'$ und für eine symmetrische Matrix \mathbf{A} die symmetrische Pseudoinverse $\mathbf{A}^+ = (\mathbf{A}^+)'$.

Satz: Es gilt $(\mathbf{A}^+)^+ = \mathbf{A}$. (153.19)

Beweis: Für die Pseudoinverse $(\mathbf{A}^+)^+$ von \mathbf{A}^+ gelten ebenso wie für die Pseudoinverse \mathbf{A}^+ von \mathbf{A} die vier Bedingungen (153.15), so daß wegen der Eindeutigkeit der Pseudoinversen $(\mathbf{A}^+)^+ = \mathbf{A}$ folgt.

Satz: Für die Matrix \mathbf{A} mit vollem Zeilenrang und die Matrix \mathbf{B} mit vollem Spaltenrang gilt $\mathbf{A}^+ = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ und $\mathbf{B}^+ = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$, wobei \mathbf{A}^+ gleichzeitig eine Rechtsinverse von \mathbf{A} und \mathbf{B}^+ eine Linksinverse von \mathbf{B} ist. (153.20)

Beweis: Aus (151.1) folgt eine Rechtsinverse \mathbf{R} von \mathbf{A} zu $\mathbf{R} = \mathbf{A}'(\mathbf{A}\mathbf{A}')^{-1}$ und aus (151.2) eine Linksinverse \mathbf{L} von \mathbf{B} zu $\mathbf{L} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$. Die Matrizen \mathbf{R} und \mathbf{L} erfüllen (153.15), so daß die Aussage folgt.

Satz: Für eine beliebige $m \times n$ Matrix \mathbf{A} gilt $\mathbf{A}^+ = \lim_{\delta \rightarrow 0} (\mathbf{A}'\mathbf{A} + \delta^2\mathbf{I})^{-1}\mathbf{A}' = \lim_{\delta \rightarrow 0} \mathbf{A}'(\mathbf{A}\mathbf{A}' + \delta^2\mathbf{I})^{-1}$. (153.21)

Der Beweis dieses Satzes befindet sich in (ALBERT 1972, S.19).

Weitere Gleichungen für die Pseudoinverse, beispielsweise diejenigen, die sich auf die Rangfaktorisierung (132.11) stützen, befinden sich in (BEN-ISRAEL und GREVILLE 1974; BOULLION und ODELL 1971; GRAYBILL 1969; RAO und MITRA 1971). Rechenformeln für generalisierte Inversen symmetrischer Matrizen werden im Kapitel 155 angegeben und weitere Eigenschaften der Pseudoinversen und einer speziellen symmetrischen reflexiven generalisierten

Inversen im Kapitel 156 behandelt.

Die Beziehung zwischen den Inversen regulärer Matrizen und den generalisierten Inversen ergibt sich aus dem

Satz: Ist \mathbf{A} eine reguläre $n \times n$ Matrix, dann gilt

$$\mathbf{A}^- = \mathbf{A}_r^- = \mathbf{A}^+ = \mathbf{A}^{-1}. \quad (153.22)$$

Beweis: Substituiert man die generalisierten Inversen in (153.1), (153.11) und (153.15) durch \mathbf{A}^{-1} , sind sämtliche Bedingungen erfüllt. Weiter folgt aus $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ für beliebige \mathbf{A}^- mit $\mathbf{A}^{-1}\mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1} = \mathbf{A}^-$, daß außer \mathbf{A}^{-1} keine weitere generalisierte Inverse von \mathbf{A} existiert.

154 Lineare Gleichungssysteme

In (133.14) wurde bereits ein lineares Gleichungssystem mit quadratischer Koeffizientenmatrix eingeführt und die Lösung für eine reguläre Koeffizientenmatrix angegeben. In dem linearen Gleichungssystem

$$\mathbf{A}\boldsymbol{\beta} = \mathbf{l} \quad (154.1)$$

sei nun \mathbf{A} eine $m \times n$ Koeffizientenmatrix mit beliebigem Rang $r = \text{rg}\mathbf{A}$, $\boldsymbol{\beta}$ der $n \times 1$ Vektor unbekannter Parameter und \mathbf{l} der $m \times 1$ Absolutgliedvektor. Ist $\mathbf{l} = \mathbf{0}$, bezeichnet man (154.1) als *homogenes Gleichungssystem*. Um die Lösungsbedingungen von (154.1) zu formulieren, benötigt man die folgende

Definition: Ein lineares Gleichungssystem $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ heißt *konsistent*, wenn \mathbf{l} Element des Spaltenraums von \mathbf{A} ist, also $\mathbf{l} \in R(\mathbf{A})$. (154.2)

Die Konsistenz bedeutet nach (135.1), daß zu jedem beliebigen Vektor \mathbf{l} ein Vektor \mathbf{w} existiert, so daß $\mathbf{A}\mathbf{w} = \mathbf{l}$ gilt.

Satz: Ein Gleichungssystem ist genau dann lösbar, wenn es konsistent ist. (154.3)

Beweis: Ist das Gleichungssystem $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ lösbar, dann existiert ein Vektor $\boldsymbol{\beta}$, so daß $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$, woraus die Konsistenz, also $\mathbf{l} \in R(\mathbf{A})$ folgt. Ist umgekehrt $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ konsistent, gilt $\mathbf{l} \in R(\mathbf{A})$, so daß für jeden Vektor \mathbf{l} ein Vektor \mathbf{w} derart existiert, daß $\mathbf{A}\mathbf{w} = \mathbf{l}$, womit sich \mathbf{w} als Lösung ergibt.

Zwei äquivalente Formulierungen der Lösungsbedingungen (154.3) werden in dem folgenden Satz gegeben.

Satz: Das Gleichungssystem $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ ist genau dann lösbar, wenn für den Spaltenraum der um den Vektor \mathbf{l} erweiterten Matrix \mathbf{A} die Beziehung $\text{rg}[\mathbf{A}, \mathbf{l}] = \text{rg}\mathbf{A}$ gilt oder wenn jede Lösung \mathbf{z} des homogenen Gleichungssystems $\mathbf{A}'\mathbf{z} = \mathbf{0}$

orthogonal zum Absolutgliedvektor \mathbf{l} ist. (154.4)

Beweis: Ist das Gleichungssystem lösbar, gilt $\mathbf{l} \in R(\mathbf{A})$ und $R(|\mathbf{A}, \mathbf{l}|) = R(\mathbf{A})$, so daß mit (135.5) $\text{rg}|\mathbf{A}, \mathbf{l}| = \text{rg}\mathbf{A}$ folgt. Gilt andererseits $\text{rg}|\mathbf{A}, \mathbf{l}| = \text{rg}\mathbf{A}$, folgt mit $R(\mathbf{A}) \subset R(|\mathbf{A}, \mathbf{l}|)$ auch $R(\mathbf{A}) = R(|\mathbf{A}, \mathbf{l}|)$, so daß $\mathbf{l} \in R(\mathbf{A})$ und damit die Gleichung lösbar ist und die erste Aussage folgt. Ist das Gleichungssystem lösbar, gilt $R(\mathbf{A}) = R(|\mathbf{A}, \mathbf{l}|)$ und nach (135.4) $N(\mathbf{A}') = R(\mathbf{A})^\perp$, so daß alle Lösungen \mathbf{z} der homogenen Gleichungen $\mathbf{A}'\mathbf{z} = \mathbf{0}$ wegen $\mathbf{z} \in N(\mathbf{A}')$ orthogonal zu sämtlichen Vektoren $\mathbf{y} \in R(|\mathbf{A}, \mathbf{l}|)$, also auch zu \mathbf{l} sind. Sind umgekehrt die Lösungen \mathbf{z} orthogonal zu \mathbf{l} , muß wegen $N(\mathbf{A}') = R(\mathbf{A})^\perp$ gelten $\mathbf{l} \in R(\mathbf{A})$, woraus die Lösbarkeit und damit die zweite Aussage folgt.

Eine Lösung von (154.1) ergibt sich mit dem

Satz: Konsistente Gleichungssysteme $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ besitzen genau dann eine Lösung $\boldsymbol{\beta} = \mathbf{A}^-\mathbf{l}$ für alle \mathbf{l} , falls \mathbf{A}^- eine generalisierte Inverse von \mathbf{A} ist. (154.5)

Beweis: Es sei $\boldsymbol{\beta} = \mathbf{A}^-\mathbf{l}$ eine Lösung. Da $\mathbf{l} \in R(\mathbf{A})$, existiert für jedes \mathbf{l} ein Vektor \mathbf{w} derart, daß $\mathbf{A}\mathbf{w} = \mathbf{l}$. Folglich ist $\mathbf{l} = \mathbf{A}\boldsymbol{\beta} = \mathbf{A}\mathbf{A}^-\mathbf{l} = \mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{w} = \mathbf{A}\mathbf{w}$ für alle \mathbf{w} , so daß $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ gelten muß. Ist umgekehrt \mathbf{A} generalisierte Inverse von \mathbf{A} , dann ist $\mathbf{A}\mathbf{A}^-\mathbf{A}\boldsymbol{\beta} = \mathbf{A}\boldsymbol{\beta}$ und $\mathbf{A}\mathbf{A}^-\mathbf{l} = \mathbf{l}$, so daß $\boldsymbol{\beta} = \mathbf{A}^-\mathbf{l}$ eine Lösung von $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ ist.

Schreibt man $\mathbf{A}\boldsymbol{\beta} = \mathbf{A}(\boldsymbol{\beta}_0 + \boldsymbol{\beta}) = \mathbf{0} + \mathbf{l}$, wird offensichtlich, daß zur Lösung $\boldsymbol{\beta} = \mathbf{A}^-\mathbf{l}$ des konsistenten Gleichungssystems $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ noch die allgemeine Lösung des homogenen Gleichungssystems $\mathbf{A}\boldsymbol{\beta}_0 = \mathbf{0}$ addiert werden muß, um die allgemeine Lösung von $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ zu erhalten.

Satz: Eine allgemeine Lösung des homogenen Gleichungssystems $\mathbf{A}\boldsymbol{\beta}_0 = \mathbf{0}$ mit der $m \times n$ Koeffizientenmatrix \mathbf{A} vom Rang r , einer generalisierten Inversen \mathbf{A}^- von \mathbf{A} und dem $n \times 1$ Vektor $\boldsymbol{\beta}_0$ unbekannter Parameter ergibt sich mit dem beliebigen $n \times 1$ Vektor \mathbf{z} zu

$$\boldsymbol{\beta}_0 = (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z} ,$$

so daß der durch diese Lösungen aufgespannte Spaltenraum $R(\mathbf{I} - \mathbf{A}^-\mathbf{A})$, der als *Lösungsraum* bezeichnet wird, mit dem Nullraum $N(\mathbf{A})$ der Matrix \mathbf{A} identisch ist. (154.6)

Beweis: Da $\mathbf{A}\boldsymbol{\beta}_0 = \mathbf{A}(\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{z} = \mathbf{0}$ wegen (153.1) gilt, ist $\boldsymbol{\beta}_0$ eine Lösung von $\mathbf{A}\boldsymbol{\beta}_0 = \mathbf{0}$ und somit $R(\mathbf{I} - \mathbf{A}^-\mathbf{A}) \subset N(\mathbf{A})$. Nach (153.4) ist $\mathbf{A}^-\mathbf{A}$ idempotent, so daß mit (135.5) und (152.4) $\text{rg}(\mathbf{I} - \mathbf{A}^-\mathbf{A}) = \dim R(\mathbf{I} - \mathbf{A}^-\mathbf{A}) = n - r = \dim N(\mathbf{A})$ und damit $R(\mathbf{I} - \mathbf{A}^-\mathbf{A}) = N(\mathbf{A})$ folgt, so daß mit $\boldsymbol{\beta}_0$ eine allgemeine Lösung des homogenen Gleichungssystems gefunden ist.

Eine allgemeine Lösung von $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ ergibt sich nun durch Addition der Lösun-

gen in (154.5) und (154.6).

Satz: Eine allgemeine Lösung des konsistenten Gleichungssystems $\mathbf{A}\boldsymbol{\beta} = \mathbf{l}$ mit der $m \times n$ Koeffizientenmatrix \mathbf{A} vom Rang r , einer generalisierten Inversen \mathbf{A}^- von \mathbf{A} , dem $n \times 1$ Vektor $\boldsymbol{\beta}$ unbekannter Parameter und dem $m \times 1$ Vektor \mathbf{l} der Absolutglieder ergibt sich mit dem beliebigen $n \times 1$ Vektor \mathbf{z} zu

$$\boldsymbol{\beta} = \mathbf{A}^- \mathbf{l} + (\mathbf{I} - \mathbf{A}^- \mathbf{A}) \mathbf{z} . \quad (154.7)$$

Die allgemeine Lösung (154.7) ist nicht eindeutig, denn es existieren wegen (154.6) $n - r$ linear unabhängige Lösungen in dem Lösungsraum der homogenen Gleichungen. Ist $n - r = 1$, besteht der Lösungsraum aus einer Linie, auf der die Lösungen sich verschieben lassen. Diese Unbestimmtheit tritt beispielsweise ein, wenn aus Messungen von Schwerkraftdifferenzen Absolutwerte der Schwerkraft bestimmt werden sollen. Ist $n - r = 2$, besteht der Lösungsraum aus einer Ebene, in der die Lösungen willkürlich zu verschieben sind. Für höhere Werte von $n - r$ ergeben sich entsprechende Interpretationen.

Gilt $n - r = 0$, ist $\text{rg} \mathbf{A} = n$ und $\mathbf{A}\boldsymbol{\beta}_0 = \mathbf{0}$ ergibt sich wegen (122.1) und (132.2) nur für $\boldsymbol{\beta}_0 = \mathbf{0}$. Die Lösung $\boldsymbol{\beta}$ in (154.7) läßt sich dann mit einer Linksinversen für \mathbf{A} aus (151.2) berechnen. Die Lösungen mit Hilfe verschiedener Linksinversen, beispielsweise $\boldsymbol{\beta}_1 = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{l}$ oder $\boldsymbol{\beta}_2 = (\mathbf{A}'\mathbf{C}\mathbf{A})^{-1}\mathbf{A}'\mathbf{C}\mathbf{l}$ mit $\text{rg}(\mathbf{A}'\mathbf{C}\mathbf{A}) = n$, stimmen sämtlich überein, denn wegen der Konsistenz des Gleichungssystems gibt es zu jedem \mathbf{l} einen Vektor \mathbf{w} , so daß $\mathbf{A}\mathbf{w} = \mathbf{l}$ gilt, womit durch Substitution $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \mathbf{w}$ folgt.

Wie im Kapitel 331 gezeigt wird, treten bei der Parameterschätzung in Modellen mit nicht vollem Rang symmetrische Gleichungssysteme der folgenden Gestalt auf. Für sie gilt der

Satz: Das Gleichungssystem

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y} , \quad (154.8)$$

in dem \mathbf{X} eine $n \times u$ Matrix mit $\text{rg} \mathbf{X} = q < u$, $\boldsymbol{\beta}$ ein $u \times 1$ Vektor unbekannter Parameter und \mathbf{y} ein gegebener $n \times 1$ Vektor bedeuten, ist immer lösbar. Seine allgemeine Lösung ist gegeben durch

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{y} + (\mathbf{I} - (\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X}) \mathbf{z} , \quad (154.9)$$

worin \mathbf{z} ein beliebiger $u \times 1$ Vektor bedeutet. Gilt $\text{rg} \mathbf{X} = u$, ergibt sich die eindeutige Lösung zu

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} . \quad (154.10)$$

Beweis: Nach (135.6) ist $R(\mathbf{X}'\mathbf{X}) = R(\mathbf{X}')$, so daß $\mathbf{X}'\mathbf{y} \in R(\mathbf{X}'\mathbf{X})$, womit nach (154.2) die Gleichung (154.8) konsistent und nach (154.3) immer lösbar

ist. (154.9) folgt aus (154.7). Weiter ist wegen (135.6) $\text{rg}\mathbf{X} = \text{rg}(\mathbf{X}'\mathbf{X})$, so daß mit $\text{rg}\mathbf{X} = u$ und (153.22) die eindeutige Lösung (154.10) aus (154.9) sich ergibt.

155 Generalisierte Inversen symmetrischer Matrizen

Es sollen nun Rechenformeln für die generalisierten Inversen symmetrischer Matrizen angegeben werden, um Lösungen nach (154.9) berechnen zu können. Hierbei wird unterschieden in Rechenformeln, die sich aufgrund der Definitionen des Kapitels 153 ergeben, und in solche, die mit Hilfe der Basis des Nullraums der Koeffizientenmatrix des Gleichungssystems abgeleitet werden.

a) Rechenformeln aufgrund der Definitionen

Es wird vorausgesetzt, daß die symmetrische $u \times u$ Matrix $\mathbf{X}'\mathbf{X} = \mathbf{N}$ mit $\text{rg}\mathbf{X} = \text{rg}\mathbf{N} = q < u$ durch Umordnen der Zeilen und Spalten sich derart in die Blockmatrix

$$\mathbf{X}'\mathbf{X} = \mathbf{N} = \begin{vmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{N}_{21} & \mathbf{N}_{22} \end{vmatrix} \quad (155.1)$$

zerlegen läßt, daß für die $q \times q$ Matrix \mathbf{N}_{11} gilt $\text{rg}\mathbf{N}_{11} = q$. Aufgrund der linearen Abhängigkeit von $u - q$ Spalten der Matrix \mathbf{N} gibt es wegen (122.3) eine $q \times (u - q)$ Matrix \mathbf{M} derart, daß

$$\mathbf{N}_{12} = \mathbf{N}_{11}\mathbf{M} \quad \text{und} \quad \mathbf{N}_{22} = \mathbf{N}_{21}\mathbf{M} \quad (155.2)$$

gilt und somit

$$\mathbf{M} = \mathbf{N}_{11}^{-1}\mathbf{N}_{12} \quad \text{und} \quad \mathbf{N}_{22} = \mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12} . \quad (155.3)$$

Eine generalisierte Inverse \mathbf{N}^- und zugleich eine symmetrische reflexive generalisierte Inverse \mathbf{N}_{rs}^- von \mathbf{N} sind gegeben durch

$$\mathbf{N}^- = \begin{vmatrix} \mathbf{N}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} = \mathbf{N}_{rs}^- , \quad (155.4)$$

denn (153.1) ist erfüllt mit

$$\mathbf{N}\mathbf{N}^-\mathbf{N} = \begin{vmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{N}_{21} & \mathbf{N}_{21}\mathbf{N}_{11}^{-1}\mathbf{N}_{12} \end{vmatrix} = \mathbf{N} \quad (155.5)$$

und weiter (153.11). Außerdem ist die reflexive generalisierte Inverse (155.4) symmetrisch.

Bezeichnet man mit \mathbf{Y} die Matrix der Eigenvektoren von \mathbf{N} , die der Zerlegung

von N in (155.1) entsprechend aufgespalten sei in $Y = |Y_1, Y_2|$, so folgt aus (142.6)

$$N = |Y_1, Y_2| \begin{vmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} \begin{vmatrix} Y_1' \\ Y_2' \end{vmatrix} = Y_1 \Lambda Y_1',$$

worin Λ die $q \times q$ Diagonalmatrix der von Null verschiedenen Eigenwerte von N bedeutet. Für die Pseudoinverse N^+ von N gilt dann

$$N^+ = Y_1 \Lambda^{-1} Y_1', \quad (155.6)$$

denn mit $Y_1' Y_1 = I$ und $NN^+ = N^+N = Y_1 Y_1'$ ist (153.15) erfüllt.

b) Rechenformeln mit Hilfe der Basis des Nullraums

Die $u \times u$ Matrix $X'X$ mit $\text{rg}X = \text{rg}X'X = q < u$ soll jetzt mit Hilfe einer Matrix B wie folgt zur Matrix D erweitert werden

$$D = \begin{vmatrix} X'X & B' \\ B & \mathbf{0} \end{vmatrix}. \quad (155.7)$$

Dies entspricht bei der Parameterschätzung, wie im Kapitel 333 gezeigt wird, der Einführung der Restriktionen

$$B\beta = \mathbf{0} \quad (155.8)$$

für den $u \times 1$ Parametervektor β im Gleichungssystem (154.8). Die Matrix B soll derart gewählt werden, daß D regulär wird, so daß β aus (154.8) eindeutig bestimmbar ist. Da $\text{rg}(X'X) = q$ gilt, müssen $u - q$ Restriktionen eingeführt werden, wie sich aus dem folgenden Satz ergibt.

Satz: Es gilt $\text{rg} \begin{vmatrix} X \\ B \end{vmatrix} = u$, worin B eine $(u - q) \times u$ Matrix bedeutet, genau dann, wenn D regulär ist. (155.9)

Beweis: Die Matrix X besitzt nach Voraussetzung q linear unabhängige Zeilen, die nach (124.4) und (124.5) um weitere $u - q$ linear unabhängige Zeilen, zusammengefaßt in der Matrix B , ergänzt werden können, so daß $\text{rg}B = u - q$ und $\text{rg} \begin{vmatrix} X \\ B \end{vmatrix} = u$ folgen. Da für den durch die Zeilen von X aufgespannten Vektorraum $R(X')$ nach (135.6) $R(X') = R(X'X)$ gilt, besitzt auch die Matrix $\begin{vmatrix} X'X \\ B \end{vmatrix}$ insgesamt u linear unabhängige Zeilen und damit nach (132.2) vollen Spaltenrang u . Die Spalten dieser Matrix bilden mit den $u - q$ Spalten der Matrix $\begin{vmatrix} B' \\ \mathbf{0} \end{vmatrix}$ insgesamt $2u - q$ linear unabhängige Spalten, denn

durch Linearkombination der Spalten von $\begin{vmatrix} \mathbf{X}'\mathbf{X} \\ \mathbf{B} \end{vmatrix}$ lassen sich zwar Vektoren erzeugen, deren unterste $u - q$ Komponenten gleich Null sind, die oberen u Komponenten bilden aber einen Vektor $\neq \mathbf{0}$ aus $R(\mathbf{X}'\mathbf{X})$, so daß die Spalten von $\begin{vmatrix} \mathbf{B}' \\ \mathbf{0} \end{vmatrix}$ nicht erzeugt werden können, folglich die Matrix \mathbf{D} den vollen Rang $2u - q$ besitzt und damit regulär ist. Ist andererseits \mathbf{D} regulär, sind nach (133.1) die Spalten von $\begin{vmatrix} \mathbf{X}'\mathbf{X} \\ \mathbf{B} \end{vmatrix}$ und somit von $\begin{vmatrix} \mathbf{X} \\ \mathbf{B} \end{vmatrix}$ linear unabhängig, so daß die Aussage sich ergibt.

Um die Inverse von \mathbf{D} angeben zu können, wird die folgende Matrix eingeführt. Aus den beiden Gleichungen in (155.2) $-\mathbf{N}_{11}\mathbf{M} + \mathbf{N}_{12} = \mathbf{0}$ und $-\mathbf{N}_{21}\mathbf{M} + \mathbf{N}_{22} = \mathbf{0}$ folgt, daß eine $(u - q) \times u$ Matrix \mathbf{E} mit

$$\mathbf{E}' = \begin{vmatrix} -\mathbf{N}_{11}^{-1}\mathbf{N}_{12} \\ \mathbf{I} \end{vmatrix} \quad (155.10)$$

derart existiert, daß

$$\mathbf{X}'\mathbf{X}\mathbf{E}' = \mathbf{0} \quad \text{und daher} \quad \mathbf{X}\mathbf{E}' = \mathbf{0}, \quad (155.11)$$

denn aus $\mathbf{X}'\mathbf{X}\mathbf{E}' = \mathbf{0}$ folgt $(\mathbf{X}\mathbf{E}')'\mathbf{X}\mathbf{E}' = \mathbf{0}$ und damit $\mathbf{X}\mathbf{E}' = \mathbf{0}$ wegen (131.11). Wie ein Vergleich mit (132.8) zeigt, besitzt \mathbf{E} vollen Zeilenrang, also $\text{rg}\mathbf{E} = \text{rg}\mathbf{E}' = u - q$. Die Spalten der Matrix \mathbf{E}' bilden daher wegen (135.5) und (155.11) eine Basis für den Nullraum $N(\mathbf{X})$ der Matrix \mathbf{X} und wegen (154.6) ebenfalls eine Basis für den Lösungsraum des aus (154.8) resultierenden homogenen Gleichungssystems. Die Basis läßt sich entweder aus (155.10) berechnen oder häufig unmittelbar angeben, da in ihr die Änderungen enthalten sind, die die Parameter in (154.8) vornehmen können, ohne daß sich die Absolutglieder des Gleichungssystems ändern. Beispiele hierzu befinden sich im Kapitel 333 und 343.

Für die Matrizen \mathbf{B} und \mathbf{E} erhält man den

Satz: Es gilt $\text{rg} \begin{vmatrix} \mathbf{X} \\ \mathbf{B} \end{vmatrix} = u$ genau dann, wenn die Matrix $\mathbf{B}\mathbf{E}'$ vollen Rang besitzt. (155.12)

Beweis: Besitzt die Matrix $\begin{vmatrix} \mathbf{X} \\ \mathbf{B} \end{vmatrix}$ vollen Spaltenrang, dann existiert nach (151.2) eine Linksinverse, die durch $[\mathbf{S}, \mathbf{U}]$ gegeben sei, so daß $\mathbf{S}\mathbf{X} + \mathbf{U}\mathbf{B} = \mathbf{I}_u$ gilt und weiter $\mathbf{S}\mathbf{X}\mathbf{E}' + \mathbf{U}\mathbf{B}\mathbf{E}' = \mathbf{E}'$ oder $\mathbf{U}\mathbf{B}\mathbf{E}' = \mathbf{E}'$ wegen (155.11). Mit (132.3) erhält man $\text{rg}(\mathbf{B}\mathbf{E}') \geq \text{rg}(\mathbf{U}\mathbf{B}\mathbf{E}') = \text{rg}\mathbf{E}' = u - q$. Da aber $\mathbf{B}\mathbf{E}'$ eine $(u - q) \times (u - q)$ Matrix ist, folgt $\text{rg}(\mathbf{B}\mathbf{E}') = u - q$. Besitzt andererseits $\mathbf{B}\mathbf{E}'$ vollen Rang und nimmt man an, daß $\text{rg} \begin{vmatrix} \mathbf{X} \\ \mathbf{B} \end{vmatrix} < u$ gilt, dann läßt

sich durch eine Linearkombination $\mathbf{h}'\mathbf{B} + \mathbf{k}'\mathbf{X}$ der Zeilen von \mathbf{B} und \mathbf{X} der Nullvektor erzeugen, also $\mathbf{h}'\mathbf{B} + \mathbf{k}'\mathbf{X} = \mathbf{0}$, wobei $\mathbf{h} \neq \mathbf{0}$ ist, da jetzt die Zeilen von \mathbf{B} als linear abhängig angenommen werden müssen. Hiermit folgt $\mathbf{h}'\mathbf{B}\mathbf{E}' = -\mathbf{k}'\mathbf{X}\mathbf{E}' = \mathbf{0}$ und $\mathbf{h} = \mathbf{0}$, da die Inverse von $\mathbf{B}\mathbf{E}'$ existiert. Dies führt aber auf einen Widerspruch, so daß $\text{rg} \begin{vmatrix} \mathbf{X} \\ \mathbf{B} \end{vmatrix} = u$ und damit die Aussage folgt.

Unter der Voraussetzung von (155.12) soll jetzt die Inverse von (155.7) berechnet werden. Setzt man

$$\begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{B}' \\ \mathbf{B} & \mathbf{0} \end{vmatrix}^{-1} = \begin{vmatrix} \mathbf{Q}_b & \mathbf{P}' \\ \mathbf{P} & \mathbf{R} \end{vmatrix}, \quad (155.13)$$

da nach (131.17) die Inverse einer symmetrischen Matrix symmetrisch ist, folgt durch Multiplikation von (155.7) mit der rechten Seite von (155.13)

$$\begin{aligned} \mathbf{X}'\mathbf{X}\mathbf{Q}_b + \mathbf{B}'\mathbf{P} &= \mathbf{I}, \quad \mathbf{X}'\mathbf{X}\mathbf{P}' + \mathbf{B}'\mathbf{R} = \mathbf{0}, \\ \mathbf{B}\mathbf{Q}_b &= \mathbf{0}, \quad \mathbf{B}\mathbf{P}' = \mathbf{I}. \end{aligned} \quad (155.14)$$

Mit $\mathbf{E}\mathbf{X}' = \mathbf{0}$ und $\text{rg}(\mathbf{E}\mathbf{B}') = u - q$ folgt aus der ersten Gleichung $\mathbf{P} = (\mathbf{E}\mathbf{B}')^{-1}\mathbf{E}$, die die vierte Gleichung erfüllt. Mit \mathbf{P} ergibt sich aus der zweiten Gleichung von (155.14) wegen (155.11) $\mathbf{B}'\mathbf{R} = \mathbf{0}$ und weiter $\mathbf{E}\mathbf{B}'\mathbf{R} = \mathbf{0}$ und somit $\mathbf{R} = \mathbf{0}$, da $\mathbf{E}\mathbf{B}'$ vollen Rang besitzt. Man erhält daher anstelle von (155.13)

$$\begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{B}' \\ \mathbf{B} & \mathbf{0} \end{vmatrix}^{-1} = \begin{vmatrix} \mathbf{Q}_b & \mathbf{E}'(\mathbf{B}\mathbf{E}')^{-1} \\ (\mathbf{E}\mathbf{B}')^{-1}\mathbf{E} & \mathbf{0} \end{vmatrix}. \quad (155.15)$$

Weiter ergibt sich aus der ersten Gleichung von (155.14)

$$\mathbf{X}'\mathbf{X}\mathbf{Q}_b = \mathbf{I} - \mathbf{B}'(\mathbf{E}\mathbf{B}')^{-1}\mathbf{E} \quad (155.16)$$

oder $(\mathbf{X}'\mathbf{X} + \mathbf{B}'\mathbf{B})(\mathbf{Q}_b + \mathbf{E}'(\mathbf{B}\mathbf{E}')^{-1}(\mathbf{E}\mathbf{B}')^{-1}\mathbf{E}) = \mathbf{I}$ wegen $\mathbf{B}\mathbf{Q}_b = \mathbf{0}$ und $\mathbf{X}\mathbf{E}' = \mathbf{0}$. Hieraus folgt, da $|\mathbf{X}', \mathbf{B}'| \begin{vmatrix} \mathbf{X} \\ \mathbf{B} \end{vmatrix}$ vollen Rang besitzt,

$$\mathbf{Q}_b = (\mathbf{X}'\mathbf{X} + \mathbf{B}'\mathbf{B})^{-1} - \mathbf{E}'(\mathbf{E}\mathbf{B}'\mathbf{B}\mathbf{E}')^{-1}\mathbf{E}. \quad (155.17)$$

(155.17) erfüllt $\mathbf{B}\mathbf{Q}_b = \mathbf{0}$, denn man erhält $\mathbf{B}\mathbf{Q}_b(\mathbf{X}'\mathbf{X} + \mathbf{B}'\mathbf{B}) = \mathbf{B} - \mathbf{B}\mathbf{E}'(\mathbf{B}\mathbf{E}')^{-1}(\mathbf{E}\mathbf{B}')^{-1}\mathbf{E}\mathbf{B}'\mathbf{B} = \mathbf{0}$.

Rechtsseitige Multiplikation von (155.16) mit $\mathbf{X}'\mathbf{X}$ ergibt $\mathbf{X}'\mathbf{X}\mathbf{Q}_b\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$ und aus linksseitiger Multiplikation mit \mathbf{Q}_b folgt $\mathbf{Q}_b\mathbf{X}'\mathbf{X}\mathbf{Q}_b = \mathbf{Q}_b$,

da $\mathbf{Q}_b \mathbf{B}' = \mathbf{0}$ gilt. Damit ist (153.11) erfüllt, und \mathbf{Q}_b ist eine symmetrische reflexive generalisierte Inverse von $\mathbf{X}'\mathbf{X}$,

$$\mathbf{Q}_b = (\mathbf{X}'\mathbf{X})_{rs}^- . \quad (155.18)$$

Die Inverse \mathbf{Q}_b ist abhängig von der Wahl von \mathbf{B} und daher nicht eindeutig.

Erfüllt die $(u - q) \times u$ Matrix \mathbf{C} (155.12), dann ist die zugehörige Matrix \mathbf{Q}_c ebenfalls eine symmetrische reflexive generalisierte Inverse von $\mathbf{X}'\mathbf{X}$. Zwischen \mathbf{Q}_b und \mathbf{Q}_c bestehen wegen (153.11) die Beziehungen

$$\mathbf{Q}_b = \mathbf{Q}_b \mathbf{X}'\mathbf{X} \mathbf{Q}_c \mathbf{X}'\mathbf{X} \mathbf{Q}_b \quad \text{und} \quad \mathbf{Q}_c = \mathbf{Q}_c \mathbf{X}'\mathbf{X} \mathbf{Q}_b \mathbf{X}'\mathbf{X} \mathbf{Q}_c \quad (155.19)$$

mit $\mathbf{Q}_b \mathbf{X}'\mathbf{X}$ und $\mathbf{Q}_c \mathbf{X}'\mathbf{X}$ aus (155.16), wobei für letzteren Ausdruck in (155.16) \mathbf{B} durch \mathbf{C} zu ersetzen ist.

Eine symmetrische reflexive generalisierte Inverse \mathbf{Q}_b läßt sich aus (155.15), (155.17) oder (155.19) berechnen. Um (155.15) anzuwenden, muß mit Hilfe einer Pivotisierung gearbeitet werden, oder Zeilen und Spalten sind zu vertauschen, wie bereits im Zusammenhang mit (133.11) erläutert wurde. Bei großen, schwach besetzten Matrizen ermittelt man zweckmäßig zunächst eine symmetrische reflexive generalisierte Inverse nach (155.4), da bei ihrer Berechnung die schwache Besetzung auszunutzen ist, und transformiert dann (155.4) mit (155.19) in die gewünschte symmetrische reflexive generalisierte Inverse. Benötigt man nur Teile dieser Inversen, braucht auch nicht die gesamte Inverse \mathbf{N}_{11}^{-1} in (155.4) berechnet zu werden (KOCH 1983C).

Anstelle von \mathbf{B} soll jetzt die Matrix \mathbf{E} eingeführt werden. Dies bedeutet, daß anstelle von (155.8) die Restriktionen

$$\mathbf{E}\boldsymbol{\beta} = \mathbf{0} \quad (155.20)$$

gewählt werden, was zulässig ist, da $\text{rg} \begin{vmatrix} \mathbf{X} \\ \mathbf{E} \end{vmatrix} = u$ wegen (124.2) und (155.11) gilt. Mit \mathbf{Q}_e anstelle von \mathbf{Q}_b erhält man aus (155.15)

$$\begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{E}' \\ \mathbf{E} & \mathbf{0} \end{vmatrix}^{-1} = \begin{vmatrix} \mathbf{Q}_e & \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1} \\ (\mathbf{E}\mathbf{E}')^{-1}\mathbf{E} & \mathbf{0} \end{vmatrix} \quad (155.21)$$

und anstelle von (155.16)

$$\mathbf{X}'\mathbf{X}\mathbf{Q}_e = \mathbf{I} - \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1}\mathbf{E} \quad (155.22)$$

sowie aus (155.17)

$$\mathbf{Q}_e = (\mathbf{X}'\mathbf{X} + \mathbf{E}'\mathbf{E})^{-1} - \mathbf{E}'(\mathbf{E}\mathbf{E}'\mathbf{E}\mathbf{E}')^{-1}\mathbf{E} . \quad (155.23)$$

Wie für Q_b gilt $X'XQ_eX'X = X'X$ und $Q_eX'XQ_e = Q_e$. Weiter sind $X'XQ_e$ und $Q_eX'X$ symmetrisch, so daß nach (153.15) Q_e die Pseudoinverse von $X'X$ ist,

$$Q_e = (X'X)^+ . \quad (155.24)$$

Entsprechend (155.19) gilt

$$Q_e = Q_eX'XQ_bX'XQ_e \quad (155.25)$$

mit $Q_eX'X$ aus (155.22).

Die Pseudoinverse Q_e läßt sich aus (155.21), (155.23) oder (155.25) berechnen, wobei das gleiche gilt, was zur Ermittlung der Inversen Q_b gesagt wurde.

Zur Berechnung von $X'X$ aus Q_e erhält man mit (153.19) $Q_e^+ = X'X$. Da nach (153.17) $\text{rg}X'X = \text{rg}Q_e$ und nach (155.14) $Q_eE' = \mathbf{0}$ gilt, bilden die Spalten der Matrix E' auch eine Basis für den Nullraum $N(Q_e)$, so daß zur Berechnung von Q_e^+ in (155.21) lediglich $X'X$ und Q_e zu vertauschen sind. Es folgt dann zum Beispiel aus (155.23)

$$X'X = (Q_e + E'E)^{-1} - E'(EE'EE')^{-1}E . \quad (155.26)$$

Eine im Vergleich zu (155.15) ergänzte Formel ergibt sich mit der positiv definiten $(u - q) \times (u - q)$ Matrix Σ

$$\begin{vmatrix} X'X & B' \\ B & -\Sigma \end{vmatrix}^{-1} = \begin{vmatrix} Q_b + E'(BE')^{-1}\Sigma(EB')^{-1}E & E'(BE')^{-1} \\ (EB')^{-1}E & \mathbf{0} \end{vmatrix} .$$

Ihre Richtigkeit folgt aus der Definition (131.12) einer inversen Matrix mit (155.11), (155.16) und $BQ_b = \mathbf{0}$ in (155.14). Mit (155.16) ergibt sich weiter $(X'X + B'\Sigma^{-1}B)(Q_b + E'(BE')^{-1}\Sigma(EB')^{-1}E) = I$ und hieraus

$$Q_b = (X'X + B'\Sigma^{-1}B)^{-1} - E'(BE')^{-1}\Sigma(EB')^{-1}E . \quad (155.27)$$

Wie aus Kapitel 32 zusammen mit (134.8) ersichtlich, lassen sich also die Restriktionen $B\beta = \mathbf{0}$ als Beobachtungen mit der Kovarianzmatrix Σ interpretieren, die beliebig zu wählen ist. Die symmetrische reflexive generalisierte Inverse Q_b folgt dann mit (155.27). Ersetzt man B durch E , ergibt sich die Pseudoinverse Q_e .

156 Eigenschaften der Pseudoinversen und einer speziellen symmetrischen reflexiven generalisierten Inversen

In der Menge der generalisierten Inversen zeichnet sich die Pseudoinverse dadurch aus, daß sie nach (153.17) eindeutig ist. Weiter ist sie für symmetrische

Matrizen wegen (153.18) ebenfalls symmetrisch. Außerdem besitzt sie die beiden folgenden, im Zusammenhang mit Parameterschätzungen wichtigen Eigenschaften.

Satz: In der Menge der symmetrischen reflexiven generalisierten Inversen symmetrischer Matrizen besitzt die Pseudoinverse minimale Spur. (156.1)

Beweis: Mit (155.22) und (155.25) erhält man unter Beachtung von (137.3) $\text{sp}(\mathbf{X}'\mathbf{X})^+ = \text{sp}(\mathbf{X}'\mathbf{X})_{rs}^- - \text{sp}[(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1}\mathbf{E}]$. Wegen (143.8) und (143.9) ist $(\mathbf{E}\mathbf{E}')^{-1}$ positiv definit, so daß nach (143.5) die Cholesky-Faktorisierung gilt, beispielsweise $(\mathbf{E}\mathbf{E}')^{-1} = \mathbf{G}\mathbf{G}'$. Dann ist $\text{sp}[(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{E}'(\mathbf{E}\mathbf{E}')^{-1}\mathbf{E}] = \text{sp}[\mathbf{G}'\mathbf{E}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{E}'\mathbf{G}] > 0$ nach (143.6), da die Matrix $\mathbf{G}'\mathbf{E}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{E}'\mathbf{G}$ wegen (153.14) und (143.7) positiv semidefinit ist. Folglich gilt $\text{sp}(\mathbf{X}'\mathbf{X})^+ < \text{sp}(\mathbf{X}'\mathbf{X})_{rs}^-$.

Satz: Das lineare Gleichungssystem $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ wird genau dann mit Hilfe der Pseudoinversen durch $\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^+ \mathbf{X}'\mathbf{y}$ gelöst, wenn $\boldsymbol{\beta}'\boldsymbol{\beta}$ unter allen Lösungen im Lösungsraum des Gleichungssystems minimal wird. (156.2)

Beweis: Mit der Pseudoinversen und mit der Matrix \mathbf{E}' aus (155.10), deren Spalten eine Basis des Nullraums der Matrix \mathbf{X} enthält, ergibt sich als allgemeine Lösung $\boldsymbol{\beta}$ im Lösungsraum von $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ mit (154.6) und (154.9) $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^+ \mathbf{X}'\mathbf{y} + \mathbf{E}'\mathbf{z}$, worin \mathbf{z} ein beliebiger $(u - q) \times 1$ Vektor ist. Der Vektor \mathbf{z} wird derart bestimmt, daß $\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\beta}}$ minimal wird. Dies entspricht, wie im Kapitel 323 gezeigt wird, der Anwendung der Methode der kleinsten Quadrate, so daß man mit (323.3) erhält $\mathbf{z} = -(\mathbf{E}\mathbf{E}')^{-1}\mathbf{E}(\mathbf{X}'\mathbf{X})^+ \mathbf{X}'\mathbf{y} = \mathbf{0}$ wegen $\mathbf{E}(\mathbf{X}'\mathbf{X})^+ = \mathbf{0}$. Damit ergibt sich das Minimum für $\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^+ \mathbf{X}'\mathbf{y}$. Ist andererseits $\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^+ \mathbf{X}'\mathbf{y}$ die Lösung des Gleichungssystems, dann ist $\boldsymbol{\beta}'\boldsymbol{\beta}$ minimal, denn aus der allgemeinen Lösung folgt $\tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}'\boldsymbol{\beta} + \mathbf{z}'\mathbf{E}\mathbf{E}'\mathbf{z}$. Der zweite Summand auf der rechten Seite ist für alle $\mathbf{z} \neq \mathbf{0}$ größer als Null, da $\mathbf{E}\mathbf{E}'$ wegen (143.8) positiv definit ist, so daß $\boldsymbol{\beta}'\boldsymbol{\beta} < \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{\beta}}$ gilt.

Wie schon durch (155.20) angedeutet wurde und wie mit (333.10) gezeigt wird, erhält man die Lösung des linearen Gleichungssystems $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ mit der Pseudoinversen auch durch Einführung der Restriktionen $\mathbf{E}\boldsymbol{\beta} = \mathbf{0}$.

Entsprechende Eigenschaften wie für die Pseudoinverse, beschränkt allerdings im Vergleich zu (156.1) auf eine Teilspur und im Vergleich zu (156.2) auf eine Teilnorm, lassen sich auch für eine symmetrische reflexive generalisierte Inverse zeigen, wenn sie mit Hilfe einer Matrix \mathbf{B} in (155.13) berechnet wird, die aus der Matrix \mathbf{E} in (155.10) dadurch entsteht, daß nur die der Teilspur oder Teilnorm zugehörigen Spalten übernommen und die restlichen Spalten mit Nullen aufgefüllt werden.

Satz: In der Menge der symmetrischen reflexiven generalisierten Inversen

symmetrischer Matrizen besitzt die symmetrische reflexive generalisierte Inverse $[(\mathbf{X}'\mathbf{X})_{rs}^-]_b$, die mit der Matrix \mathbf{B} berechnet wird, für die (155.12) gelte und die mit $\mathbf{B} = \mathbf{E}\mathbf{S}$ bestimmt sei, wobei \mathbf{S} eine Diagonalmatrix mit $\mathbf{S} = \text{diag}(0, \dots, 0, 1, \dots, 1, 0, \dots, 0)$ ist, deren Elemente Eins an beliebigen Stellen stehen können, minimale Teilspur und zwar

$$\text{sp}\{\mathbf{S}[(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{S}\} < \text{sp}\{\mathbf{S}(\mathbf{X}'\mathbf{X})_{rs}^-\mathbf{S}\}. \quad (156.3)$$

Beweis: Mit (155.16) und (155.19) erhält man $\text{sp}\{\mathbf{S}[(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{S}\} = \text{sp}\{\mathbf{S}(\mathbf{I} - \mathbf{E}'(\mathbf{B}\mathbf{E}')^{-1}\mathbf{B})(\mathbf{X}'\mathbf{X})_{rs}^-(\mathbf{I} - \mathbf{B}'(\mathbf{E}\mathbf{B}')^{-1}\mathbf{E})\mathbf{S}\}$, und wegen $\mathbf{B} = \mathbf{E}\mathbf{S}$, $\mathbf{S} = \mathbf{S}'$, $\mathbf{S}\mathbf{S} = \mathbf{S}$ folgt $\text{sp}\{\mathbf{S}[(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{S}\} = \text{sp}\{\mathbf{S}(\mathbf{X}'\mathbf{X})_{rs}^-\mathbf{S}\} - \text{sp}\{\mathbf{S}(\mathbf{X}'\mathbf{X})_{rs}^-\mathbf{S}\mathbf{S}\mathbf{E}'(\mathbf{E}\mathbf{S}\mathbf{E}')^{-1}\mathbf{E}\mathbf{S}\}$. Nach Voraussetzung besitzt \mathbf{B} vollen Zeilenrang, so daß $\mathbf{E}\mathbf{S}\mathbf{E}' = \mathbf{B}\mathbf{B}'$ positiv definit ist. Der Rest des Beweises läuft daher ab wie im Beweis zu (156.1).

Satz: Das lineare Gleichungssystem $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ wird genau dann mit Hilfe einer symmetrischen reflexiven generalisierten Inversen $[(\mathbf{X}'\mathbf{X})_{rs}^-]_b$ durch $\boldsymbol{\beta} = [(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{X}'\mathbf{y}$ gelöst, wenn $\boldsymbol{\beta}'\mathbf{S}'\mathbf{S}\boldsymbol{\beta}$ unter allen Lösungen im Lösungsraum des Gleichungssystems minimal wird, wobei $[(\mathbf{X}'\mathbf{X})_{rs}^-]_b$ und \mathbf{S} wie in (156.3) definiert sind. (156.4)

Beweis: Mit $[(\mathbf{X}'\mathbf{X})_{rs}^-]_b$ ergibt sich wie im Beweis von (156.2) als allgemeine Lösung $\bar{\boldsymbol{\beta}}$ im Lösungsraum des Gleichungssystems $\bar{\boldsymbol{\beta}} = [(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{X}'\mathbf{y} + \mathbf{E}'\mathbf{z}$ oder $\mathbf{S}\bar{\boldsymbol{\beta}} = \mathbf{S}[(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{X}'\mathbf{y} + \mathbf{S}\mathbf{E}'\mathbf{z}$. Der Vektor \mathbf{z} wird nun derart bestimmt, daß $\bar{\boldsymbol{\beta}}'\mathbf{S}'\mathbf{S}\bar{\boldsymbol{\beta}}$ minimal wird, so daß man $\mathbf{z} = -(\mathbf{E}\mathbf{S}'\mathbf{S}\mathbf{E}')^{-1}\mathbf{E}\mathbf{S}'\mathbf{S}[(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{X}'\mathbf{y} = \mathbf{0}$ wegen $\mathbf{B} = \mathbf{E}\mathbf{S}'\mathbf{S}$ und $\mathbf{B}[(\mathbf{X}'\mathbf{X})_{rs}^-]_b = \mathbf{0}$ erhält. Damit ergibt sich das Minimum für $\boldsymbol{\beta} = [(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{X}'\mathbf{y}$. Der zweite Teil des Beweises entspricht dem von (156.2).

Wie bereits mit (155.8) angedeutet und wie aus (333.9) ersichtlich, ergibt sich die Lösung des linearen Gleichungssystems $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ mit einer symmetrischen reflexiven generalisierten Inversen auch durch die Restriktionen $\mathbf{B}\boldsymbol{\beta} = \mathbf{0}$.

16 Projektionen

161 Allgemeine Projektionen

Mit Hilfe der Projektionen lassen sich, wie im Kapitel 323 gezeigt wird, die Methoden der Parameterschätzung geometrisch interpretieren.

Definition: Der Vektorraum V lasse sich als direkte Summe $V = V_1 \oplus V_2$ der Unterräume V_1 und V_2 ausdrücken, so daß nach (121.10) $\mathbf{x} \in V$ sich eindeutig

in $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ mit $\mathbf{x}_1 \in V_1$ und $\mathbf{x}_2 \in V_2$ zerlegen läßt. Die Transformation $\mathbf{R}\mathbf{x} = \mathbf{x}_1$ bezeichnet man dann als *Projektion* des Vektorraums V auf V_1 entlang V_2 und \mathbf{R} als *Projektionsoperator*. (161.1)

Der Projektionsoperator besitzt die folgenden Eigenschaften.

Satz: Eine Matrix \mathbf{R} ist genau dann Projektionsoperator, falls \mathbf{R} idempotent, also $\mathbf{R}^2 = \mathbf{R}$ ist. (161.2)

Beweis: Es sei $\mathbf{x}_1 = \mathbf{R}\mathbf{x}$ eine Projektion von $\mathbf{x} \in V$ auf $\mathbf{x}_1 \in V_1$. Eine weitere Projektion von $\mathbf{x}_1 \in V_1$ auf V_1 muß \mathbf{x}_1 ergeben, also $\mathbf{R}\mathbf{x}_1 = \mathbf{R}\mathbf{R}\mathbf{x} = \mathbf{R}\mathbf{x} = \mathbf{x}_1$ für alle $\mathbf{x} \in V$, woraus $\mathbf{R}^2 = \mathbf{R}$ folgt. Gilt andererseits $\mathbf{R}^2 = \mathbf{R}$, definiert man $V_1 = R(\mathbf{R})$ und $V_2 = R(\mathbf{I} - \mathbf{R})$, so daß jedes $\mathbf{x} \in V$ eindeutig durch $\mathbf{x} = \mathbf{R}\mathbf{x} + (\mathbf{I} - \mathbf{R})\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ dargestellt werden kann, V also in die direkte Summe $V = V_1 \oplus V_2$ zerfällt wegen $\mathbf{R}\mathbf{x}_2 = (\mathbf{R} - \mathbf{R}^2)\mathbf{x} = \mathbf{0}$. Mit $\mathbf{R}\mathbf{x} = \mathbf{R}\mathbf{x}_1 + \mathbf{R}\mathbf{x}_2 = \mathbf{R}^2\mathbf{x} + \mathbf{0} = \mathbf{R}\mathbf{x} = \mathbf{x}_1$ wird \mathbf{R} also Projektionsoperator von V auf $R(\mathbf{R})$ entlang $R(\mathbf{I} - \mathbf{R})$.

Satz: Ist \mathbf{R} der Projektionsoperator für die Projektion von V auf V_1 entlang V_2 , ist $\mathbf{I} - \mathbf{R}$ der Projektionsoperator für die Projektion von V auf V_2 entlang V_1 . (161.3)

Beweis: Nach (152.4) ist $\mathbf{I} - \mathbf{R}$ idempotent, falls \mathbf{R} idempotent ist, so daß $\mathbf{I} - \mathbf{R}$ Projektionsoperator ist. Definiert man wie im Beweis zu (161.2) $V_1 = R(\mathbf{R})$ und $V_2 = (\mathbf{I} - \mathbf{R})$, ergibt sich analog zu diesem Beweis die Aussage.

162 Orthogonale Projektionen

Von besonderem Interesse sind die Projektionen des Vektorraums E^n auf einen Unterraum U und sein orthogonales Komplement U^\perp . Nach (124.7) gilt $E^n = U \oplus U^\perp$ und die eindeutige Zerlegung $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ mit $\mathbf{x} \in E^n$, $\mathbf{x}_1 \in U$ und $\mathbf{x}_2 \in U^\perp$.

Definition: Der Vektorraum E^n werde durch einen Unterraum U und sein orthogonales Komplement U^\perp gebildet. Dann bezeichnet man die Transformation $\mathbf{R}\mathbf{x} = \mathbf{x}_1$ mit $\mathbf{x} \in E^n$ und $\mathbf{x}_1 \in U$ als *orthogonale Projektion* von E^n auf U entlang U^\perp und \mathbf{R} als *orthogonalen Projektionsoperator*. (162.1)

Satz: Die Matrix \mathbf{R} ist genau dann orthogonaler Projektionsoperator, wenn \mathbf{R} idempotent und symmetrisch ist. (162.2)

Beweis: Ist \mathbf{R} orthogonaler Projektionsoperator, gilt $\mathbf{R}\mathbf{x} = \mathbf{x}_1 \in U$ mit $\mathbf{x} \in E^n$ und $(\mathbf{I} - \mathbf{R})\mathbf{y} = \mathbf{y}_2 \in U^\perp$ mit $\mathbf{y} \in E^n$ wegen (161.3). Da die Vektoren in U und U^\perp zueinander orthogonal sind, gilt $\mathbf{x}'\mathbf{R}'(\mathbf{I} - \mathbf{R})\mathbf{y} = 0$ für alle $\mathbf{x}, \mathbf{y} \in E^n$, woraus $\mathbf{R}'(\mathbf{I} - \mathbf{R})\mathbf{y} = \mathbf{0}$ für alle $\mathbf{y} \in E^n$ und $\mathbf{R}'(\mathbf{I} - \mathbf{R}) = \mathbf{0}$

oder $\mathbf{R}' = \mathbf{R}'\mathbf{R}$ oder $\mathbf{R} = (\mathbf{R}'\mathbf{R})' = \mathbf{R}'\mathbf{R} = \mathbf{R}'$ folgt, so daß \mathbf{R} symmetrisch und idempotent sein muß. Andererseits erhält man aus $\mathbf{R}' = \mathbf{R}$ und $\mathbf{R}^2 = \mathbf{R}$ die Beziehung $\mathbf{R}'(\mathbf{I} - \mathbf{R}) = \mathbf{0}$ und damit $\mathbf{x}'\mathbf{R}'(\mathbf{I} - \mathbf{R})\mathbf{y} = 0$, so daß $R(\mathbf{R}) = U$ und $R(\mathbf{I} - \mathbf{R})$ zueinander orthogonal sind, also $R(\mathbf{I} - \mathbf{R}) = U^\perp$ gilt, so daß die Aussage folgt.

Für die orthogonale Projektion auf den Spaltenraum $R(\mathbf{A})$ einer $m \times n$ Matrix \mathbf{A} , der nach (135.2) Unterraum des E^m ist, gilt der

Satz: Ist \mathbf{A} eine $m \times n$ Matrix, dann ist der orthogonale Projektionsoperator für die orthogonale Projektion des E^m auf den Spaltenraum $R(\mathbf{A})$ beziehungsweise auf $R(\mathbf{A})^\perp$ gegeben durch

$$\mathbf{R} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' \quad \text{beziehungsweise durch} \quad \mathbf{I} - \mathbf{R},$$

so daß gilt

$$\mathbf{R}\mathbf{A} = \mathbf{A} \quad \text{beziehungsweise} \quad (\mathbf{I} - \mathbf{R})\mathbf{A} = \mathbf{0}. \quad (162.3)$$

Beweis: \mathbf{R} ist mit $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ wegen (153.5) idempotent und wegen (153.9) symmetrisch. Außerdem ist $R(\mathbf{R}) = R(\mathbf{A})$, da aus $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A} = \mathbf{R}\mathbf{A} = \mathbf{A}$ wegen (153.5) $R(\mathbf{A}) \subset R(\mathbf{R})$ und aus $\mathbf{R} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ die Beziehung $R(\mathbf{R}) \subset R(\mathbf{A})$ folgt. Die übrigen Aussagen ergeben sich mit (161.3).

Orthogonale Projektionsoperatoren lassen sich nicht nur bezüglich des mit (123.1) eingeführten Skalarproduktes $\mathbf{x}'\mathbf{y}$ definieren, sondern auch bezüglich des im Zusammenhang mit (141.8) verallgemeinerten Skalarproduktes $\mathbf{x}'\mathbf{V}\mathbf{y}$. Es gilt der

Satz: Die Matrix \mathbf{R} ist genau dann orthogonaler Projektionsoperator bezüglich des durch $\mathbf{x}'\mathbf{V}\mathbf{y}$ definierten Skalarproduktes, in dem \mathbf{V} eine positiv definite Matrix bedeutet, falls \mathbf{R} idempotent und $\mathbf{V}\mathbf{R}$ symmetrisch ist. (162.4)

Beweis: Der Beweis verläuft analog zu dem für (162.2). Aus $\mathbf{x}'\mathbf{R}'\mathbf{V}(\mathbf{I} - \mathbf{R})\mathbf{y} = 0$ für alle $\mathbf{x}, \mathbf{y} \in E^n$ folgt $\mathbf{R}'\mathbf{V}(\mathbf{I} - \mathbf{R}) = \mathbf{0}$ sowie $\mathbf{R}'\mathbf{V} = \mathbf{R}'\mathbf{V}\mathbf{R} = (\mathbf{R}'\mathbf{V}\mathbf{R})' = \mathbf{V}\mathbf{R}$ und somit $\mathbf{R}'\mathbf{V} = \mathbf{V}\mathbf{R}$ sowie $\mathbf{R}'\mathbf{V}\mathbf{R} = \mathbf{V}\mathbf{R}\mathbf{R}$ oder $\mathbf{V}\mathbf{R} = \mathbf{V}\mathbf{R}\mathbf{R}$, so daß $\mathbf{V}\mathbf{R}$ symmetrisch und \mathbf{R} idempotent ist.

Satz: Ist \mathbf{A} eine $m \times n$ Matrix, so ist der orthogonale Projektionsoperator für die orthogonale Projektion des E^m auf den Spaltenraum $R(\mathbf{A})$ beziehungsweise $R(\mathbf{A})^\perp$ bezüglich des durch $\mathbf{x}'\mathbf{V}\mathbf{y}$ definierten Skalarproduktes gegeben durch

$$\mathbf{R} = \mathbf{A}(\mathbf{A}'\mathbf{V}\mathbf{A})^{-1}\mathbf{A}'\mathbf{V} \quad \text{beziehungsweise durch} \quad \mathbf{I} - \mathbf{R},$$

so daß gilt

$$\mathbf{R}\mathbf{A} = \mathbf{A} \quad \text{sowie} \quad (\mathbf{I} - \mathbf{R})\mathbf{A} = \mathbf{0}. \quad (162.5)$$

Beweis: $\mathbf{R}^2 = \mathbf{R}$, $\mathbf{R}\mathbf{A} = \mathbf{A}$ und $(\mathbf{I} - \mathbf{R})\mathbf{A} = \mathbf{0}$ folgen aus (153.6). Die Symmetrie von $\mathbf{V}\mathbf{R}$ ergibt sich aus (153.10), da \mathbf{V} symmetrisch ist, und $R(\mathbf{R}) = R(\mathbf{A})$ erhält man mit den gleichen Überlegungen wie im Beweis zu (162.3).

17 Differentiation und Integration von Vektoren und Matrizen

171 Extrema von Funktionen

Im folgenden wird häufig das Problem auftreten, Extremwerte, also Maxima oder Minima von Funktionen von Vektoren oder Matrizen zu bestimmen. Diese Aufgabe soll daher kurz behandelt werden.

Definition: Die Funktion $f(\mathbf{x})$ des Vektors $\mathbf{x} \in E^n$ besitzt im Punkt $\mathbf{x}_0 \in E^n$ ein *Maximum* beziehungsweise *Minimum*, wenn $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ beziehungsweise $f(\mathbf{x}) \geq f(\mathbf{x}_0)$ für alle $\mathbf{x} \in E^n$ gilt. Sind diese Beziehungen nur in der Nachbarschaft von \mathbf{x}_0 erfüllt, liegen *lokale Extrema* vor.

Obere und untere Schranken einer Menge A bezeichnet man als *Supremum* und *Infimum* von A und schreibt $\sup A$ und $\inf A$. (171.1)

Es sollen nun die notwendigen Bedingungen für das Auftreten von Extrema angegeben werden. Hierzu müssen die Ableitungen von Funktionen eines Vektors oder einer Matrix definiert werden, was für eine Matrix geschehen soll.

Definition: Es sei $f(\mathbf{A})$ eine reelle, differenzierbare Funktion der $m \times n$ Matrix $\mathbf{A} = (a_{ij})$. Dann ist $\partial f(\mathbf{A})/\partial \mathbf{A}$ die $m \times n$ Matrix der partiellen Ableitungen von $f(\mathbf{A})$ nach den Elementen von \mathbf{A} , also $\partial f(\mathbf{A})/\partial \mathbf{A} = (\partial f(\mathbf{A})/\partial a_{ij})$. (171.2)

Als einfache Beispiele von Funktionen von Matrizen seien $f(\mathbf{A}) = \text{sp}\mathbf{A}$ oder $f(\mathbf{A}) = \det \mathbf{A}$ genannt. Für reelle differenzierbare Funktionen gilt folgendes (BLATTER 1974, II, S.168).

Satz: Es sei $f(\mathbf{x})$ eine reelle differenzierbare Funktion von $\mathbf{x} \in E^n$ und $\mathbf{x} + \Delta \mathbf{x} \in E^n$ mit $\Delta \mathbf{x} = |\Delta x_1, \Delta x_2, \dots, \Delta x_n|'$ ein Nachbarpunkt. Dann ergibt die *Taylor-Entwicklung* bei Vernachlässigung von Termen zweiter und höherer Potenzen von Δx_i

$$f(\mathbf{x} + \Delta \mathbf{x}) = f(\mathbf{x}) + \left| \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right|' \Delta \mathbf{x}. \quad (171.3)$$

Mit Hilfe der Taylor-Entwicklung kann nun der Satz über lokale Extrema bewiesen werden.

Satz: Besitzt die Funktion $f(\mathbf{x})$ im Punkt \mathbf{x}_0 ein lokales Extremum, so ist sie dort *stationär*, das heißt

$$\partial f(\mathbf{x})/\partial \mathbf{x}|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{0} . \quad (171.4)$$

Beweis: Damit \mathbf{x}_0 Punkt eines lokalen Maximums oder Minimums wird, muß in der Nachbarschaft von \mathbf{x}_0 die Differenz $f(\mathbf{x}) - f(\mathbf{x}_0)$ immer gleiches Vorzeichen aufweisen. Nach (171.3) ist dieses Vorzeichen, falls $\Delta \mathbf{x}$ klein genug gewählt wird, abhängig vom Vorzeichen von $|\partial f(\mathbf{x})/\partial \mathbf{x}' \Delta \mathbf{x}$, das seinerseits von dem von $\Delta \mathbf{x}$ abhängt, das sowohl positiv als auch negativ sein kann. Eine notwendige Bedingung für das Auftreten eines Extremwertes ist daher

$$|\partial f(\mathbf{x})/\partial \mathbf{x}' \Delta \mathbf{x} = \mathbf{0} , \quad (171.5)$$

womit die Aussage folgt.

Die Prüfung, ob ein lokaler Extremwert ein Maximum, Minimum oder Sattelpunkt darstellt, kann mit Hilfe der zweiten Ableitungen erfolgen. Diese Methode wird im folgenden nicht angewendet, da durch zusätzliche Überlegungen geprüft wird, ob nicht nur lokale, sondern absolute Maxima oder Minima vorliegen.

Zusätzlich wird das Problem der Extremwertbestimmung unter Restriktionen zu lösen sein. Es gilt der

Satz: Es sei $f(\mathbf{x})$ eine reelle differenzierbare Funktion des Vektors $\mathbf{x} \in E^n$. Weiter gelte $m < n$ und

$$g_i(\mathbf{x}) = 0 \quad \text{für } i \in \{1, \dots, m\} ,$$

wobei die Funktionen $g_i(\mathbf{x})$ reell und differenzierbar seien und die $m \times n$ Matrix $\mathbf{B} = (\partial g_i/\partial x_j)$ vollen Zeilenrang m besitze. Unter den Nebenbedingungen $g_i(\mathbf{x}) = 0$ besitze $f(\mathbf{x})$ ein lokales Extremum im Punkte \mathbf{x}_0 . Es existiert dann der $m \times 1$ Vektor $\mathbf{k} = (k_i)$ der *Lagrangeschen Multiplikatoren*, so daß die *Lagrangesche Funktion* $w(\mathbf{x})$ mit

$$w(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^m k_i g_i(\mathbf{x})$$

im Punkt \mathbf{x}_0 stationär wird, also $\partial w(\mathbf{x})/\partial \mathbf{x}|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{0}$. (171.6)

Beweis: An der Stelle des lokalen Extremums gilt mit (171.5), falls dort $\Delta \mathbf{x} = |dx_1, \dots, dx_n|'$ gesetzt wird,

$$\frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \dots + \frac{\partial f}{\partial x_n} dx_n = 0 . \quad (171.7)$$

Wegen der Nebenbedingungen $g_i(\mathbf{x}) = 0$ sind die dx_i nicht voneinander unabhängig, sondern man erhält durch Differentiation

$$\frac{\partial g_i}{\partial x_1} dx_1 + \frac{\partial g_i}{\partial x_2} dx_2 + \dots + \frac{\partial g_i}{\partial x_n} dx_n = 0 \quad \text{für } i \in \{1, \dots, m\}.$$

Multipliziert man die für $i = 1$ sich ergebende Gleichung mit dem zunächst unbestimmten Skalar k_1 , die für $i = 2$ mit k_2 und so fort und addiert sie zu (171.7), erhält man

$$\sum_{j=1}^n \left(\frac{\partial f}{\partial x_j} + k_1 \frac{\partial g_1}{\partial x_j} + k_2 \frac{\partial g_2}{\partial x_j} + \dots + k_m \frac{\partial g_m}{\partial x_j} \right) dx_j = 0. \quad (171.8)$$

Die m Gleichungen

$$\frac{\partial f}{\partial x_j} + k_1 \frac{\partial g_1}{\partial x_j} + k_2 \frac{\partial g_2}{\partial x_j} + \dots + k_m \frac{\partial g_m}{\partial x_j} = 0$$

für $j \in \{1, \dots, m\}$ (171.9)

bilden für k_1, k_2, \dots, k_m ein reguläres Gleichungssystem, da die Koeffizientenmatrix $(\partial g_i / \partial x_j)$ nach Voraussetzung vollen Rang besitzt. Die k_1, k_2, \dots, k_m lassen sich daher nach (133.15) eindeutig bestimmen und ergeben in (171.8) eingesetzt

$$\sum_{j=m+1}^n \left(\frac{\partial f}{\partial x_j} + k_1 \frac{\partial g_1}{\partial x_j} + k_2 \frac{\partial g_2}{\partial x_j} + \dots + k_m \frac{\partial g_m}{\partial x_j} \right) dx_j = 0.$$

Da die x_j für $j \in \{m+1, \dots, n\}$ voneinander unabhängig sind, erhält man

$$\frac{\partial f}{\partial x_j} + k_1 \frac{\partial g_1}{\partial x_j} + k_2 \frac{\partial g_2}{\partial x_j} + \dots + k_m \frac{\partial g_m}{\partial x_j} = 0$$

für $j \in \{m+1, \dots, n\}$. (171.10)

(171.9) zusammen mit (171.10) bedeutet $\partial w(\mathbf{x}) / \partial \mathbf{x} |_{\mathbf{x}=\mathbf{x}_0} = \mathbf{0}$, so daß die Aussage folgt.

172 Differentialquotienten spezieller Funktionen

Im folgenden sollen die Ableitungen häufig benutzter Funktionen von Vektoren und Matrizen angegeben werden.

Satz: Es sei $c = \mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$. Dann ist $\partial c / \partial \mathbf{x} = \mathbf{y}$. (172.1)

Beweis: Für $\mathbf{x}, \mathbf{y} \in E^n$ ist nach (123.1) $c = \sum_{i=1}^n x_i y_i$ und somit $\partial c / \partial x_i = y_i$

und nach (171.2) $\partial c/\partial \mathbf{x} = \mathbf{y}$.

Satz: Es sei \mathbf{x} ein $n \times 1$ Vektor und \mathbf{A} eine symmetrische $n \times n$ Matrix. Dann ist $\partial(\mathbf{x}'\mathbf{A}\mathbf{x})/\partial \mathbf{x} = 2\mathbf{A}\mathbf{x}$. (172.2)

Beweis: Mit $\mathbf{x} = (x_i)$, $\mathbf{A} = (a_{ij})$, $\mathbf{A}\mathbf{x} = (c_i)$ und (131.6) erhält man, da $\mathbf{A} = \mathbf{A}'$ gilt,

$$\begin{aligned} \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_i} &= \frac{\partial}{\partial x_i} \sum_{j=1}^n \sum_{k=1}^n x_j x_k a_{jk} \\ &= \frac{\partial}{\partial x_i} \left(\sum_{j=1}^n x_j^2 a_{jj} + \sum_{\substack{j=1 \\ j \neq k}}^n \sum_{k=1}^n x_j x_k a_{jk} \right) \\ &= 2x_i a_{ii} + 2 \sum_{\substack{k=1 \\ i \neq k}}^n x_k a_{ik} = 2 \sum_{k=1}^n x_k a_{ik} = 2c_i . \end{aligned}$$

Satz: Ist die $n \times n$ Matrix \mathbf{A} regulär, gilt

$$\partial \det \mathbf{A} / \partial \mathbf{A} = (\mathbf{A}^{-1})' \det \mathbf{A} .$$

Ist \mathbf{A} außerdem symmetrisch, erhält man

$$\partial \det \mathbf{A} / \partial \mathbf{A} = (2\mathbf{A}^{-1} - \text{diag} \mathbf{A}^{-1}) \det \mathbf{A} ,$$

wobei $\text{diag} \mathbf{A}^{-1}$ die Diagonalmatrix bedeutet, die die Diagonalelemente von \mathbf{A}^{-1} enthält. (172.3)

Beweis: Mit $\mathbf{A} = (a_{ij})$ und (136.5) gilt $\det \mathbf{A} = \sum_i (-1)^{i+j} a_{ij} \det \mathbf{A}_{ij}$ für festes j und mit (136.15) $\partial \det \mathbf{A} / \partial a_{ij} = a_{ij}^*$ oder $\partial \det \mathbf{A} / \partial \mathbf{A} = \bar{\mathbf{A}}'$ für eine Matrix \mathbf{A} mit funktional unabhängigen Elementen. Für symmetrisches \mathbf{A} erhält man mit (136.5) für festes i und j

$$\begin{aligned} \det \mathbf{A} &= (-1)^{1+j} a_{1j} \det \mathbf{A}_{1j} + (-1)^{2+j} a_{2j} \det \mathbf{A}_{2j} + \dots \\ &\quad + (-1)^{i+j} a_{ij} \det \mathbf{A}_{ij} + \dots \\ &= (-1)^{1+j} a_{1j} \sum_{\substack{k \\ k \neq 1}}^n (-1)^{k+i} a_{ki} \det (\mathbf{A}_{1j})_{ki} + (-1)^{2+j} a_{2j} \sum_{\substack{k \\ k \neq 2}}^n (-1)^{k+i} \\ &\quad a_{ki} \det (\mathbf{A}_{2j})_{ki} + \dots + (-1)^{i+j} a_{ij} \sum_{\substack{k \\ k \neq i}}^n (-1)^{k+i} a_{ki} \det (\mathbf{A}_{ij})_{ki} + \dots \end{aligned}$$

Bei der Differentiation nach dem Element a_{ij} für $i \neq j$ ist $a_{ij} = a_{ji}$ zu berücksichtigen, also

$$\begin{aligned} \partial \det \mathbf{A} / \partial a_{ij} &= (-1)^{i+j} \det \mathbf{A}_{ij} + (-1)^{1+j} a_{1j} [(-1)^{j+i} \det(\mathbf{A}_{1j})_{ji}] + \\ &\quad (-1)^{2+j} a_{2j} [(-1)^{j+i} \det(\mathbf{A}_{2j})_{ji}] + \dots \\ &\quad + (-1)^{i+j} a_{ij} [(-1)^{j+i} \det(\mathbf{A}_{ij})_{ji}] + \dots \\ &= (-1)^{i+j} \det \mathbf{A}_{ij} + (-1)^{j+i} \sum_{\substack{k \\ k \neq j}} (-1)^{k+j} a_{kj} \det(\mathbf{A}_{ji})_{kj} \\ &= (-1)^{i+j} \det \mathbf{A}_{ij} + (-1)^{j+i} \det \mathbf{A}_{ji} = a_{ij}^* + a_{ji}^* . \end{aligned}$$

Somit erhält man

$$\begin{aligned} \partial \det \mathbf{A} / \partial a_{ij} &= a_{ij}^* + a_{ji}^* \quad \text{für } i \neq j \quad \text{und} \\ \partial \det \mathbf{A} / \partial a_{ij} &= a_{ii}^* \quad \text{für } i = j . \end{aligned}$$

Daraus folgt

$$\partial \det \mathbf{A} / \partial \mathbf{A} = 2\bar{\mathbf{A}} - \text{diag} \bar{\mathbf{A}}$$

und mit (136.16) die Aussage.

Satz: Es seien \mathbf{A} eine $m \times n$ und \mathbf{B} eine $n \times m$ Matrix. Dann gilt

$$\partial \text{sp}(\mathbf{AB}) / \partial \mathbf{A} = \mathbf{B}' .$$

Falls \mathbf{A} und \mathbf{B} quadratisch sind und \mathbf{A} außerdem symmetrisch ist, erhält man

$$\partial \text{sp}(\mathbf{AB}) / \partial \mathbf{A} = \mathbf{B} + \mathbf{B}' - \text{diag} \mathbf{B} . \quad (172.4)$$

Beweis: Mit (131.6) und (137.1) folgt

$$\text{sp}(\mathbf{AB}) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ji} \quad \text{und} \quad \partial \text{sp}(\mathbf{AB}) / \partial a_{ij} = b_{ji} .$$

Sind \mathbf{A} und \mathbf{B} quadratisch und ist \mathbf{A} symmetrisch, gilt

$$\begin{aligned} \partial \text{sp}(\mathbf{AB}) / \partial a_{ij} &= b_{ji} + b_{ij} \quad \text{für } i \neq j \quad \text{und} \\ \partial \text{sp}(\mathbf{AB}) / \partial a_{ij} &= b_{ii} \quad \text{für } i = j , \end{aligned}$$

so daß die Aussage folgt.

Satz: Es seien \mathbf{A} eine $m \times n$ Matrix und \mathbf{B} sowie \mathbf{C} zwei $n \times m$ Matrizen. Dann gilt

$$\partial \text{sp}(\mathbf{ABAC}) / \partial \mathbf{A} = (\mathbf{BAC} + \mathbf{CAB})' . \quad (172.5)$$

Beweis: Es sei $\mathbf{D} = \mathbf{AB}$, $\mathbf{E} = \mathbf{AC}$, $\mathbf{F} = \mathbf{DE} = \mathbf{ABAC}$, $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$ und so fort. Dann gilt

$$f_{ij} = \sum_{k=1}^m d_{ik} e_{kj} \quad \text{mit} \quad d_{ik} = \sum_{l=1}^n a_{il} b_{lk} \quad \text{und} \quad e_{kj} = \sum_{r=1}^n a_{kr} c_{rj} ,$$

so daß

$$f_{ij} = \sum_{k=1}^m \sum_{l=1}^n \sum_{r=1}^n a_{il} b_{lk} a_{kr} c_{rj} \quad \text{für} \quad i, j \in \{1, \dots, m\}$$

und

$$\text{sp} \mathbf{F} = \sum_{i=1}^m \sum_{k=1}^m \sum_{l=1}^n \sum_{r=1}^n a_{il} b_{lk} a_{kr} c_{ri} .$$

Damit folgt

$$\begin{aligned} \partial \text{sp} \mathbf{F} / \partial a_{op} &= \partial \text{sp} \mathbf{F} / \partial a_{il} + \partial \text{sp} \mathbf{F} / \partial a_{kr} \\ &= \sum_{k=1}^m \sum_{r=1}^n b_{lk} a_{kr} c_{ri} + \sum_{i=1}^m \sum_{l=1}^n a_{il} b_{lk} c_{ri} \\ &= \sum_{k=1}^m b_{lk} g_{ki} + \sum_{i=1}^m c_{ri} h_{ik} = u_{li} + v_{rk} = u_{po} + v_{po} , \end{aligned}$$

falls $\mathbf{G} = \mathbf{AC}$, $\mathbf{H} = \mathbf{AB}$, $\mathbf{U} = \mathbf{BAC}$ und $\mathbf{V} = \mathbf{CAB}$. Hieraus folgt dann die Aussage.

Satz: Sind \mathbf{A} , \mathbf{B} und \mathbf{C} quadratische Matrizen, gilt

$$\partial \text{sp}(\mathbf{ABA}'\mathbf{C}) / \partial \mathbf{A} = (\mathbf{BA}'\mathbf{C})' + \mathbf{CAB} . \quad (172.6)$$

Beweis: Mit $\mathbf{F} = \mathbf{ABA}'\mathbf{C}$ erhält man

$$\text{sp} \mathbf{F} = \sum_i \sum_l \sum_k \sum_r a_{il} b_{lk} a_{rk} c_{ri} .$$

Dann folgt

$$\begin{aligned} \partial \text{sp} \mathbf{F} / \partial a_{op} &= \partial \text{sp} \mathbf{F} / \partial a_{il} + \partial \text{sp} \mathbf{F} / \partial a_{kr} \\ &= \sum_k \sum_r b_{lk} a_{rk} c_{ri} + \sum_i \sum_l a_{il} b_{lr} c_{ki} \end{aligned}$$

und damit die Aussage.

173 Integration und Variablentransformation

Bei der Berechnung von Wahrscheinlichkeiten aus Wahrscheinlichkeitsverteilungen muß über Funktionen von Vektoren oder Matrizen integriert werden. Der Integrand $f(\mathbf{x})$ sei eine reelle Funktion des Vektors \mathbf{x} mit $\mathbf{x} = |x_1, x_2, \dots, x_n|'$, dann ist das Integral I über den Bereich B definiert als

$$I = \int \dots \int_B f(\mathbf{x}) dx_1 \dots dx_n . \quad (173.1)$$

Für die Herleitung von Wahrscheinlichkeitsverteilungen werden häufig Variablentransformationen erforderlich, so daß (173.1) in ein Integral neuer Variablen transformiert werde. Hierzu sei die injektive, also eineindeutige Abbildung mittels $x_i = g_i(\mathbf{y})$ mit $i \in \{1, \dots, n\}$ und $\mathbf{y} = |y_1, \dots, y_n|'$ gegeben. Die Funktionen g_i werden als einmal stetig differenzierbar vorausgesetzt. Es existiert dann die *Jacobische Matrix* \mathbf{J} ,

$$\mathbf{J} = \begin{vmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} & \dots & \frac{\partial g_1}{\partial y_n} \\ \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} & \dots & \frac{\partial g_2}{\partial y_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_n}{\partial y_1} & \frac{\partial g_n}{\partial y_2} & \dots & \frac{\partial g_n}{\partial y_n} \end{vmatrix} , \quad (173.2)$$

deren Determinante, die *Funktionaldeterminante* oder *Jacobische Determinante*, von Null verschieden ist, also $\det \mathbf{J} \neq 0$. Die Umkehrabbildung $y_i = h_i(\mathbf{x})$ ist dann eindeutig. Wird durch $x_i = g_i(\mathbf{y})$ der Bereich S auf den Bereich B abgebildet, so ist das Integral I in den neuen Variablen gegeben durch (BLATTER 1974, III, S.83)

$$\begin{aligned} & \int \dots \int_B f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int \dots \int_S f(g_1(\mathbf{y}), \dots, g_n(\mathbf{y})) |\det \mathbf{J}| dy_1 \dots dy_n , \end{aligned} \quad (173.3)$$

worin $|\det \mathbf{J}|$ den Absolutbetrag von $\det \mathbf{J}$ bedeutet.

Es sollen noch die Funktionaldeterminanten $\det \mathbf{J}$ für zwei Transformationen angegeben werden.

Satz: Für die Transformation des $n \times 1$ Variablenvektors \mathbf{x} in den $n \times 1$ Variablenvektor \mathbf{y} mittels $\mathbf{y} = \mathbf{A}^{-1}(\mathbf{x} - \mathbf{c})$, wobei die $n \times n$ Matrix \mathbf{A} von Konstanten regulär sei und der $n \times 1$ Vektor \mathbf{c} Konstanten enthalte, gilt

$$\det \mathbf{J} = \det \mathbf{A} . \quad (173.4)$$

Beweis: Aus $\mathbf{y} = \mathbf{A}^{-1}(\mathbf{x} - \mathbf{c})$ folgt $\mathbf{x} - \mathbf{c} = \mathbf{A}\mathbf{y}$ und mit (173.2) die Aussage, da die Komponenten von $\mathbf{x} - \mathbf{c}$ sich aus dem Skalarprodukt der Zeilen von \mathbf{A} und \mathbf{y} ergeben, so daß (172.1) anwendbar ist.

Satz: Für die Transformation der symmetrischen $n \times n$ Matrix \mathbf{Q} von Variablen in die symmetrische $n \times n$ Matrix \mathbf{V} von Variablen mittels $\mathbf{V} = \mathbf{G}^{-1}\mathbf{Q}(\mathbf{G}')^{-1}$, wobei \mathbf{G} eine reguläre untere Dreiecksmatrix bedeutet, gilt

$$\det \mathbf{J} = (\det \mathbf{G})^{n+1}. \quad (173.5)$$

Beweis: Aus $\mathbf{V} = \mathbf{G}^{-1}\mathbf{Q}(\mathbf{G}')^{-1}$ folgt $\mathbf{Q} = \mathbf{G}\mathbf{V}\mathbf{G}'$ und mit $\mathbf{Q} = (q_{ij})$, $\mathbf{G} = (g_{ij})$ und $\mathbf{V} = (v_{ij})$ aus (131.6) $q_{ij} = \sum_{l=1}^n (\sum_{k=1}^n g_{ik}v_{kl})g_{jl}$. Hieraus ergeben sich die Differentialquotienten

$$\frac{\partial q_{ij}}{\partial v_{kl}} = g_{ik}g_{jl} + g_{il}g_{jk} \quad \text{für } k \neq l,$$

da $v_{kl} = v_{lk}$ ist und

$$\frac{\partial q_{ij}}{\partial v_{kk}} = g_{ik}g_{jk} \quad \text{für } k = l.$$

Die Zeilen der Jacobischen Matrix \mathbf{J} in (173.2) enthalten die Elemente $\partial q_{ij}/\partial v_{11}, \partial q_{ij}/\partial v_{12}, \dots, \partial q_{ij}/\partial v_{1n}, \partial q_{ij}/\partial v_{22}, \dots, \partial q_{ij}/\partial v_{nn}$. Da \mathbf{G} eine untere Dreiecksmatrix darstellt, ist auch \mathbf{J} , wie sich aus den Differentialquotienten ergibt, eine untere Dreiecksmatrix. Ihre Determinante berechnet sich nach (136.10) aus dem Produkt der Diagonalelemente

$$\begin{aligned} \det \mathbf{J} &= \frac{\partial q_{11}}{\partial v_{11}} \frac{\partial q_{12}}{\partial v_{12}} \cdots \frac{\partial q_{1n}}{\partial v_{1n}} \frac{\partial q_{22}}{\partial v_{22}} \cdots \frac{\partial q_{nn}}{\partial v_{nn}} \\ &= g_{11}^2 g_{11} g_{22} \cdots g_{11} g_{nn} g_{22}^2 \cdots g_{nn}^2 = \prod_{i=1}^n g_{ii}^{n+1} \\ &= (\det \mathbf{G})^{n+1}. \end{aligned}$$

2 Wahrscheinlichkeitstheorie

Die Beobachtungen, mit deren Hilfe Parameter sowie Bereiche für die Parameter geschätzt und Hypothesenprüfungen vorgenommen werden, ergeben sich als Ergebnisse von Zufallsexperimenten und stellen somit zufällige Ereignisse dar. Sind aber die Beobachtungen zufälliger Natur, so erhebt sich die Frage, mit welcher Wahrscheinlichkeit sie eintreffen. Mit diesem Problem beschäftigt sich die Wahrscheinlichkeitstheorie, auf deren Grundbegriffe im folgenden eingegangen wird.

Den Zufallsereignissen werden zunächst durch Axiome Wahrscheinlichkeiten zugeordnet. Da es bequemer ist, mit Zufallsvariablen anstelle von Zufallsereignissen zu arbeiten, werden Zufallsvariable als Abbildungen von Ereignissen auf die reelle x -Achse eingeführt. Genügt die Abbildung auf eine Achse, liegt eine eindimensionale Zufallsvariable mit einer univariaten Verteilung vor, aus der die Wahrscheinlichkeit zufälliger Ereignisse zu ermitteln ist. Als Abbildungen auf mehrere Achsen folgen die mehrdimensionalen Zufallsvariablen mit multivariaten Verteilungen. Behandelt werden die univariate und multivariate Normalverteilung sowie die univariate Beta- und Gammaverteilung. Ferner werden die aus der Normalverteilung folgende χ^2 -, F - und t -Verteilung sowie die Wishart-Verteilung abgeleitet, die für Hypothesentests und Bereichsschätzungen in univariaten und multivariaten Modellen der Parameterschätzung benötigt werden.

21 Wahrscheinlichkeit

211 Einführung

Gegenstand der Wahrscheinlichkeitstheorie bilden die zufälligen Ereignisse. Diese Zufallsereignisse sind das Ergebnis von Messungen oder Experimenten, die zur Sammlung von Daten über ein bestimmtes Phänomen ausgeführt werden, wobei die Resultate vom Zufall abhängen. Das Würfeln beispielsweise ist ein solches Experiment, dessen Ergebnis die Zahl darstellt, die oben auf dem Würfel erscheint. Dieses Ergebnis ist vom Zufall abhängig, denn falls der Würfel symmetrisch oder ausbalanciert ist, werden bei mehrfachem Würfeln die geworfenen Zahlen variieren.

Da die Ergebnisse von Messungen und Experimenten zufällig sind, möchte man die Wahrscheinlichkeit angeben, mit der die Ereignisse eintreffen. Bei

bestimmten Experimenten ist die Wahrscheinlichkeit vorweg angebbbar. Zum Beispiel wird beim Werfen einer symmetrischen Münze nach vielen Wiederholungen die „Zahl“ ebenso häufig oben liegen wie der „Adler“ oder anders ausgedrückt, in einem von zwei möglichen Fällen wird die Zahl oben liegen, so daß die Wahrscheinlichkeit des Eintreffens einer Zahl gleich $1/2$ gesetzt werden kann. Dann beträgt auch die Wahrscheinlichkeit des Eintreffens des Adlers $1/2$. Beim Werfen eines symmetrischen Würfels wird zum Beispiel die Drei in einem von sechs möglichen Fällen oben liegen, so daß die Wahrscheinlichkeit des Würfels der Drei sich zu $1/6$ ergibt. Aufgrund dieser Überlegungen folgt die

Definition (Klassische Definition der Wahrscheinlichkeit): Falls bei einem Experiment n sich gegenseitig ausschließende und gleichmögliche Ergebnisse erzielt werden können und falls die Anzahl n_A dieser Ergebnisse mit dem Ereignis A verbunden ist, dann ist die *Wahrscheinlichkeit* $P(A)$ des Ereignisses A gegeben durch

$$P(A) = n_A/n . \quad (211.1)$$

Beispiel: Die Wahrscheinlichkeit P ist gesucht, mit der die Zahlen Eins oder Zwei bei einmaligem Würfeln fallen. Man erhält $P = 2/6 = 0,333$.

Aufgrund von (211.1) ergibt sich als Wahrscheinlichkeit $P(A)$ eines Ereignisses A eine rationale Zahl zwischen Null und Eins, $0 \leq P(A) \leq 1$. Diese Wahrscheinlichkeit kann aber nur für Experimente angegeben werden, deren Ereignisse vorhersagbar sind. Die klassische Wahrscheinlichkeitsdefinition versagt beispielsweise beim Würfeln mit unsymmetrischen Würfeln. Die Definition muß daher ergänzt werden, was mit Hilfe der relativen Häufigkeit eines Ereignisses geschehen kann.

Definition: Die *relative Häufigkeit* $h(A)$ eines Ereignisses A ergibt sich mit der Anzahl n_A des Eintreffens des Ereignisses A unter n Versuchen zu

$$h(A) = n_A/n . \quad (211.2)$$

1. Beispiel: Bei 300 Würfeln mit einem symmetrischen Würfel wurde die Zahl Eins 48 mal und die Zahl Zwei 51 mal geworfen, so daß sich die relative Häufigkeit h des Würfels einer Eins oder Zwei zu $h = (48 + 51)/300 = 0,330$ ergibt.

2. Beispiel: Für den Grauwert g eines Bildelementes, auch Pixel genannt, eines digitalen Bildes gilt $0 \leq g \leq 255$. Besitzen in einem digitalen Bild mit $512 \cdot 512$ Pixeln insgesamt 100 Pixel den Grauwert $g = 0$, beträgt seine relative Häufigkeit $h(g = 0) = 100/512^2$. Die Verteilung der relativen Häufigkeiten der Grauwerte $g = 0, g = 1, \dots, g = 255$ bezeichnet man als *Histogramm*.

Man nimmt nun an, und Versuche wie die des ersten Beispiels unterstützen die Hypothese, daß sich die relative Häufigkeit nach (211.2) bei einer großen Anzahl von Wiederholungen von Experimenten, deren Ereignisse vorhersagbar sind, der nach (211.1) definierten Wahrscheinlichkeit annähert, so daß die Wahrscheinlichkeit eines Ereignisses mit Hilfe der relativen Häufigkeit unter Vorschrift eines Grenzprozesses, bei dem die Anzahl der Versuche gegen unendlich geht, definiert werden könnte. Man verzichtet aber auf eine solche Definition und führt die Wahrscheinlichkeit durch Axiome ein, wie im Kapitel 213 gezeigt wird. Die Axiome stellen Aussagen über unmittelbar einzusehende Wahrheiten dar, die, als richtig akzeptiert, nicht bewiesen zu werden brauchen. Zum besseren Verständnis der Axiome soll noch darauf hingewiesen werden, daß sich die Wahrscheinlichkeit beziehungsweise die relative Häufigkeit mehrerer sich gegenseitig ausschließender Ereignisse aus der Summe der Wahrscheinlichkeiten beziehungsweise der relativen Häufigkeiten der einzelnen Ereignisse ergeben, wie das Beispiel zu (211.1) und das erste Beispiel zu (211.2) zeigen. Bevor aber die Axiome der Wahrscheinlichkeit eingeführt werden, müssen die zufälligen Ereignisse definiert werden.

212 Zufällige Ereignisse

Das Ergebnis eines bestimmten Experimentes bezeichnet man als *Elementarereignis* und faßt sämtliche vorstellbaren Elementarereignisse in der Menge S der Elementarereignisse zusammen, die auch als *Ergebnisraum* oder *Grundraum* bezeichnet wird.

Beispiel: Beim Würfeln ergeben sich als Elementarereignisse die Zahlen 1 bis 6, so daß die Menge S der Elementarereignisse sechs Elemente enthält.

Teilmengen von S bezeichnet man als zufällige Ereignisse und die Menge aller Teilmengen von S als Menge Z der zufälligen Ereignisse. Da die leere Menge \emptyset Teilmenge jeder Menge ist, gehört sie zur Menge aller Teilmengen, so daß die Menge Z der zufälligen Ereignisse auch die leere Menge \emptyset enthält, die als unmögliches Ereignis bezeichnet wird. Schließlich ist auch die Menge S selbst, die als sicheres Ereignis bezeichnet wird, Element der Menge Z der zufälligen Ereignisse.

Beispiel: Die Menge Z der zufälligen Ereignisse beim Würfeln enthält das unmögliche Ereignis, daß keine Zahl zwischen 1 und 6 gewürfelt wird, die Ereignisse der Zahlen 1 bis 6, die die Elemente der Menge S der Elementarereignisse darstellen, die Ereignisse der Zahlen 1 oder 2, 1 oder 3 und so fort, die Ereignisse 1 oder 2 oder 3, 1 oder 2 oder 4 und so fort und schließlich das sichere Ereignis 1 oder 2 oder ... oder 6, somit $Z = \{\emptyset, \{1\}, \{2\}, \dots, \{6\}, \{1, 2\}, \{1, 3\}, \dots, \{1, 2, \dots, 6\}\}$.

Für die Teilmengen von S , die die Elemente von Z bilden, lassen sich die im Kapitel 112 definierten Mengenverknüpfungen vornehmen. Da im folgenden nicht nur die Mengen S von endlich vielen oder abzählbar unendlich vielen Elementarereignissen betrachtet werden, sondern auch die Mengen S von Elementarereignissen, die ein Kontinuum bilden, wird Z als spezielles System von Teilmengen von S eingeführt, das als σ -Algebra bezeichnet wird (HINDERER 1972, S.75). Eine σ -Algebra Z besitzt die Eigenschaft, daß sie als Elemente die Teilmengen von S und als Element die leere Menge \emptyset sowie als Element die Menge S der Elementarereignisse enthält. Gehören weiter endlich viele oder abzählbar unendlich viele Elemente zu Z , gehören auch die Vereinigung und der Durchschnitt dieser Elemente zu Z . Gleiches gilt für die Differenz zweier Elemente. Z enthält dann im allgemeinen nicht mehr alle Teilmengen von S , wie das bei endlich vielen Elementarereignissen der Fall ist. Die Elemente von Z sind meßbare Mengen, für die die im folgenden Kapitel definierten Wahrscheinlichkeitsmaße eingeführt werden können.

Definition: Die Elemente einer σ -Algebra Z von Teilmengen einer bestimmten Menge S von Elementarereignissen bezeichnet man als *zufällige Ereignisse*.
(212.1)

Definition: Es sei $A \in Z$ und $B \in Z$, dann sagt man, daß die Ereignisse A und B sich *gegenseitig ausschließen*, wenn sie kein Elementarereignis gemeinsam haben, wenn also $A \cap B = \emptyset$ gilt.
(212.2)

Definition: Das Ereignis, das kein Element der Menge S der Elementarereignisse enthält, das Ereignis also, das durch die leere Menge \emptyset gekennzeichnet ist, bezeichnet man als *unmögliches Ereignis*.
(212.3)

Definition: Das Ereignis, das alle Elemente von S enthält, bezeichnet man als *sicheres Ereignis*.
(212.4)

Denkt man sich die Menge S der Elementarereignisse durch die Vereinigung der sich gegenseitig ausschließenden Elementarereignisse gebildet, zum Beispiel beim Würfeln $S = \{1, 2, \dots, 6\} = \{1\} \cup \{2\} \cup \dots \cup \{6\}$, so läßt sich das sichere Ereignis dadurch interpretieren, daß zumindest ein Elementarereignis eintritt.

Ist $A \in Z$, so gilt $A \subset S$ und wegen $S \in Z$ bildet auch die Menge $S \setminus A$ nach (212.1) ein Ereignis. In Übereinstimmung mit (112.4) ergibt sich dann die

Definition: Es sei $A \in Z$, dann bezeichnet die Differenzmenge $S \setminus A$ das zu A *komplementäre* Ereignis \bar{A} in S , $\bar{A} = S \setminus A$.
(212.5)

213 Axiome der Wahrscheinlichkeit

Für jedes zufällige Ereignis A der σ -Algebra Z von Teilmengen der Menge S der Elementarereignisse wird jetzt die Wahrscheinlichkeit $P(A)$ eingeführt, wobei die folgenden Axiome gelten.

Axiom 1: Jedem Ereignis A von Z ist eine reelle Zahl $P(A) \geq 0$ zugeordnet, die die *Wahrscheinlichkeit* von A heißt. (213.1)

Axiom 2: Die Wahrscheinlichkeit des sicheren Ereignisses ist gleich Eins, $P(S) = 1$. (213.2)

Axiom 3: Ist A_1, A_2, \dots eine Folge von endlich vielen oder abzählbar unendlich vielen Ereignissen von Z , die sich gegenseitig ausschließen, $A_i \cap A_j = \emptyset$ für $i \neq j$, dann gilt

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots \quad (213.3)$$

Das Tripel (S, Z, P) heißt dann *Wahrscheinlichkeitsraum*.

Aus den Axiomen lassen sich eine Reihe von Sätzen ableiten, auf die im folgenden häufig zurückgegriffen wird.

Satz: Die Wahrscheinlichkeit, daß das Ereignis $A \in Z$ nicht eintritt, oder die Wahrscheinlichkeit $P(\bar{A})$ des zu A komplementären Ereignisses \bar{A} beträgt $P(\bar{A}) = 1 - P(A)$. (213.4)

Beweis: Mit (212.5) ergibt sich $A \cap \bar{A} = \emptyset$ und $A \cup \bar{A} = S$, so daß mit (213.2) und (213.3) $P(A) + P(\bar{A}) = 1$ folgt, woraus die Aussage erhalten wird.

Satz: Die Wahrscheinlichkeit des unmöglichen Ereignisses ist gleich Null. (213.5)

Beweis: Nach (212.3) ist das unmögliche Ereignis durch $A = \emptyset$ definiert. Weiter gilt $\bar{S} = \emptyset, S \cup \bar{S} = S, S \cap \bar{S} = \emptyset$ und somit wegen (213.3) $P(S \cup \bar{S}) = P(S) + P(\emptyset) = 1$, woraus mit (213.2) $P(\emptyset) = 0$ folgt.

Satz: Ist A ein beliebiges Ereignis von Z , gilt

$$0 \leq P(A) \leq 1. \quad (213.6)$$

Beweis: Nach (213.1) ist $P(A) \geq 0$, so daß noch $P(A) \leq 1$ gezeigt werden muß, was aus (213.4) mit $P(A) = 1 - P(\bar{A}) \leq 1$ wegen $P(\bar{A}) \geq 0$ folgt.

Das dritte Axiom (213.3) gibt nur die Wahrscheinlichkeit der Vereinigung sich gegenseitig ausschließender Ereignisse an. Für die Vereinigung beliebiger Ereignisse gilt der

Satz: Sind A und B zwei beliebige Ereignisse von Z , dann gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \quad (213.7)$$

Beweis: Das Ereignis $A \cup B$ läßt sich durch die Vereinigung der drei sich gegenseitig ausschließenden Ereignisse $A \cap \bar{B}$, $A \cap B$ und $\bar{A} \cap B$ gewinnen, wie im Venn-Diagramm der Abbildung 213-1 dargestellt ist. Mit (213.3) folgt dann $P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B)$. Entsprechend gilt $P(A) = P(A \cap \bar{B}) + P(A \cap B)$ sowie $P(B) = P(A \cap B) + P(\bar{A} \cap B)$ und nach Addition $P(A) + P(B) - 2P(A \cap B) = P(A \cap \bar{B}) + P(\bar{A} \cap B)$. Diesen Ausdruck in die erste Gleichung eingesetzt ergibt die Aussage.

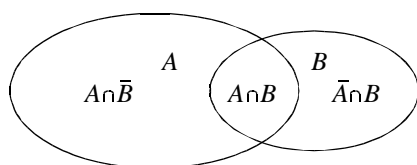


Abb. 213-1: Vereinigung beliebiger Ereignisse

Beispiel: Entnimmt man eine Karte einem Kartenspiel von 52 Spielkarten und fragt nach der Wahrscheinlichkeit, daß sie ein As oder Karo ist, dann erhält man aus (213.7), da die Wahrscheinlichkeit $P(A)$ für das Ziehen eines Asses nach (211.1) $P(A) = 4/52$, die Wahrscheinlichkeit $P(B)$ für das Ziehen einer Karo-Karte $P(B) = 13/52$ und für das Ziehen eines Karo-Asses $P(A \cap B) = 1/52$ beträgt, die Wahrscheinlichkeit $P(A \cup B) = 4/52 + 13/52 - 1/52 = 4/13$.

214 Bedingte Wahrscheinlichkeit und Bayessche Formel

Es gibt Situationen, in denen nach der Wahrscheinlichkeit eines Ereignisses unter der Bedingung gefragt wird, daß ein anderes Ereignis bereits eingetroffen ist. Als Beispiel seien Textuntersuchungen genannt, in denen die Häufigkeit von Buchstaben- oder Lautkombinationen untersucht wird, so daß die Wahrscheinlichkeit eines Buchstabens unter der Bedingung gesucht wird, daß bestimmte Buchstaben vorangegangen sind. Man benutzt das Symbol $A|B$ und sagt A unter der Bedingung, daß B eingetroffen ist, um das bedingte Eintreffen von A auszudrücken.

Man stelle sich vor, daß bei k Wiederholungen eines Versuches l Ereignisse B erzielt werden, unter denen m Ereignisse A mit $m \leq l$ erhalten wurden, beispielsweise bei Textuntersuchungen l Buchstaben e , gefolgt von m Buchstaben i . Die relative Häufigkeit $h(A \cap B)$ des Ereignisses $A \cap B$ beträgt nach (211.2)

m/k , während die relative Häufigkeit $h(B)$ des Ereignisses B sich zu l/k ergibt. Die relative Häufigkeit $h(A|B)$ des Ereignisses A unter der Bedingung, daß B eingetreten ist, beträgt m/l , folglich

$$h(A|B) = \frac{m}{l} = \frac{m}{k} / \frac{l}{k} = \frac{h(A \cap B)}{h(B)} .$$

Entsprechend wird die bedingte Wahrscheinlichkeit eingeführt.

Definition: Die Wahrscheinlichkeit $P(B)$ des zufälligen Ereignisses B sei ungleich Null, dann bezeichnet man als *bedingte Wahrscheinlichkeit* $P(A|B)$ des Ereignisses A unter der Bedingung, daß B eingetroffen ist, das Verhältnis

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{mit} \quad P(B) > 0 . \quad (214.1)$$

Aus (214.1) folgt unmittelbar die Wahrscheinlichkeit des Durchschnitts $A \cap B$ der Ereignisse A und B

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A) . \quad (214.2)$$

Beispiel: Ein Kasten enthalte 15 rote und 5 schwarze Kugeln. Gefragt wird nach der Wahrscheinlichkeit, in zwei aufeinanderfolgenden Ziehungen ohne Zurücklegen eine rote und eine schwarze Kugel zu erhalten. Die Wahrscheinlichkeit $P(A)$, eine rote Kugel zu ziehen, ist nach (211.1) $P(A) = 15/20 = 3/4$. Die Wahrscheinlichkeit $P(B|A)$ eine schwarze Kugel unter der Bedingung zu ziehen, daß eine rote gezogen wurde, beträgt $P(B|A) = 5/19$. Die Wahrscheinlichkeit, in zwei Ziehungen ohne Zurücklegen eine rote und eine schwarze Kugel zu ziehen, beträgt daher nach (214.2) $P(A \cap B) = (3/4)(5/19) = 15/76$.

Die im folgenden abgeleitete Bayessche Formel ermittelt aus der gegebenen *Priori-Wahrscheinlichkeit* $P(A_i)$ eines Ereignisses A_i die *Posteriori-Wahrscheinlichkeit* $P(A_i|B)$ von A_i , die sich durch das Eintreffen eines Ereignisses B ergibt. Das der Bayesschen Formel entsprechende Bayes-Theorem, das für Verteilungen gilt, bildet den Ausgangspunkt für die Bayes-Statistik, deren praktische Anwendung sich immer weiter ausbreitet (z.B. BOSSLER 1972; BROEMELING 1985; KOCH und RIESMEIER 1985; KOCH 1990; KOCH und SCHMIDT 1994; KOCH 2000).

Satz: Für die zufälligen Ereignisse A_1, A_2, \dots, A_n im Wahrscheinlichkeitsraum (S, Z, P) gelte $A_i \cap A_j = \emptyset$ für $i \neq j$, $A_1 \cup A_2 \cup \dots \cup A_n = S$ und $P(A_i) > 0$, dann ergibt sich für jedes beliebige Ereignis $B \in Z$ mit $P(B) > 0$ die *Bayessche Formel*

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^n P(A_j)P(B|A_j)} . \quad (214.3)$$

Beweis: Den zweiten Ausdruck von (214.2) in (214.1) substituiert ergibt $P(A_i|B) = P(A_i)P(B|A_i)/P(B)$. Da $A_i \cap A_j = \emptyset$ für $i \neq j$ und $A_1 \cup \dots \cup A_n = S$ gelten, folgt $B = (B \cap A_1) \cup \dots \cup (B \cap A_n)$ und wegen $(B \cap A_i) \cap (B \cap A_j) = \emptyset$ für $i \neq j$ mit (213.3) $P(B) = \sum_{j=1}^n P(B \cap A_j)$, woraus mit (214.2) die Aussage folgt.

215 Unabhängige Ereignisse

Wenn die bedingte Wahrscheinlichkeit $P(A|B)$ nicht vom Eintreffen des Ereignisses B abhängt, gilt die

Definition: Die Ereignisse A und B sind voneinander *unabhängig*, falls gilt

$$P(A|B) = P(A) \quad \text{oder} \quad P(B|A) = P(B) . \quad (215.1)$$

In diesem Fall hat also das Eintreffen des Ereignisses B keinerlei Einfluß auf die bedingte Wahrscheinlichkeit des Ereignisses A und umgekehrt. Beispielsweise führt das zweimalige Würfeln auf zwei unabhängige Ereignisse, da die Wahrscheinlichkeit des Ergebnisses des zweiten Wurfes unter der Bedingung, daß der erste Wurf ein bestimmtes Ergebnis aufweist, nicht von diesem Ergebnis abhängt.

Mit (215.1) vereinfacht sich (214.2) zu

$$P(A \cap B) = P(A)P(B) . \quad (215.2)$$

22 Zufallsvariable

221 Definition

In den vorangegangenen Beispielen des Werfens einer Münze waren es die Begriffe Zahl und Adler, die die Elementarereignisse bildeten. Im allgemeinen ist es aber vorteilhafter, den Elementarereignissen reelle Zahlen zuzuordnen. Hierzu wird eine Funktion eingeführt, die die Menge S der Elementarereignisse auf die Menge \mathbb{R} der reellen Zahlen abbildet.

Definition: Man bezeichnet eine eindeutige reellwertige Funktion $X(s_i)$, die im Wahrscheinlichkeitsraum (S, Z, P) auf der Menge S der Elementarereignisse s_i definiert ist, als *Zufallsvariable*, falls für jedes beliebige $x \in \mathbb{R}$ das Ereignis, für das $X(s_i) < x$ gilt, zu den zufälligen Ereignissen von Z gehört. (221.1)

Beispiel: Beim zweimaligen Werfen einer Münze besteht die Menge S der Elementarereignisse aus den vier Elementen $s_1 = AA, s_2 = AB, s_3 = BA, s_4 = BB$, falls A das Erscheinen des Adlers und B das der Zahl bedeutet. Zufällige

Ereignisse sind beispielsweise das Werfen keines oder eines Adlers. Als Zufallsvariable $X(s_i)$ läßt sich also die Summe der Adler einführen, so daß man erhält $X(s_1) = 2, X(s_2) = 1, X(s_3) = 1$ und $X(s_4) = 0$. Für $x = 2$ bezeichnet dann $X(s_i) < 2$ das zufällige Ereignis, bei dem kein Adler oder ein Adler geworfen wird.

Die Zufallsvariable bildet also die Menge S der Elementarereignisse zusammen mit Z auf die Menge \mathbb{R} der reellen Zahlen mit der σ -Algebra $Z_{\mathbb{R}}$ der sogenannten Borelschen Mengen ab, die durch die Intervalle $(-\infty, x)$ für beliebiges $x \in \mathbb{R}$ gebildet werden (FISZ 1976, S.49). Für das Urbild X^{-1} eines Intervalles wird gefordert

$$X^{-1}(-\infty, x) \in Z \quad \text{für jedes beliebige } x \in \mathbb{R}. \quad (221.2)$$

Die Zufallsvariable heißt dann meßbar (HINDERER 1972, S.98). Weiter ist das Urbild der Vereinigung, des Durchschnitts oder der Differenz von Intervallen gleich der Vereinigung, dem Durchschnitt oder der Differenz der Urbilder der Intervalle und damit gleich der Vereinigung, dem Durchschnitt oder der Differenz der entsprechenden Elemente von Z , so daß die Intervalle wie zufällige Ereignisse zu behandeln sind.

Die Wahrscheinlichkeit für das zufällige Ereignis, für das $X(s_i) < x$ gilt, ist wegen (221.2) definiert durch

$$P^*(X(s_i) < x) = P(X^{-1}(-\infty, x)). \quad (221.3)$$

Damit bildet die Zufallsvariable den Wahrscheinlichkeitsraum (S, Z, P) auf $(\mathbb{R}, Z_{\mathbb{R}}, P^*)$ ab, für den im folgenden die Wahrscheinlichkeit wieder mit P bezeichnet wird, also

$$P(X < x) = P^*(X(s_i) < x). \quad (221.4)$$

Bei einem Experiment, dessen Ergebnisse durch reelle Zahlen oder durch ein Meßinstrument mit digitaler Anzeige registriert werden, erfolgt die Abbildung der Menge der Elementarereignisse auf die Menge der reellen Zahlen durch die digitale Aufzeichnung. Eine Zufallsvariable, die auf diese Weise definiert wird, bezeichnet man als *Messung* oder *Beobachtung*.

Mit (221.1) wird die *eindimensionale Zufallsvariable* definiert. Sind anstelle einer Funktion mehrere Funktionen zur Abbildung erforderlich, ergeben sich die *mehrdimensionalen Zufallsvariablen*, die im Kapitel 225 behandelt werden.

222 Verteilungsfunktion

Aufgrund von (221.3) und (221.4) wird jetzt die Verteilungsfunktion einer Zufallsvariablen eingeführt.

Definition: Es sei $X(s_i)$ eine Zufallsvariable und $P(X < x)$ die Wahrscheinlichkeit des zufälligen Ereignisses $X(s_i) < x$, dann bezeichnet man die durch

$$F(x) = P(X < x) = P(X^{-1}(-\infty, x)) \quad \text{für alle } x \in \mathbb{R}$$

definierte Funktion $F(x)$ als *Verteilungsfunktion* der Zufallsvariablen X .

(222.1)

Beispiel: Ein Experiment bestehe im Werfen dreier symmetrischer Münzen, wobei die Anzahl der oben liegenden Adler die Ereignisse bilden, deren Wahrscheinlichkeit zu berechnen ist. Die Zufallsvariable X bezeichne daher diese Anzahl von Adlern. Die Gesamtanzahl der Möglichkeiten, mit der drei Münzen fallen können, beträgt 2^3 . Die Anzahl der Möglichkeiten, daß x Adler mit $x \in \{0, 1, 2, 3\}$ oben liegen, berechnet sich nach (136.2) zu $\binom{3}{x}$. Somit folgt wegen (211.1)

$$P(X = x) = \binom{3}{x} / 2^3 \quad \text{für } x \in \{0, 1, 2, 3\}$$

oder

$$P(0) = 1/8, P(1) = 3/8, P(2) = 3/8, P(3) = 1/8.$$

Weiter ergibt sich aus (222.1) $F(0) = P(X < 0) = 0$ wegen (213.5), $F(1) = P(X < 1) = P(X = 0) = 1/8$ und mit (213.3), da die Elementarereignisse sich gegenseitig ausschließen, $F(2) = P(X < 2) = P(X = 0) + P(X = 1) = 4/8$, $F(3) = P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 7/8$ und schließlich $F(4) = P(X < 4) = 1$, so daß die in Abbildung 222-1 dargestellte Treppenfunktion für die Verteilungsfunktion erhalten wird.

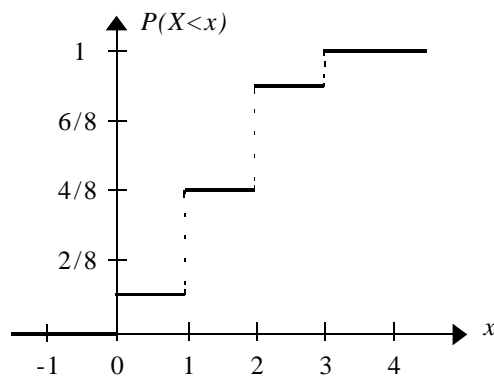


Abb. 222-1: Verteilungsfunktion

In dem Beispiel zeigen sich drei wichtige Eigenschaften der Verteilungsfunktion, die in dem folgenden Satz genannt sind.

Satz: Ist $F(x)$ Verteilungsfunktion der Zufallsvariablen X , dann gilt mit $x_1 < x_2$

$$F(x_1) \leq F(x_2) \quad \text{und} \quad F(-\infty) = 0, \quad F(\infty) = 1. \quad (222.2)$$

Beweis: Da wegen $x_1 < x_2$ das Intervall $(-\infty, x_2)$ das Intervall $(-\infty, x_1)$ enthält, die Ereignisse, für die $-\infty < X < x_1$ und $x_1 \leq X < x_2$ gelten, sich aber gegenseitig ausschließen, gilt mit (213.1) und (213.3) $P(X < x_2) \geq P(X < x_1)$, so daß mit (222.1) $F(x_2) \geq F(x_1)$ folgt. Weiter entspricht $X < -\infty$ dem unmöglichen Ereignis und $X < \infty$ dem sicheren Ereignis, so daß mit (213.5) und (213.2) die beiden restlichen Aussagen folgen.

223 Diskrete und stetige Zufallsvariable

Eine diskrete Zufallsvariable X nimmt endlich viele oder abzählbar unendlich viele Werte x_j an. Die Wahrscheinlichkeit $f(x_j)$ eines solchen Wertes ergibt sich mit (221.3) als Differenz zweier Intervalle zu $P(X(s_i) = x_j) = P(X^{-1}(x_j))$, und es folgt die

Definition: Man bezeichnet X als *diskrete Zufallsvariable*, falls sie lediglich endlich viele oder abzählbar unendlich viele Werte annimmt. Sind $x_1, x_2, \dots, x_n, \dots$ diese Werte und $f(x_1), f(x_2), \dots, f(x_n), \dots$ ihre Wahrscheinlichkeiten, nennt man $f(x_i)$ die *Dichte* oder die *Verteilung* von X . (223.1)

Anstelle der kurzen Bezeichnung Dichte oder Verteilung benutzt man auch *Dichtefunktion*, *Wahrscheinlichkeitsdichte* oder *Wahrscheinlichkeitsverteilung*.

Damit die Funktion $f(x_i)$ der Werte x_i einer diskreten Zufallsvariablen X die Wahrscheinlichkeit angibt, ist es notwendig, daß $f(x_i)$ die Axiome (213.1) bis (213.3) erfüllt. Da die Werte x_i von X sich gegenseitig ausschließende Ereignisse angeben, erhält man mit (213.3) aus (213.1) und (213.2) für die Dichte $f(x_i)$ die Bedingungen

$$f(x_i) \geq 0 \quad \text{und} \quad \sum_{i=1}^n f(x_i) = 1 \quad \text{oder} \quad \sum_{i=1}^{\infty} f(x_i) = 1, \quad (223.2)$$

falls X insgesamt n Werte x_i oder falls X abzählbar unendlich viele Werte x_i annimmt.

Für das Ereignis $X < x_i$ erhält man mit (223.1) wegen (213.3)

$$P(X < x_i) = \sum_{j < i} f(x_j). \quad (223.3)$$

Hieraus folgt nach (222.1) für die Verteilungsfunktion einer diskreten Zufallsvariablen

$$P(X < x_i) = F(x_i) = \sum_{j < i} f(x_j) . \quad (223.4)$$

Die Verteilungsfunktion einer diskreten Zufallsvariablen bildet also eine Treppenfunktion, die in Abbildung 222-1 für ein Beispiel dargestellt wurde.

Im Gegensatz zur Verteilungsfunktion einer diskreten Zufallsvariablen besitzt die Verteilungsfunktion einer stetigen Zufallsvariablen keine Sprungstellen, sondern ist stetig.

Definition: Man bezeichnet X als *stetige Zufallsvariable*, falls eine nicht-negative integrierbare Funktion $f(x)$ existiert, die für beliebiges $x \in \mathbb{R}$ die Beziehung

$$F(x) = \int_{-\infty}^x f(t) dt$$

erfüllt, wobei $F(x)$ die Verteilungsfunktion von X und t eine Integrationsvariable bedeuten. Die Funktion $f(x)$ nennt man Dichte oder *univariate Verteilung* von X . (223.5)

In Abbildung 223-1 sind die Verteilungsfunktion $F(x)$ und die Dichte $f(x)$ einer stetigen Zufallsvariablen X dargestellt.

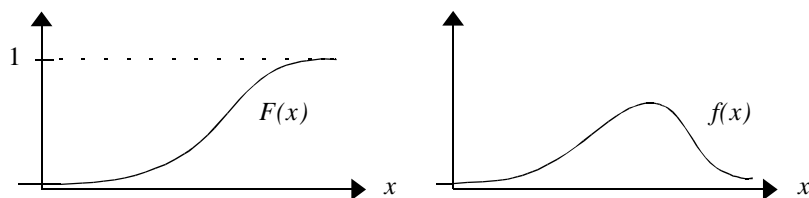


Abb. 223-1: Verteilungs- und Dichtefunktion

Die Wahrscheinlichkeit des Ereignisses $X < x$ ergibt sich aus (223.5) zu

$$P(X < x) = F(x) = \int_{-\infty}^x f(t) dt . \quad (223.6)$$

Anstelle der Summation (223.4) über die Dichte zur Berechnung der Wahrscheinlichkeit eines zufälligen Ereignisses für eine diskrete Zufallsvariable tritt

also bei einer stetigen Zufallsvariablen die Integration (223.6) über die Dichte. Für das Ereignis $a \leq X < b$ erhält man

$$P(a \leq X < b) = F(b) - F(a) = \int_a^b f(t) dt . \quad (223.7)$$

Da $P(a \leq X < b) = P(a < X < b)$ gilt, wird im folgenden mit offenen Intervallen gearbeitet. Für das Intervall $x < X < x + dx$ erhält man

$$P(x < X < x + dx) = f(x) dx . \quad (223.8)$$

Ist $f(x)$ im Punkte x stetig, ergibt sich aus der Ableitung des Integrals auf der rechten Seite von (223.6) nach seiner oberen Grenze die Beziehung zwischen Verteilungsfunktion und Dichte einer stetigen Zufallsvariablen zu

$$dF(x)/dx = f(x) . \quad (223.9)$$

Damit die integrierbare Funktion $f(x)$ der Werte x einer stetigen Zufallsvariablen X die Dichte angibt, ist es notwendig, daß die aus (223.5) sich ergebende Verteilungsfunktion $F(x)$ die Axiome (213.1) bis (213.3) erfüllt. Mit $f(x) \geq 0$ sind (213.1) und (213.3) erfüllt, da sich gegenseitig ausschließende Ereignisse auf elementenfremde Intervalle abgebildet werden. Weiter ist mit $F(\infty) = 1$ aus (222.2) das 2. Axiom erfüllt, so daß für $f(x)$ die Bedingungen erhalten werden

$$f(x) \geq 0 \quad \text{und} \quad \int_{-\infty}^{\infty} f(x) dx = 1 . \quad (223.10)$$

Beispiele für univariate Verteilungen stetiger Zufallsvariablen werden im Kapitel 24 behandelt, während diskrete Verteilungen im nächsten Kapitel folgen.

224 Binomialverteilung und Poisson-Verteilung

Von den diskreten Verteilungen ist die Binomialverteilung die wichtigste. Sie gibt die Wahrscheinlichkeit an, daß bei n voneinander unabhängigen Wiederholungen eines Experimentes, wobei das einzelne Experiment nur in einem Erfolg oder Mißerfolg enden kann und der Erfolg die Wahrscheinlichkeit p aufweist, x Erfolge eintreffen.

Definition: Die Zufallsvariable X besitzt die *Binomialverteilung*, falls ihre Dichte gegeben ist durch

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \\ \text{für } x \in \{0, 1, \dots, n\} \quad \text{und} \quad 0 < p < 1 . \quad (224.1)$$

Zunächst wird geprüft, ob (223.2) erfüllt ist. Das ist der Fall, denn mit $p > 0$ und $1 - p > 0$ ist auch $f(x) > 0$, und mit der binomischen Reihe ergibt sich

$$1 = ((1 - p) + p)^n = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} = \sum_{x=0}^n f(x).$$

Die Binomialverteilung soll nun hergeleitet werden. Das Ereignis A bezeichne den Erfolg, dann ergibt das zu A komplementäre Ereignis \bar{A} den Mißerfolg. Weiter sei $P(A) = p$, so daß mit (213.4) $P(\bar{A}) = 1 - p$ folgt. Die Wahrscheinlichkeit, daß bei n aufeinander folgenden Versuchen zunächst k Erfolge und dann $n - k$ Mißerfolge eintreten, ergibt sich wegen der Unabhängigkeit der einzelnen Versuche mit (215.2) zu

$$P(A \cap A \cap \dots \cap A \cap \bar{A} \cap \bar{A} \cap \dots \cap \bar{A}) = p^k (1 - p)^{n-k}.$$

Es brauchen nun die ersten k Versuche nicht auch k Erfolge zu bewirken, sondern nach (136.2) gibt es $\binom{n}{k}$ Möglichkeiten, daß k Erfolge bei n Versuchen eintreffen. Mit (213.3) ergibt sich daher die Wahrscheinlichkeit von k Erfolgen unter n Versuchen zu $\binom{n}{k} p^k (1 - p)^{n-k}$. Ersetzt man k durch x , erhält man die Dichte in (224.1).

Beispiel: Gesucht ist die Wahrscheinlichkeit, daß in einer Fertigung von 4 Produkten x Produkte mit $x \in \{0, 1, 2, 3, 4\}$ fehlerhaft sind, falls die Wahrscheinlichkeit, daß ein bestimmtes Produkt defekt ist, $p = 0,3$ beträgt und die Fertigungen der einzelnen Produkte voneinander unabhängig sind. Aus (224.1) erhält man

$$f(x) = \binom{4}{x} 0,3^x \cdot 0,7^{4-x} \quad \text{für } x \in \{0, 1, 2, 3, 4\}$$

und somit

$$f(0) = 0,240, f(1) = 0,412, f(2) = 0,264, f(3) = 0,076, f(4) = 0,008.$$

Aus (223.3) folgt dann beispielsweise für die Wahrscheinlichkeit $P(X < 2)$, daß weniger als 2 Produkte fehlerhaft sind, $P(X < 2) = 0,652$. Die Dichte $f(x)$ und die Verteilungsfunktion $F(x)$ für dieses Beispiel sind in Abbildung 224-1 dargestellt.

Geht die Anzahl der Wiederholungen eines Experimentes gegen unendlich und die Wahrscheinlichkeit des Eintreffens eines Erfolges gegen Null, ergibt sich aus der Binomialverteilung die Poisson-Verteilung.

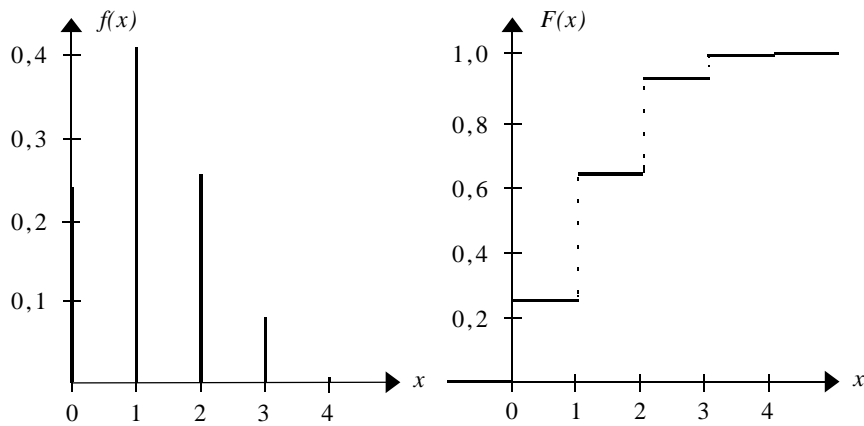


Abb. 224-1: Dichte und Verteilungsfunktion

Definition: Die Zufallsvariable X besitzt die *Poisson-Verteilung* mit dem reellen Parameter λ , falls ihre Dichte gegeben ist durch

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{für } x \in \{0, 1, 2, \dots\} \quad \text{und } \lambda > 0. \quad (224.2)$$

Da $\lambda > 0$ ist auch $f(x) > 0$ und mit der Reihenentwicklung der Exponentialfunktion e^λ

$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \quad (224.3)$$

ergibt sich $\sum_{x=0}^{\infty} f(x) = 1$, so daß (223.2) erfüllt ist.

Zur Herleitung der Poisson-Verteilung aus der Binomialverteilung setzt man $p = \lambda/n$, so daß $p \rightarrow 0$ bei $n \rightarrow \infty$ geht. Hiermit ergibt sich aus (224.1)

$$\frac{n(n-1)\dots(n-x+1)}{x!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}.$$

Mit $n \rightarrow \infty$ folgen die Grenzwerte

$$\lim_{n \rightarrow \infty} \left(\frac{n}{n}\right) \left(\frac{n-1}{n}\right) \dots \left(\frac{n-x+1}{n}\right) = 1 \quad \text{und} \quad \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-x} = 1$$

sowie mit der Definition der Zahl e (SMIRNOW 1975, Teil I, S.93)

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda},$$

so daß die Dichte in (224.2) folgt.

Der Parameter λ ist der Erwartungswert einer Zufallsvariablen mit der Poisson-Verteilung, denn, wie in (231.1) definiert wird, berechnet sich der Erwartungswert unter Beachtung von (224.3) zu

$$\sum_{x=0}^{\infty} xf(x) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda. \quad (224.4)$$

Die Werte einer Zufallsvariablen mit einer Poisson-Verteilung sind positive ganze Zahlen. Zufallsphänomene, bei denen gezählt wird, zum Beispiel die Anzahl der in einer Telefonzentrale pro Stunde eintreffenden Anrufe, ergeben sich daher häufig als Poisson-verteilt. Ein Beispiel hierzu befindet sich im Kapitel 244.

225 Mehrdimensionale stetige Zufallsvariable

Wie bereits im Kapitel 221 erwähnt, muß zwischen ein- und mehrdimensionalen Zufallsvariablen unterschieden werden. Mehrdimensionale Zufallsvariable sollen an den folgenden beiden Beispielen erläutert werden.

1. Beispiel: Mit einem Bogen wird auf eine Zielscheibe geschossen, die zum Registrieren der Treffer mit einem orthogonalen (x, y) -Koordinatensystem versehen wurde, dessen Ursprung mit dem Mittelpunkt der Scheibe zusammenfällt. Jeder Treffer, der ein Elementarereignis darstellt, wird durch seine (x, y) -Koordinaten festgelegt. Damit ordnet man jedem Elementarereignis einen Punkt in der (x, y) -Ebene zu, so daß eine zweidimensionale Zufallsvariable erhalten wird.

2. Beispiel: Ein Elementarereignis bestehe darin, daß die Längen von n Strecken mit einem Streckenmeßgerät jeweils einmal gemessen werden. Dem Elementarereignis wird dann ein Punkt mit n Koordinaten beziehungsweise ein Vektor in dem mit (123.6) definierten n -dimensionalen Euklidischen Raum E^n zugeordnet, so daß eine n -dimensionale Zufallsvariable erhalten wird.

Entsprechend (221.1) ergibt sich die

Definition: Man bezeichnet eine eindeutige Funktion $X(s_i) = (X_1(s_i), X_2(s_i), \dots, X_n(s_i))$ mit Werten im \mathbb{R}^n , die im Wahrscheinlichkeitsraum (S, Z, P) auf der Menge S der Elementarereignisse s_i definiert ist, als *n -dimensionale Zufallsvariable*, falls für jedes beliebige $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ das Ereignis, für das $X_1(s_i) < x_1, X_2(s_i) < x_2, \dots, X_n(s_i) < x_n$ gilt, zu den zufälligen Ereignissen von Z gehört. (225.1)

Von den mehrdimensionalen Zufallsvariablen sollen im folgenden nur die stetigen Zufallsvariablen behandelt werden. (223.5) entsprechend ergibt sich die

Definition: Man bezeichnet X_1, \dots, X_n als stetige n -dimensionale Zufallsvariable, falls eine nichtnegative integrierbare Funktion $f(x_1, \dots, x_n)$ existiert, die für beliebige $x_1, \dots, x_n \in \mathbb{R}$ die Beziehung

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_1 \dots dt_n$$

erfüllt, wobei $F(x_1, \dots, x_n)$ die Verteilungsfunktion und t_1, \dots, t_n die Integrationsvariablen bedeuten. Die Funktion $f(x_1, \dots, x_n)$ nennt man Dichte oder *multivariate Verteilung* von X_1, \dots, X_n . (225.2)

Die Wahrscheinlichkeit des Ereignisses $X_1 < x_1, \dots, X_n < x_n$ ergibt sich mit (225.2) zu

$$P(X_1 < x_1, \dots, X_n < x_n) = F(x_1, \dots, x_n) \quad (225.3)$$

und des Ereignisses $x_{1u} \leq X_1 < x_{1o}, \dots, x_{nu} \leq X_n < x_{no}$

$$\begin{aligned} P(x_{1u} \leq X_1 < x_{1o}, \dots, x_{nu} \leq X_n < x_{no}) \\ = \int_{x_{nu}}^{x_{no}} \dots \int_{x_{1u}}^{x_{1o}} f(x_1, \dots, x_n) dx_1 \dots dx_n, \end{aligned} \quad (225.4)$$

wobei wie bei den eindimensionalen Zufallsvariablen $P(x_{1u} \leq X_1 < x_{1o}, \dots, x_{nu} \leq X_n < x_{no}) = P(x_{1u} < X_1 < x_{1o}, \dots, x_{nu} < X_n < x_{no})$ gilt. Durch Differentiation des Integrals in (225.2) nach seiner oberen Grenze folgt, sofern $f(x_1, \dots, x_n)$ an der Stelle x_1, \dots, x_n stetig ist,

$$\partial^n F(x_1, \dots, x_n) / \partial x_1 \dots \partial x_n = f(x_1, \dots, x_n). \quad (225.5)$$

Damit die Funktion $f(x_1, \dots, x_n)$ die Dichte einer n -dimensionalen Zufallsvariablen X_1, \dots, X_n angibt, ist es notwendig, daß (223.10) entsprechend gilt

$$f(x_1, \dots, x_n) \geq 0 \quad \text{und} \quad \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1.$$

Mehrdimensionale Zufallsvariable definiert man (121.1) entsprechend als ^(225.6) *Zufallsvektoren*, beispielsweise die n -dimensionale Zufallsvariable X_1, \dots, X_n als $n \times 1$ Zufallsvektor $\mathbf{x} = |X_1, \dots, X_n|'$. Der Zufallsvektor \mathbf{x} wird bewußt mit einem kleinen Buchstaben bezeichnet, um ihn von einer Matrix zu unterscheiden. Der Zufallsvektor \mathbf{x} enthält also als Komponenten die Zufallsvariablen X_i , nicht dagegen die Werte x_i , die die Zufallsvariablen X_i annehmen können.

Diese Unterscheidung wird allerdings zur Vereinfachung der Bezeichnung beginnend mit dem Kapitel 251 fortfallen. Zufallsvariable und die Werte, die sie annehmen können, werden dann einheitlich mit kleinen Buchstaben bezeichnet, sofern aus dem Zusammenhang zu entnehmen ist, welche Größe gemeint ist.

226 Randverteilung

Es sei $f(x_1, x_2)$ die Dichte einer zweidimensionalen Zufallsvariablen X_1, X_2 . Ist man lediglich an der Zufallsvariablen X_1 interessiert und fragt nach der Wahrscheinlichkeit des Ereignisses $X_1 < a$, ergibt sich mit (225.2) und (225.3)

$$P(X_1 < a) = F(a, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^a f(x_1, x_2) dx_1 dx_2 = \int_{-\infty}^a g(x_1) dx_1 . \quad (226.1)$$

Die Dichte $g(x_1)$ mit

$$g(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \quad (226.2)$$

bezeichnet man als Randverteilung von X_1 . Sie ist, da über x_2 in den Grenzen von $-\infty$ bis ∞ integriert wird, nur von x_1 abhängig. Für n Dimensionen ergibt sich die

Definition: Es sei $f(x_1, \dots, x_n)$ die Verteilung des $n \times 1$ Zufallsvektors $\mathbf{x} = |X_1, \dots, X_n|'$. Dann ist die *Randverteilungsfunktion* $G(\infty, \dots, \infty, x_{i+1}, \dots, x_n)$ der Zufallsvariablen X_{i+1}, \dots, X_n gegeben durch

$$G(\infty, \dots, \infty, x_{i+1}, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_{i+1}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) dt_1 \dots dt_n . \quad (226.3)$$

Hieraus folgt die *Randverteilung* $g(x_{i+1}, \dots, x_n)$ der Zufallsvariablen X_{i+1}, \dots, X_n mit

$$g(x_{i+1}, \dots, x_n) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(t_1, \dots, t_i, t_{i+1}, \dots, t_n) dt_1 \dots dt_i . \quad (226.4)$$

Die Reihenfolge und die Anzahl der Zufallsvariablen in (226.3) und (226.4) ist beliebig. Mit Hilfe der Randverteilung läßt sich also die Dichte mehrdimensionaler Zufallsvariablen auf die Dichte von Zufallsvariablen niedriger Dimension zurückführen, was bei der Herleitung von Verteilungen benutzt wird. Beispiele für Randverteilungen befinden sich in (253.1) und in den Beweisen zu (245.1), (263.1), (264.1) und (265.1).

227 Bedingte Verteilung

Mit (214.1) war die bedingte Wahrscheinlichkeit eines Ereignisses definiert worden. Entsprechend läßt sich auch die bedingte Verteilung von Zufallsvariablen einführen, was zunächst für die zweidimensionale Zufallsvariable X_1, X_2 mit der Dichte $f(x_1, x_2)$ gezeigt werden soll. Es wird das Ereignis $X_1 < x_1$ unter der Bedingung betrachtet, daß das Ereignis $x_2 \leq X_2 < x_2 + \Delta x_2$ eingetroffen ist. Mit (214.1), (225.2) und (225.3) ergibt sich seine Wahrscheinlichkeit zu

$$\begin{aligned} P(X_1 < x_1 | x_2 \leq X_2 < x_2 + \Delta x_2) &= \frac{P(X_1 < x_1, x_2 \leq X_2 < x_2 + \Delta x_2)}{P(x_2 \leq X_2 < x_2 + \Delta x_2)} \\ &= \frac{\int_{x_2}^{x_2 + \Delta x_2} \int_{-\infty}^{x_1} f(x_1, x_2) dx_1 dx_2}{\int_{x_2}^{x_2 + \Delta x_2} \int_{-\infty}^{\infty} f(x_1, x_2) dx_1 dx_2} \quad \text{für } P(x_2 \leq X_2 < x_2 + \Delta x_2) > 0. \end{aligned} \quad (227.1)$$

Die bedingte Wahrscheinlichkeit $P(X_1 < x_1 | X_2 = x_2)$ ist gesucht. Sie muß durch einen Grenzprozeß definiert werden, da mit (223.7) für stetige Zufallsvariable $P(X = x) = 0$ gilt. Dividiert man Zähler und Nenner der rechten Seite von (227.1) durch Δx_2 , erhält man mit der Randverteilung $g(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1$ für die bedingte Verteilungsfunktion

$$\begin{aligned} F(x_1 | x_2) &= \lim_{\Delta x_2 \rightarrow 0} P(X_1 < x_1 | x_2 \leq X_2 < x_2 + \Delta x_2) \\ &= \frac{\int_{-\infty}^{x_1} f(x_1, x_2) dx_1}{g(x_2)}, \end{aligned} \quad (227.2)$$

so daß die bedingte Verteilung $f(x_1 | x_2)$ folgt mit

$$f(x_1 | x_2) = \frac{f(x_1, x_2)}{g(x_2)}. \quad (227.3)$$

Definition: Es sei $f(x_1, \dots, x_n)$ die Dichte des $n \times 1$ Zufallsvektors $\mathbf{x} = |X_1, \dots, X_n|'$. Dann ist die *bedingte Verteilungsfunktion* $F(x_1, \dots, x_i | x_{i+1}, \dots, x_n)$ der Zufallsvariablen X_1, \dots, X_i unter der Bedingung, daß $X_{i+1} = x_{i+1}, \dots, X_n = x_n$ gilt, gegeben durch

$$F(x_1, \dots, x_i | x_{i+1}, \dots, x_n) = \frac{\int_{-\infty}^{x_i} \dots \int_{-\infty}^{x_1} f(x_1, \dots, x_n) dx_1 \dots dx_i}{g(x_{i+1}, \dots, x_n)},$$

worin $g(x_{i+1}, \dots, x_n)$ die Randverteilung der Zufallsvariablen X_{i+1}, \dots, X_n bedeutet. (227.4)

Aus (227.4) ergibt sich die *bedingte Verteilung* zu

$$f(x_1, \dots, x_i | x_{i+1}, \dots, x_n) = \frac{f(x_1, \dots, x_n)}{g(x_{i+1}, \dots, x_n)}. \quad (227.5)$$

Ein Beispiel für die bedingte Verteilung befindet sich in (253.2).

228 Unabhängige Zufallsvariable

Der bedingten Wahrscheinlichkeit entsprechend wurde im vorangegangenen Kapitel die bedingte Verteilung definiert. Mit Hilfe der bedingten Wahrscheinlichkeit ergab sich in (215.1) die Unabhängigkeit von Ereignissen, so daß unabhängige Zufallsvariable entsprechend einzuführen sind.

Definition: Es sei $\mathbf{x} = |X_1, \dots, X_n|'$ ein $n \times 1$ Zufallsvektor mit der Dichte $f(x_1, \dots, x_n)$. Ist die bedingte Verteilungsfunktion $F(x_1, \dots, x_i | x_{i+1}, \dots, x_n)$ lediglich eine Funktion $H(x_1, \dots, x_i)$ der Zufallsvariablen X_1, \dots, X_i

$$F(x_1, \dots, x_i | x_{i+1}, \dots, x_n) = H(x_1, \dots, x_i),$$

dann heißen die beiden Mengen X_1, \dots, X_i und X_{i+1}, \dots, X_n von Zufallsvariablen voneinander *unabhängig*. (228.1)

Auf den folgenden Satz wird später bei der Herleitung von Verteilungen häufig zurückgegriffen werden.

Satz: Zwei Mengen X_1, \dots, X_i und X_{i+1}, \dots, X_n von Zufallsvariablen sind genau dann voneinander unabhängig, wenn ihre gemeinsame Verteilung aus dem Produkt ihrer Randverteilungen sich bestimmt. (228.2)

Beweis: Die beiden Mengen von Zufallsvariablen seien voneinander unabhängig, so daß (228.1) gilt und nach Substitution in (227.4)

$$H(x_1, \dots, x_i)g(x_{i+1}, \dots, x_n) = \int_{-\infty}^{x_i} \dots \int_{-\infty}^{x_1} f(x_1, \dots, x_n) dx_1 \dots dx_i.$$

Integriert man diese Gleichung über x_{i+1}, \dots, x_n in den Grenzen von $-\infty$ bis ∞ , ergibt sich mit (225.6) und (226.3) $H(x_1, \dots, x_i) = G(x_1, \dots, x_i, \infty, \dots, \infty)$ oder wegen (228.1)

$$F(x_1, \dots, x_i | x_{i+1}, \dots, x_n) = G(x_1, \dots, x_i, \infty, \dots, \infty)$$

und den Schritten von (226.3) nach (226.4) und (227.4) nach (227.5) entsprechend

$$f(x_1, \dots, x_i | x_{i+1}, \dots, x_n) = g(x_1, \dots, x_i). \quad (228.3)$$

Dies in (227.5) substituiert ergibt die erste Aussage. Gilt andererseits $f(x_1, \dots, x_n) = g(x_1, \dots, x_i)g(x_{i+1}, \dots, x_n)$ und wird nach (227.4) die bedingte Verteilungsfunktion $F(x_1, \dots, x_i | x_{i+1}, \dots, x_n)$ gebildet, so ist sie nur von X_1, \dots, X_i abhängig, so daß nach (228.1) die beiden Mengen von Zufallsvariablen voneinander unabhängig sind.

229 Transformation von Verteilungen

Ist die Verteilung eines Zufallsvektors bekannt, so stellt sich häufig das Problem, die Verteilung einer Funktion des Zufallsvektors abzuleiten.

Satz: Für die Transformation der Werte des Zufallsvektors $\mathbf{y} = |Y_1, \dots, Y_n|'$ in die von $\mathbf{x} = |X_1, \dots, X_n|'$ gelte die eineindeutige Abbildung $x_i = g_i(y_1, \dots, y_n)$ mit $i \in \{1, \dots, n\}$ und der einmal stetig differenzierbaren Funktion g_i , und es existiere die Jacobische Matrix $\mathbf{J} = (\partial g_i / \partial y_j)$ mit $\det \mathbf{J} \neq 0$ und $i, j \in \{1, \dots, n\}$, so daß die inverse Transformation $y_i = h_i(x_1, \dots, x_n)$ eindeutig ist. Weiter sei $f(x_1, \dots, x_n)$ die Verteilung von \mathbf{x} , dann folgt die Verteilung $b(y_1, \dots, y_n)$ von \mathbf{y} mit

$$b(y_1, \dots, y_n) = f(g_1(y_1, \dots, y_n), \dots, g_n(y_1, \dots, y_n)) |\det \mathbf{J}|. \quad (229.1)$$

Beweis: Die Wahrscheinlichkeit, daß die Zufallsvariablen X_1, \dots, X_n im Bereich B liegen, ergibt sich aus (225.2). Transformiert man in dem gewonnenen Integral die Variablen mit $x_i = g_i(y_1, \dots, y_n)$, wodurch der Bereich S auf den Bereich B abgebildet werde, erhält man mit (173.3)

$$\begin{aligned} & \int \dots \int_B f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int \dots \int_S f(g_1(y_1, \dots, y_n), \dots, g_n(y_1, \dots, y_n)) |\det \mathbf{J}| dy_1 \dots dy_n, \end{aligned}$$

so daß sich als Dichte $b(y_1, \dots, y_n)$ die angegebene Funktion ergibt, die sicher nicht negativ ist und somit (225.6) erfüllt.

23 Erwartungswerte und Momente von Zufallsvariablen

231 Erwartungswert

Man erhält den Erwartungswert einer Zufallsvariablen, indem der Durchschnittswert aller möglichen Werte der Variablen unter Berücksichtigung ihrer Wahrscheinlichkeitsdichten gebildet wird. Den Erwartungswert kann man sich

daher auch als einen Mittelwert vorstellen.

Beispiel: Zur Erläuterung der Binomialverteilung (224.1) war für eine Fertigung von 4 Produkten die Wahrscheinlichkeit von x fehlerhaften Produkten mit $x \in \{0, 1, 2, 3, 4\}$ zu $f(0) = 0,240$, $f(1) = 0,412$, $f(2) = 0,264$, $f(3) = 0,076$ und $f(4) = 0,008$ berechnet worden. Interpretiert man die Wahrscheinlichkeit nach (211.2) als relative Häufigkeit, erhält man zum Beispiel bei 1000 Fertigungen 240 mal kein fehlerhaftes Produkt, 412 mal ein fehlerhaftes Produkt und so fort. Die Gesamtanzahl der fehlerhaften Produkte beträgt

$$240 \cdot 0 + 412 \cdot 1 + 264 \cdot 2 + 76 \cdot 3 + 8 \cdot 4 = 1200,$$

so daß der Erwartungswert der fehlerhaften Produkte einer Fertigung sich als Durchschnitt zu 1,2 berechnet. Der Erwartungswert ergibt sich auch unmittelbar aus der Summe der Produkte der Werte der Zufallsvariablen und ihrer Wahrscheinlichkeitsdichten zu

$$0,240 \cdot 0 + 0,412 \cdot 1 + 0,264 \cdot 2 + 0,076 \cdot 3 + 0,008 \cdot 4 = 1,2.$$

Definition: Die diskrete Zufallsvariable X besitze die Dichte $f(x_i)$, dann bezeichnet man μ oder $E(X)$

$$\mu = E(X) = \sum_{i=1}^n x_i f(x_i) \quad \text{für } i \in \{1, \dots, n\}$$

und

$$\mu = E(X) = \sum_{i=1}^{\infty} x_i f(x_i) \quad \text{für } i \in \{1, \dots, \infty\}$$

als den *Erwartungswert* der Zufallsvariablen X . (231.1)

Für n -dimensionale stetige Zufallsvariable gilt die

Definition: Der $n \times 1$ Zufallsvektor $\mathbf{x} = |X_1, \dots, X_n|'$ besitze die Dichte $f(x_1, \dots, x_n)$, dann bezeichnet man μ_i oder $E(X_i)$

$$\mu_i = E(X_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i f(x_1, \dots, x_n) dx_1 \dots dx_n$$

als den Erwartungswert der Zufallsvariablen X_i , falls das mehrfache Integral existiert. Ersetzt man X_i und x_i durch eine Funktion dieser Größen, ergibt sich der Erwartungswert einer Funktion der Zufallsvariablen X_i . (231.2)

Für die eindimensionale Zufallsvariable X folgt aus (231.2)

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad (231.3)$$

und anstelle von (231.2) mit Hilfe der Randverteilung $g(x_i)$ von X_i aus (226.4)

$$\mu_i = E(X_i) = \int_{-\infty}^{\infty} x_i g(x_i) dx_i . \quad (231.4)$$

Der Erwartungswert liegt im Zentrum einer Verteilung, wie für die Zufallsvariable X_i in Abbildung 231-1 dargestellt ist, denn die x_i -Achse kann man

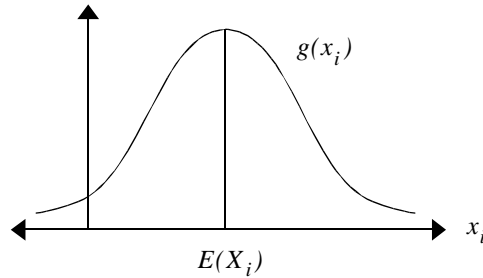


Abb. 231-1: Erwartungswert

sich als Stab mit der Dichte $g(x_i)$ vorstellen. Für das Massenzentrum x_s des Stabes gilt nach den Gesetzen der Mechanik

$$x_s = \int_{-\infty}^{\infty} x_i g(x_i) dx_i / \int_{-\infty}^{\infty} g(x_i) dx_i ,$$

woraus $x_s = E(X_i)$ wegen (225.6) folgt.

Für den Erwartungswert einer linearen Transformation von Zufallsvektoren gilt der folgende

Satz: Es seien \mathbf{A} und \mathbf{B} zwei $m \times n$ und $m \times o$ Matrizen sowie \mathbf{c} ein $m \times 1$ Vektor von Konstanten, $\mathbf{x} = |X_1, \dots, X_n|'$ und $\mathbf{y} = |Y_1, \dots, Y_o|'$ zwei $n \times 1$ und $o \times 1$ Zufallsvektoren, dann gilt

$$E(\mathbf{Ax} + \mathbf{By} + \mathbf{c}) = \mathbf{AE}(\mathbf{x}) + \mathbf{BE}(\mathbf{y}) + \mathbf{c} . \quad (231.5)$$

Beweis: Es sei $\mathbf{d} = \mathbf{Ax}$ mit $\mathbf{d} = (d_i)$, $\mathbf{e} = \mathbf{By}$ mit $\mathbf{e} = (e_i)$ sowie $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ik})$ und $\mathbf{c} = (c_i)$. Definiert man die Konstante c_i mit $c_i = h(X_l)$ als Funktion $h(X_l)$ einer Zufallsvariablen X_l mit der Dichte $g(x_l)$, folgt mit (231.2) und (225.6) $E(c_i) = E(h(X_l)) = c_i \int_{-\infty}^{\infty} g(x_l) dx_l = c_i$ und weiter mit

(131.6)

$$\begin{aligned} E(d_i + e_i + c_i) &= E\left(\sum_{j=1}^n a_{ij} X_j + \sum_{k=1}^o b_{ik} Y_k + c_i\right) \\ &= \sum_{j=1}^n a_{ij} E(X_j) + \sum_{k=1}^o b_{ik} E(Y_k) + c_i \end{aligned}$$

und damit die Aussage.

Beispiel: Es ist der Erwartungswert $E((X - \mu)^2)$ zu berechnen, worin X eine Zufallsvariable und $\mu = E(X)$ bedeuten. Man erhält

$$E((X - \mu)^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2 .$$

232 Multivariate Momente

Die Erwartungswerte von Zufallsvariablen sind Sonderfälle der Momente von Zufallsvariablen.

Definition: Der $n \times 1$ Zufallsvektor $\mathbf{x} = |X_1, \dots, X_n|'$ besitze die multivariate Verteilung $f(x_1, \dots, x_n)$, dann bezeichnet man $\mu_{x_1 \dots x_n}^{(k)}$ mit $k = \sum_{i=1}^n k_i$ und $k_i \in \mathbb{N}$

$$\begin{aligned} \mu_{x_1 \dots x_n}^{(k)} &= E(X_1^{k_1} X_2^{k_2} \dots X_n^{k_n}) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{k_1} x_2^{k_2} \dots x_n^{k_n} f(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

als k -tes *multivariate Moment* von X_1, \dots, X_n , falls das mehrfache Integral existiert. (232.1)

Das k -te Moment von X_i folgt aus (232.1) zu

$$\mu_{x_i}^{(k)} = E(X_i^k) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_i^k f(x_1, \dots, x_n) dx_1 \dots dx_n \quad (232.2)$$

und das k -te Moment der eindimensionalen Zufallsvariablen X zu

$$\mu_x^{(k)} = E(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx . \quad (232.3)$$

Für $k = 1$ ergeben sich die Erwartungswerte $\mu_i = \mu_{x_i}^{(1)}$ und $\mu = \mu_x^{(1)}$.

Definition: Momente in bezug auf die Erwartungswerte

$$E((X_1 - \mu_1)^{k_1} (X_2 - \mu_2)^{k_2} \dots (X_n - \mu_n)^{k_n})$$

bezeichnet man als *zentrale Momente*. (232.4)

Von besonderer Bedeutung sind die zweiten zentralen Momente.

Definition: Als *Kovarianz* σ_{ij} oder $C(X_i, X_j)$ bezeichnet man das zweite zentrale Moment der Zufallsvariablen X_i und X_j des $n \times 1$ Zufallsvektors $\mathbf{x} = |X_1, \dots, X_n|'$ mit der multivariaten Verteilung $f(x_1, \dots, x_n)$

$$\begin{aligned}\sigma_{ij} &= C(X_i, X_j) = E((X_i - \mu_i)(X_j - \mu_j)) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f(x_1, \dots, x_n) dx_1 \dots dx_n\end{aligned}$$

und als *Varianz* oder *Dispersion* σ_i^2 beziehungsweise $V(X_i)$ das zweite zentrale Moment der Zufallsvariablen X_i

$$\begin{aligned}\sigma_i^2 &= V(X_i) = \sigma_{ii} = C(X_i, X_i) = E((X_i - \mu_i)^2) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_i - \mu_i)^2 f(x_1, \dots, x_n) dx_1 \dots dx_n\end{aligned}$$

sowie als *Standardabweichung* σ_i die positive Quadratwurzel der Varianz. (232.5)

Wie aus der Definition der Varianz σ_i^2 zusammen mit (225.6) ersichtlich, gilt $\sigma_i^2 \geq 0$, so daß die Standardabweichung σ_i aus σ_i^2 mit $\sigma_i = \sqrt{\sigma_i^2}$ folgt. Der positiven und negativen Wurzel der Varianz entspricht in der Fehlertheorie der Ausgleichsrechnung der *mittlere Fehler*, wobei aber auch Schätzwerte dieser Größe als mittlere Fehler bezeichnet werden.

Die Varianz σ_i^2 ist ein Maß für die Streuung einer Zufallsvariablen X_i um ihren Erwartungswert, denn mit der Randverteilung $g(x_i)$ von X_i aus (226.4) ergibt sich anstelle von (232.5)

$$\sigma_i^2 = \int_{-\infty}^{\infty} (x_i - \mu_i)^2 g(x_i) dx_i. \quad (232.6)$$

Wenn der Hauptanteil der Fläche unter der in Abbildung 231-1 dargestellten Randverteilung $g(x_i)$ in der Nähe des Erwartungswertes $E(X_i)$ liegt, ist die Varianz σ_i^2 klein. Umgekehrt ist σ_i^2 groß, wenn die Fläche ausgebreitet ist.

Die Kovarianz σ_{ij} gibt ein Maß für die Abhängigkeit zwischen den Zufallsvariablen X_i und X_j an, wie mit den beiden folgenden Sätzen deutlich wird.

Satz: Sind die Zufallsvariablen X_i und X_j voneinander unabhängig, gilt $\sigma_{ij} = 0$. (232.7)

Beweis: Aus (232.5) ergibt sich $\sigma_{ij} = E(X_i X_j) - \mu_j E(X_i) - \mu_i E(X_j) + \mu_i \mu_j$ und mit (231.2)

$$\sigma_{ij} = E(X_i X_j) - E(X_i)E(X_j). \quad (232.8)$$

Weiter gilt mit (232.1) und der Randverteilung $g(x_i, x_j)$ für X_i und X_j aus (226.4)

$$E(X_i X_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j g(x_i, x_j) dx_i dx_j$$

und wegen der Unabhängigkeit von X_i und X_j mit (228.2) $g(x_i, x_j) = g(x_i)g(x_j)$, so daß mit (231.4) $E(X_i X_j) = E(X_i)E(X_j)$ und damit die Aussage folgt.

Umgekehrt bedeutet eine verschwindende Kovarianz im allgemeinen keine Unabhängigkeit der Zufallsvariablen, wie beispielsweise die in (321.3) bis (321.5) vorgenommene Transformation eines Zufallsvektors zeigt. Eine Ausnahme bilden, wie in (254.1) nachgewiesen wird, die normalverteilten Zufallsvariablen.

Zum besseren Vergleich der Kovarianzen werden sie normiert.

Definition: Für die Varianzen der Zufallsvariablen X_i und X_j gelte $\sigma_i^2 > 0$ und $\sigma_j^2 > 0$, dann bezeichnet man ρ_{ij}

$$\rho_{ij} = \sigma_{ij} / (\sigma_i \sigma_j)$$

als *Korrelationskoeffizienten* von X_i und X_j . (232.9)

Gilt $\sigma_{ij} = \rho_{ij} = 0$, so sagt man, daß die Zufallsvariablen X_i und X_j nicht miteinander korreliert oder *unkorreliert* seien.

Satz: Für den Korrelationskoeffizienten ρ_{ij} folgt

$$-1 \leq \rho_{ij} \leq 1,$$

wobei $\rho_{ij} = \pm 1$ genau dann gilt, wenn mit einer Wahrscheinlichkeit von Eins eine lineare Beziehung zwischen den Zufallsvariablen X_i und X_j besteht, also $P(X_j = cX_i + d) = 1$ mit den Konstanten c und d . (232.10)

Beweis: Aufgrund der Integraldarstellung (SMIRNOW 1975, *Teil II*, S.487)

$$\left(\int_A f g dx \right)^2 \leq \int_A f^2 dx \int_A g^2 dx$$

der Schwarzischen Ungleichung, angewendet auf die Integrale in (232.5), folgt $\sigma_{ij}^2 \leq \sigma_i^2 \sigma_j^2$, woraus mit (232.9) die erste Aussage sich ergibt. Definiert man

$$E([c(X_i - \mu_i) - (X_j - \mu_j)]^2) = h \geq 0,$$

so ist die linke Seite eine quadratische Funktion von c , die der Konstanten h gleicht, die größer oder gleich Null ist. Mit (231.5) und (232.5) folgt daraus die quadratische Gleichung für c

$$\sigma_i^2 c^2 - 2\sigma_{ij} c + \sigma_j^2 - h = 0$$

mit reellen oder komplexen Lösungen, wie aus der Diskriminante (ABRAMOWITZ und STEGUN 1972, S.17)

$$4\sigma_{ij}^2 - 4\sigma_i^2(\sigma_j^2 - h) = 4(\sigma_{ij}^2 - \sigma_i^2\sigma_j^2 + \sigma_i^2 h)$$

ersichtlich, da $\sigma_i^2 h \geq 0$ und aufgrund der ersten Aussage $\sigma_{ij}^2 - \sigma_i^2\sigma_j^2 \leq 0$ ist. Für $h = 0$ und $\sigma_{ij}^2 = \sigma_i^2\sigma_j^2$ oder $\rho_{ij} = \pm 1$ ergeben sich zwei identische reelle Lösungen, beispielsweise c_0 , folglich mit der Randverteilung $g(x_i, x_j)$ von X_i und X_j

$$\begin{aligned} E([c_0(X_i - \mu_i) - (X_j - \mu_j)]^2) \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [c_0(x_i - \mu_i) - (x_j - \mu_j)]^2 g(x_i, x_j) dx_i dx_j = 0, \end{aligned}$$

was wegen $g(x_i, x_j) \geq 0$ nur erfüllt sein kann, wenn mit einer Wahrscheinlichkeit von Eins $c_0(X_i - \mu_i) = X_j - \mu_j$ oder $X_j = c_0 X_i + d$ mit $d = \mu_j - c_0 \mu_i$ gilt, also $P(X_j = c_0 X_i + d) = 1$. Besteht andererseits eine lineare Beziehung zwischen den Zufallsvariablen X_i und X_j , ergibt sich die quadratische Gleichung $E([c(X_i - \mu_i) - (X_j - \mu_j)]^2) = 0$, die für c zwei identische reelle Lösungen besitzt, falls $\sigma_{ij}^2 = \sigma_i^2\sigma_j^2$ gilt, woraus die Aussage folgt.

233 Kovarianzmatrix, Fehlerfortpflanzungsgesetz, Korrelationsmatrix und Gewichtsmatrix

a) Kovarianzmatrix und Fehlerfortpflanzungsgesetz

Die Varianzen und Kovarianzen eines Zufallsvektors faßt man in einer Matrix zusammen.

Definition: Es sei $\mathbf{x} = |X_1, \dots, X_n|'$ ein $n \times 1$ Zufallsvektor, dann bezeichnet man

$$\begin{aligned} D(\mathbf{x}) &= (\sigma_{ij}) = (C(X_i, X_j)) = E((\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))') \\ &= \begin{vmatrix} V(X_1) & C(X_1, X_2) & \dots & C(X_1, X_n) \\ C(X_2, X_1) & V(X_2) & \dots & C(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ C(X_n, X_1) & C(X_n, X_2) & \dots & V(X_n) \end{vmatrix} \end{aligned}$$

und daher $\Delta\mathbf{y} = \mathbf{A}\Delta\mathbf{x}$ mit

$$\mathbf{A} = \begin{vmatrix} \frac{\partial h_1}{\partial x_1} |_{\mathbf{x}=\mathbf{x}_0} & \cdots & \frac{\partial h_1}{\partial x_n} |_{\mathbf{x}=\mathbf{x}_0} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \frac{\partial h_m}{\partial x_1} |_{\mathbf{x}=\mathbf{x}_0} & \cdots & \frac{\partial h_m}{\partial x_n} |_{\mathbf{x}=\mathbf{x}_0} \end{vmatrix}. \tag{233.4}$$

Aus (233.2) folgt $D(\Delta\mathbf{y}) = \mathbf{A}D(\Delta\mathbf{x})\mathbf{A}'$ und $D(\Delta\mathbf{y}) = D(\mathbf{y})$ und $D(\Delta\mathbf{x}) = D(\mathbf{x})$ wegen $\Delta\mathbf{y} = \mathbf{y} - \mathbf{y}_0$ und $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}_0$ allerdings nur für kleine Werte für $\Delta\mathbf{x}$ und $E(\Delta\mathbf{x})$, so daß für \mathbf{x}_0 genäherte Erwartungswerte von \mathbf{x} gewählt werden müssen. Dann läßt sich bei nichtlinearen Beziehungen (233.3) die Koeffizientenmatrix \mathbf{A} in (233.2) aus (233.4) gewinnen.

Beispiel: Der $m \times 1$ Zufallsvektor \mathbf{y} ergebe sich durch die lineare Transformation $\mathbf{y} = \mathbf{B}\mathbf{x}$ mit der $m \times n$ Matrix $\mathbf{B} = (b_{ij})$ aus dem $n \times 1$ Zufallsvektor \mathbf{x} . Die Elemente der Matrix \mathbf{A} aus (233.4) folgen dann mit

$$\mathbf{A} = \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{vmatrix} = \mathbf{B}$$

in Übereinstimmung mit (233.2).

Den Satz (233.2) in Verbindung mit (233.3) und (233.4) bezeichnet man in der Ausgleichsrechnung als *Fehlerfortpflanzungsgesetz*, denn er erlaubt, aus Varianzen beziehungsweise mittleren Fehlern und Kovarianzen von Zufallsvariablen die Varianzen und Kovarianzen von Funktionen dieser Zufallsvariablen abzuleiten. Weitere Beispiele zum Fehlerfortpflanzungsgesetz befinden sich am Ende dieses Kapitels.

Mit Hilfe von (233.2) lassen sich die folgenden Eigenschaften einer Kovarianzmatrix beweisen.

Satz: Eine Kovarianzmatrix ist positiv definit oder positiv semidefinit. (233.5)

Beweis: Die Zufallsvariable Y ergebe sich mit dem $n \times 1$ Vektor \mathbf{a} von Konstanten durch die lineare Transformation $Y = \mathbf{a}'\mathbf{x}$ aus dem $n \times 1$ Zufallsvektor $\mathbf{x} = |X_1, \dots, X_n|'$. Dann gilt mit (233.2) für die Varianz $V(Y) = \mathbf{a}'D(\mathbf{x})\mathbf{a}$, für die wegen (225.6) und (232.5) $\mathbf{a}'D(\mathbf{x})\mathbf{a} \geq 0$ gilt, woraus mit (143.1) die Aussage folgt, da $D(\mathbf{x})$ aus (233.1) wegen $C(X_i, X_j) = C(X_j, X_i)$ symmetrisch ist und \mathbf{a} beliebig gewählt werden kann.

Satz: Die Kovarianzmatrix $D(\mathbf{x})$ des $n \times 1$ Zufallsvektors $\mathbf{x} = |X_1, \dots, X_n|'$ ist genau dann positiv semidefinit, wenn die Wahrscheinlichkeit gleich Eins

ist, daß zwischen den Zufallsvariablen X_1, \dots, X_n eine lineare Beziehung besteht, daß also $\mathbf{a}'\mathbf{x} = c$ für einen $n \times 1$ Vektor $\mathbf{a} \neq \mathbf{0}$ und eine Konstante c gilt. (233.6)

Beweis: Ist die Kovarianzmatrix $D(\mathbf{x})$ positiv semidefinit, so existiert nach (143.1) ein $n \times 1$ Vektor $\mathbf{a} \neq \mathbf{0}$, für den mit (233.2) $\mathbf{a}'D(\mathbf{x})\mathbf{a} = D(\mathbf{a}'\mathbf{x}) = 0$ gilt. Hieraus folgt für $Y = \mathbf{a}'\mathbf{x}$ mit (232.5) $\int_{-\infty}^{\infty} (y - E(Y))^2 f(y) dy = 0$. Dies kann wegen $f(y) \geq 0$ nur für $Y = E(Y)$ mit einer Wahrscheinlichkeit von Eins erfüllt sein. Somit ergibt sich $P(Y = E(Y)) = 1$ oder $P(\mathbf{a}'\mathbf{x} = E(\mathbf{a}'\mathbf{x})) = 1$, worin c eine Konstante bedeutet. Gilt andererseits $P(\mathbf{a}'\mathbf{x} = c = E(\mathbf{a}'\mathbf{x})) = 1$ für einen Vektor $\mathbf{a} \neq \mathbf{0}$, so folgt $\int_{-\infty}^{\infty} (y - E(Y))^2 f(y) dy = 0$ für $Y = \mathbf{a}'\mathbf{x}$ und weiter $D(\mathbf{a}'\mathbf{x}) = 0$ oder $\mathbf{a}'D(\mathbf{x})\mathbf{a} = 0$. Dann ist wegen (233.5) $D(\mathbf{x})$ positiv semidefinit, so daß die Aussage folgt.

Zur Interpretation des Satzes (233.6) sei beispielsweise $P(a_1X_1 + a_2X_2 = c) = 1$ angenommen. Die beiden Zufallsvariablen X_1 und X_2 liegen dann mit einer Wahrscheinlichkeit von Eins auf einer Geraden, so daß lediglich eine Zufallsvariable zu berücksichtigen wäre, wenn die Gerade als Koordinatenachse gewählt würde.

b) Korrelationsmatrix

Faßt man die in (232.9) definierten Korrelationskoeffizienten ρ_{ij} in der *Korrelationsmatrix* $\mathbf{R} = (\rho_{ij})$ mit

$$\mathbf{R} = \begin{vmatrix} 1 & \rho_{12} & \dots & \rho_{1n} \\ \rho_{21} & 1 & \dots & \rho_{2n} \\ \dots & \dots & \dots & \dots \\ \rho_{n1} & \rho_{n2} & \dots & 1 \end{vmatrix} \quad (233.7)$$

zusammen, so gilt

$$\mathbf{R} = \mathbf{F}D(\mathbf{x})\mathbf{F} \quad \text{mit} \quad \mathbf{F} = \text{diag}(1/\sigma_1, \dots, 1/\sigma_n). \quad (233.8)$$

Für die Korrelationsmatrix gilt der

Satz: Die Korrelationsmatrix ist positiv definit oder positiv semidefinit. (233.9)

Beweis: Da in (232.9) $\sigma_i^2 > 0$ vorausgesetzt wurde, besitzt die Matrix \mathbf{F} in (233.8) vollen Rang, so daß die Korrelationsmatrix \mathbf{R} in Abhängigkeit von $D(\mathbf{x})$ wegen (143.7) und (233.5) positiv definit oder positiv semidefinit ist.

Satz: Die Korrelationsmatrix des $n \times 1$ Zufallsvektors $\mathbf{x} = |X_1, \dots, X_n|'$ ist genau dann positiv semidefinit, wenn mit einer Wahrscheinlichkeit von Eins eine lineare Beziehung zwischen den Zufallsvariablen X_1, \dots, X_n besteht. (233.10)

Beweis: Ist die Kovarianzmatrix $D(\mathbf{x})$ des Zufallsvektors \mathbf{x} positiv semidefinit, dann ist auch die Korrelationsmatrix \mathbf{R} positiv semidefinit, wie sich aus dem Beweis von (233.9) ergibt. Die Aussage folgt dann mit (233.6).

c) Kovarianzen von Zufallsvektoren

Die mit (233.1) definierte $m \times m$ Kovarianzmatrix $D(\mathbf{z})$ des $m \times 1$ Zufallsvektors \mathbf{z} werde in die folgenden vier Blöcke zerlegt

$$D(\mathbf{z}) = \left| \begin{array}{ccc|ccc} V(Z_1) & \dots & C(Z_1, Z_n) & C(Z_1, Z_{n+1}) & \dots & C(Z_1, Z_m) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ C(Z_n, Z_1) & \dots & V(Z_n) & C(Z_n, Z_{n+1}) & \dots & C(Z_n, Z_m) \\ \hline C(Z_{n+1}, Z_1) & \dots & C(Z_{n+1}, Z_n) & V(Z_{n+1}) & \dots & C(Z_{n+1}, Z_m) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ C(Z_m, Z_1) & \dots & C(Z_m, Z_n) & C(Z_m, Z_{n+1}) & \dots & V(Z_m) \end{array} \right|.$$

Dieser Zerlegung entsprechend soll der Zufallsvektor \mathbf{z} nun mit $\mathbf{z} = |\mathbf{x}', \mathbf{y}'|'$ und $m = n + p$ in den $n \times 1$ Vektor \mathbf{x} und den $p \times 1$ Vektor \mathbf{y} aufgeteilt werden. Für die Kovarianzmatrix $D(\mathbf{z}) = D(|\mathbf{x}', \mathbf{y}'|')$ folgt dann

$$D\left(\begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array} \right) = \left| \begin{array}{ccc|ccc} V(X_1) & \dots & C(X_1, X_n) & C(X_1, Y_1) & \dots & C(X_1, Y_p) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ C(X_n, X_1) & \dots & V(X_n) & C(X_n, Y_1) & \dots & C(X_n, Y_p) \\ \hline C(Y_1, X_1) & \dots & C(Y_1, X_n) & V(Y_1) & \dots & C(Y_1, Y_p) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ C(Y_p, X_1) & \dots & C(Y_p, X_n) & C(Y_p, Y_1) & \dots & V(Y_p) \end{array} \right|.$$

Auf diese Weise lassen sich aufgrund von (232.5) auch Kovarianzen von Zufallsvektoren einführen.

Definition: Wird der $m \times 1$ Zufallsvektor \mathbf{z} mit der multivariaten Verteilung $f(z_1, \dots, z_m)$ in den $n \times 1$ Vektor \mathbf{x} und den $p \times 1$ Vektor \mathbf{y} aufgeteilt, wobei $\mathbf{z} = |\mathbf{x}', \mathbf{y}'|'$ und $m = n + p$ gelte, dann bezeichnet $C(\mathbf{x}, \mathbf{y})$

$$C(\mathbf{x}, \mathbf{y}) = (C(X_i, Y_j)) = E((\mathbf{x} - E(\mathbf{x}))(\mathbf{y} - E(\mathbf{y}))')$$

die $n \times p$ Kovarianzmatrix der Zufallsvektoren \mathbf{x} und \mathbf{y} . (233.11)

Wie gezeigt, ergibt sich für den Zufallsvektor \mathbf{z} mit $\mathbf{z} = |\mathbf{x}', \mathbf{y}'|'$ aus (233.11) die Kovarianzmatrix $D(\mathbf{z})$ wegen $C(\mathbf{x}, \mathbf{x}) = D(\mathbf{x})$ und $C(\mathbf{y}, \mathbf{y}) = D(\mathbf{y})$ zu

$$D(\mathbf{z}) = D\left(\begin{array}{c} \mathbf{x} \\ \mathbf{y} \end{array} \right) = \left| \begin{array}{cc} D(\mathbf{x}) & C(\mathbf{x}, \mathbf{y}) \\ C(\mathbf{y}, \mathbf{x}) & D(\mathbf{y}) \end{array} \right|. \tag{233.12}$$

Gilt wiederum $\mathbf{z} = |\mathbf{x}', \mathbf{y}'|'$ und sind \mathbf{x} sowie \mathbf{y} zwei $n \times 1$ Zufallsvektoren und $\mathbf{u} = \mathbf{x} - \mathbf{y}$ der $n \times 1$ Zufallsvektor ihrer Differenz, dann folgt die Kovarianzmatrix $D(\mathbf{u})$ mit

$$\mathbf{u} = \mathbf{x} - \mathbf{y} = \begin{vmatrix} \mathbf{I} & -\mathbf{I} \\ \mathbf{x} & \mathbf{y} \end{vmatrix}$$

aus (233.2) und (233.12)

$$D(\mathbf{x} - \mathbf{y}) = D(\mathbf{x}) - C(\mathbf{x}, \mathbf{y}) - C(\mathbf{y}, \mathbf{x}) + D(\mathbf{y}) . \quad (233.13)$$

Sind \mathbf{x} und \mathbf{y} voneinander unabhängige Zufallsvektoren, so daß nach (232.7) $C(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ und $C(\mathbf{y}, \mathbf{x}) = \mathbf{0}$ gelten, ergibt sich anstelle von (233.13)

$$D(\mathbf{x} - \mathbf{y}) = D(\mathbf{x}) + D(\mathbf{y}) . \quad (233.14)$$

Eine Verallgemeinerung von (233.2) erhält man mit dem

Satz: Es seien \mathbf{x} und \mathbf{y} zwei Zufallsvektoren mit einer gemeinsamen multivariaten Verteilung und $\mathbf{A}, \mathbf{B}, \mathbf{a}, \mathbf{b}$ Matrizen und Vektoren von Konstanten, dann gilt

$$C(\mathbf{A}\mathbf{x} + \mathbf{a}, \mathbf{B}\mathbf{y} + \mathbf{b}) = \mathbf{A}C(\mathbf{x}, \mathbf{y})\mathbf{B}' . \quad (233.15)$$

Beweis: Aufgrund von (233.11) verläuft der Beweis analog zu dem von (233.2).

Zerlegt man zum Beispiel in der Kovarianzmatrix $C(\mathbf{u}, \mathbf{z})$ der Zufallsvektoren \mathbf{u} und \mathbf{z} den Vektor \mathbf{u} in $\mathbf{u} = |\mathbf{x}', \mathbf{y}'|'$, so folgt aus (233.11)

$$C(\mathbf{u}, \mathbf{z}) = \begin{vmatrix} C(\mathbf{x}, \mathbf{z}) \\ C(\mathbf{y}, \mathbf{z}) \end{vmatrix} . \quad (233.16)$$

Besitzen nun die Zufallsvektoren \mathbf{x} und \mathbf{y} gleiche Dimensionen, dann folgt für $C(\mathbf{v}, \mathbf{z})$ mit

$$\mathbf{v} = \mathbf{x} - \mathbf{y} = \begin{vmatrix} \mathbf{I} & -\mathbf{I} \\ \mathbf{x} & \mathbf{y} \end{vmatrix}$$

aus (233.15) und (233.16)

$$C(\mathbf{x} - \mathbf{y}, \mathbf{z}) = C(\mathbf{x}, \mathbf{z}) - C(\mathbf{y}, \mathbf{z}) . \quad (233.17)$$

d) Gewichtsmatrix

Es sei $\mathbf{x} = |X_1, \dots, X_n|'$ ein $n \times 1$ Zufallsvektor und $D(\mathbf{x}) = \mathbf{\Sigma}$ seine positiv definite Kovarianzmatrix. Dann bezeichnet man die $n \times n$ Matrix \mathbf{P}

$$\mathbf{P} = c\mathbf{\Sigma}^{-1} , \quad (233.18)$$

worin c eine Konstante bedeutet, als *Gewichtsmatrix* und das Diagonalelement p_{ii} von \mathbf{P} mit $\mathbf{P} = (p_{ij})$ als das *Gewicht* der Zufallsvariablen X_i . Sind die Zufallsvariablen voneinander unabhängig, vereinfacht sich die Kovarianzmatrix wegen (232.7) zur Diagonalmatrix $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$. Das Gewicht p_i von X_i ergibt sich dann zu

$$p_i = p_{ii} = c/\sigma_i^2. \tag{233.19}$$

Je kleiner die Varianzen der Zufallsvariablen X_i ausfallen, desto größer sind ihre Gewichte und je höher ist die Präzision für Zufallsvariablen, die Messungen repräsentieren.

e) Beispiele zum Fehlerfortpflanzungsgesetz und ein Beispiel zum Gewicht

1. Beispiel: Die n Zufallsvariablen X_i des $n \times 1$ Zufallsvektors \mathbf{x} mit $\mathbf{x} = |X_1, \dots, X_n|'$ seien voneinander unabhängig, so daß mit (232.7) folgt $D(\mathbf{x}) = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$. Das Fehlerfortpflanzungsgesetz für Funktionen voneinander unabhängiger Zufallsvariablen ergibt sich dann aus (233.2) zu

$$D(\mathbf{y}) = \mathbf{A} \text{diag}(\sigma_1^2, \dots, \sigma_n^2) \mathbf{A}'.$$

Bestehen nichtlineare Beziehungen zwischen den Zufallsvektoren \mathbf{x} und \mathbf{y} , folgt mit (233.4)

$$D(\mathbf{y}) = \begin{vmatrix} \sigma_1^2 \frac{\partial h_1}{\partial x_1} \Big|_{\mathbf{x}=\mathbf{x}_0} & \dots & \sigma_n^2 \frac{\partial h_1}{\partial x_n} \Big|_{\mathbf{x}=\mathbf{x}_0} \\ \dots & \dots & \dots \\ \sigma_1^2 \frac{\partial h_m}{\partial x_1} \Big|_{\mathbf{x}=\mathbf{x}_0} & \dots & \sigma_n^2 \frac{\partial h_m}{\partial x_n} \Big|_{\mathbf{x}=\mathbf{x}_0} \end{vmatrix} \mathbf{A}'$$

und somit die Varianz $V(Y_i)$ der Zufallsvariablen Y_i mit $\mathbf{y} = (Y_i)$

$$V(Y_i) = \sigma_1^2 \left(\frac{\partial h_i}{\partial x_1} \Big|_{\mathbf{x}=\mathbf{x}_0} \right)^2 + \dots + \sigma_n^2 \left(\frac{\partial h_i}{\partial x_n} \Big|_{\mathbf{x}=\mathbf{x}_0} \right)^2 \tag{233.20}$$

und die Kovarianz $C(Y_i, Y_j)$ entsprechend. In der Ausgleichsrechnung bezeichnet man (233.20) als Fehlerfortpflanzungsgesetz voneinander unabhängiger Beobachtungen.

2. Beispiel: Es seien X_i mit $V(X_i) = \sigma_i^2$ die n voneinander unabhängigen Zufallsvariablen, die durch die n Messungen beispielsweise einer Strecke definiert werden. Die Varianz des Mittels $X = \frac{1}{n} \sum_{i=1}^n X_i$ folgt dann aus (233.20) mit

$$V(X) = \frac{1}{n^2} \sigma_1^2 + \dots + \frac{1}{n^2} \sigma_n^2$$

und für $\sigma_i^2 = \sigma^2$

$$V(X) = \frac{\sigma^2}{n} .$$

Für diesen Fall ist also das mit (233.19) definierte Gewicht $c/V(X) = cn/\sigma^2$ des Mittels X proportional zur Anzahl n der Messungen.

3. Beispiel: Durch die Messungen der drei Seiten eines Dreiecks sei der 3×1 Zufallsvektor \mathbf{x} definiert, wobei die Beobachtungen die folgenden Werte x_1, x_2, x_3 für \mathbf{x} in der Dimension Millimeter ergeben haben

$$x_1 = 271\,346, \quad x_2 = 389\,423, \quad x_3 = 522\,118 .$$

Für die Kovarianzmatrix $D(\mathbf{x})$ gelte in der Dimension Millimeter²

$$D(\mathbf{x}) = \begin{vmatrix} 1,0 & 0,5 & 0,9 \\ 0,5 & 2,3 & 2,1 \\ 0,9 & 2,1 & 4,0 \end{vmatrix} .$$

Gesucht ist die Kovarianzmatrix der beiden der Seite x_3 anliegenden Winkel α und β in der Dimension 0,0001 gon. (Der Mittelpunktswinkel eines Kreises beträgt 400 gon.) Mit $\mathbf{y} = |\alpha, \beta|'$ ist also $D(\mathbf{y})$ zu berechnen.

Mit Hilfe des Kosinussatzes erhält man

$$x_1^2 = x_2^2 + x_3^2 - 2x_2x_3 \cos \alpha \quad \text{und} \quad x_2^2 = x_1^2 + x_3^2 - 2x_1x_3 \cos \beta ,$$

so daß für die Koeffizientenmatrix \mathbf{A} mit (233.4) folgt

$$\mathbf{A} = \begin{vmatrix} \left. \frac{\partial \alpha}{\partial x_1} \right|_{\mathbf{x}=\mathbf{x}_0} & \left. \frac{\partial \alpha}{\partial x_2} \right|_{\mathbf{x}=\mathbf{x}_0} & \left. \frac{\partial \alpha}{\partial x_3} \right|_{\mathbf{x}=\mathbf{x}_0} \\ \left. \frac{\partial \beta}{\partial x_1} \right|_{\mathbf{x}=\mathbf{x}_0} & \left. \frac{\partial \beta}{\partial x_2} \right|_{\mathbf{x}=\mathbf{x}_0} & \left. \frac{\partial \beta}{\partial x_3} \right|_{\mathbf{x}=\mathbf{x}_0} \end{vmatrix} .$$

Die Differentialquotienten erhält man einfach mit dem vollständigen Differential, das für den ersten Ausdruck lautet

$$x_1 dx_1 = (x_2 - x_3 \cos \alpha) dx_2 + (x_3 - x_2 \cos \alpha) dx_3 + x_2 x_3 \sin \alpha d\alpha .$$

Als Näherungswerte \mathbf{x}_0 wählt man die für \mathbf{x} gemessenen Werte. Die Elemente der Matrix \mathbf{A} in der Dimension 0,0001 gon/mm folgen dann aus

$$\begin{aligned} \mathbf{A} &= \frac{400 \cdot 10^4}{2\pi} \begin{vmatrix} \frac{x_1}{x_2 x_3 \sin \alpha} & \frac{x_3 \cos \alpha - x_2}{x_2 x_3 \sin \alpha} & \frac{x_2 \cos \alpha - x_3}{x_2 x_3 \sin \alpha} \\ \frac{x_3 \cos \beta - x_1}{x_1 x_3 \sin \beta} & \frac{x_2}{x_1 x_3 \sin \beta} & \frac{x_1 \cos \beta - x_3}{x_1 x_3 \sin \beta} \end{vmatrix} \\ &= \begin{vmatrix} 1,677 & 0,376 & -1,152 \\ 0,539 & 2,407 & -2,076 \end{vmatrix} . \end{aligned}$$

Mit den gemessenen Werten erhält man α und β in der Dimension 0,0001 gon zu $\alpha = 338\,124$ und $\beta = 518\,087$, und $D(\mathbf{y})$ folgt aus (233.2) zu

$$D(\mathbf{y}) = \begin{vmatrix} 3,78 & 3,52 \\ 3,52 & 9,15 \end{vmatrix}.$$

Die Standardabweichung σ_α des Winkels α ergibt sich daher in der Dimension 0,0001 gon zu $\sigma_\alpha = \sqrt{3,78}$ und die von β zu $\sigma_\beta = \sqrt{9,15}$.

234 Momenterzeugende Funktion

Mit Hilfe der momenterzeugenden Funktion lassen sich Momente von Zufallsvariablen angeben und Verteilungen ableiten.

Definition: Es sei $\mathbf{x} = |X_1, \dots, X_n|'$ ein $n \times 1$ Zufallsvektor mit der multivariaten Verteilung $f(x_1, \dots, x_n)$ und $\mathbf{t} = |t_1, \dots, t_n|'$ ein $n \times 1$ Vektor von Konstanten. Dann bezeichnet man $M_{\mathbf{x}}(\mathbf{t})$

$$\begin{aligned} M_{\mathbf{x}}(\mathbf{t}) &= E(e^{\mathbf{t}'\mathbf{x}}) = E(\exp(t_1 X_1 + \dots + t_n X_n)) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(t_1 x_1 + \dots + t_n x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

als *momenterzeugende Funktion* des Zufallsvektors \mathbf{x} , falls der Erwartungswert für $-h < t_i < h$ mit $i \in \{1, \dots, n\}$ und ein gewisses $h > 0$ existiert. Ersetzt man in $\mathbf{t}'\mathbf{x}$ den Vektor \mathbf{x} durch Funktionen von \mathbf{x} , ergibt sich die momenterzeugende Funktion von Funktionen der Zufallsvariablen. (234.1)

Als *charakteristische Funktion*, die im folgenden nicht verwendet wird, bezeichnet man $E(\exp(i\mathbf{t}'\mathbf{x}))$ mit $i = \sqrt{-1}$.

Die Momente einer Zufallsvariablen erhält man aus der momenterzeugenden Funktion wie folgt.

Satz: Ist $M_{\mathbf{x}}(\mathbf{t})$ die momenterzeugende Funktion des Zufallsvektors \mathbf{x} , dann ergibt sich mit $k = \sum_{i=1}^n k_i$ und $k_i \in \mathbb{N}$ das k -te multivariate Moment zu

$$\mu_{x_1 \dots x_n}^{(k)} = \frac{\partial^k M_{\mathbf{x}}(\mathbf{t})}{\partial t_1^{k_1} \dots \partial t_n^{k_n}} \Big|_{\mathbf{t}=\mathbf{0}}. \quad (234.2)$$

Beweis: Durch k_1 -fache Differentiation von $M_{\mathbf{x}}(\mathbf{t})$ nach t_1 , k_2 -fache Differentiation von $M_{\mathbf{x}}(\mathbf{t})$ nach t_2 und so fort erhält man mit $k = \sum_{i=1}^n k_i$ aus (234.1)

$$\begin{aligned} \frac{\partial^k M_{\mathbf{x}}(\mathbf{t})}{\partial t_1^{k_1} \dots \partial t_n^{k_n}} &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{k_1} \dots x_n^{k_n} \exp(t_1 x_1 + \dots + t_n x_n) \\ &\quad f(x_1, \dots, x_n) dx_1 \dots dx_n, \end{aligned}$$

woraus mit $\mathbf{t} = \mathbf{0}$ wegen (232.1) das k -te multivariate Moment folgt.

Sind die Verteilungsfunktionen zweier Zufallsvektoren gleicher Dimensionen identisch, so sind ihre momenterzeugenden Funktionen identisch. Auch der umgekehrte Schluß ist zulässig, wovon im folgenden bei den Ableitungen von Verteilungen häufig Gebrauch gemacht wird.

Satz: Sind \mathbf{x}_1 und \mathbf{x}_2 zwei $n \times 1$ Zufallsvektoren mit den Verteilungsfunktionen $F_1(\mathbf{x})$ und $F_2(\mathbf{x})$ sowie den momenterzeugenden Funktionen $M_1(\mathbf{t})$ und $M_2(\mathbf{t})$, so ist genau dann $F_1(\mathbf{x}) = F_2(\mathbf{x})$, wenn $M_1(\mathbf{t}) = M_2(\mathbf{t})$ für $-h < t_i < h$ mit $i \in \{1, \dots, n\}$ und ein gewisses $h > 0$ gilt. (234.3)

Der Beweis des Satzes befindet sich bei (WILKS 1962, S.118 und 120).

Satz: Für den Zufallsvektor \mathbf{x} gelte $\mathbf{x} = |\mathbf{x}'_1, \mathbf{x}'_2|'$ und entsprechend $\mathbf{t} = |\mathbf{t}'_1, \mathbf{t}'_2|'$. Die Zufallsvektoren \mathbf{x}_1 und \mathbf{x}_2 sind genau dann voneinander unabhängig, wenn gilt

$$M_{\mathbf{x}_1, \mathbf{x}_2}(\mathbf{t}_1, \mathbf{t}_2) = M_{\mathbf{x}_1}(\mathbf{t}_1)M_{\mathbf{x}_2}(\mathbf{t}_2). \quad (234.4)$$

Beweis: Sind die Zufallsvektoren \mathbf{x}_1 und \mathbf{x}_2 voneinander unabhängig, zerfällt nach (228.2) ihre gemeinsame Verteilung in das Produkt ihrer Randverteilungen, so daß mit (234.1) ihre momenterzeugende Funktion sich aus dem Produkt der momenterzeugenden Funktionen $M_{\mathbf{x}_1}(\mathbf{t}_1)$ und $M_{\mathbf{x}_2}(\mathbf{t}_2)$ berechnet. Gilt andererseits $M_{\mathbf{x}_1, \mathbf{x}_2}(\mathbf{t}_1, \mathbf{t}_2) = M_{\mathbf{x}_1}(\mathbf{t}_1)M_{\mathbf{x}_2}(\mathbf{t}_2)$, erhält man aus (234.1)

$$\begin{aligned} & \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(\mathbf{t}'_1 \mathbf{x}_1 + \mathbf{t}'_2 \mathbf{x}_2) f(\mathbf{x}_1, \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(\mathbf{t}'_1 \mathbf{x}_1) g(\mathbf{x}_1) d\mathbf{x}_1 \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(\mathbf{t}'_2 \mathbf{x}_2) g(\mathbf{x}_2) d\mathbf{x}_2 \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(\mathbf{t}'_1 \mathbf{x}_1 + \mathbf{t}'_2 \mathbf{x}_2) g(\mathbf{x}_1) g(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2. \end{aligned}$$

Hieraus folgt wegen (234.3) $F(\mathbf{x}_1, \mathbf{x}_2) = G(\mathbf{x}_1, \infty, \dots, \infty)G(\infty, \dots, \infty, \mathbf{x}_2)$ und mit (228.2) die Aussage.

24 Univariate Verteilungen

241 Normalverteilung

Die univariate Normalverteilung und ihre multivariate Verallgemeinerung stellen die wichtigsten Verteilungen stetiger Zufallsvariablen dar.

Definition: Die Zufallsvariable X bezeichnet man als *normalverteilt* mit den Parametern μ und σ^2 , abgekürzt geschrieben $X \sim N(\mu, \sigma^2)$, wenn ihre Dichte $f(x)$ gegeben ist durch

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad \text{für} \quad -\infty < x < \infty. \quad (241.1)$$

In Abbildung 241-1 sind die Dichten der Normalverteilung für zwei Werte von σ gezeichnet. Die Variation von μ verschiebt den Graphen der Normalverteilung lediglich entlang der x -Achse.

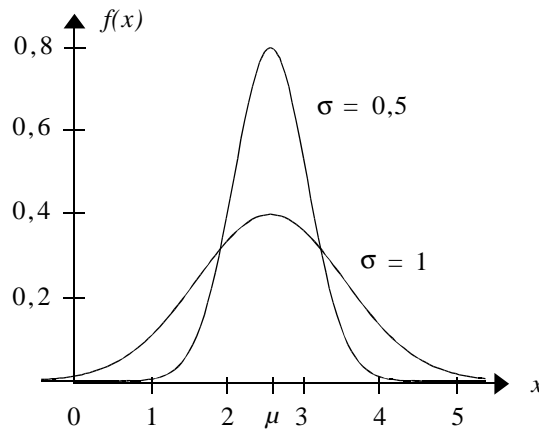


Abb. 241-1: Normalverteilungen

Die Bedingung (223.10) ist für die Normalverteilung erfüllt, denn zum einen ist $f(x) \geq 0$ und zum anderen gilt

$$A = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = 1, \quad (241.2)$$

denn substituiert man $y = (x - \mu)/\sigma$, so daß $dy = dx/\sigma$ folgt, erhält man

$$A = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1. \quad (241.3)$$

Anstatt zu zeigen, daß $A = 1$ gilt, wird $A^2 = 1$ bewiesen, woraus $A = 1$ wegen $e^{-y^2/2} > 0$ folgt. Man erhält

$$\begin{aligned} A^2 &= \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \right) \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz \right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(y^2+z^2)/2} dy dz. \end{aligned}$$

Mit den Polarkoordinaten $y = r \cos \alpha$, $z = r \sin \alpha$ folgt

$$A^2 = \frac{1}{2\pi} \int_0^\infty \int_0^{2\pi} r e^{-r^2/2} d\alpha dr = \int_0^\infty r e^{-r^2/2} dr = \left[-e^{-r^2/2} \right]_0^\infty = 1.$$

Gilt $X \sim N(\mu, \sigma^2)$, so besitzt die Zufallsvariable X den Erwartungswert μ und die Varianz σ^2 . Dies wird nicht hier, sondern allgemeiner in (252.2) für die multivariate Normalverteilung bewiesen, die die univariate Normalverteilung einschließt. Weitere Eigenschaften der univariaten Normalverteilung werden ebenfalls im Kapitel 25 im Zusammenhang mit der multivariaten Normalverteilung behandelt.

Die Bedeutung der Normalverteilung in der Statistik ergibt sich aus dem *zentralen Grenzwertsatz* (CRAMÉR 1946, S.214; WILKS 1962, S.257). Er besagt, daß für n voneinander unabhängige Zufallsvariable mit beliebigen Verteilungen unter gewissen, recht allgemeinen Bedingungen die Verteilung der Summe dieser Zufallsvariablen asymptotisch gegen eine Normalverteilung strebt, wenn n gegen unendlich geht. Für die im Kapitel 25 zu behandelnde multivariate Normalverteilung gilt ein entsprechender Satz (CRAMÉR 1946, S.316). Da man sich Zufallsvariable häufig aus der Summe sehr vieler voneinander unabhängiger Zufallsvariablen unterschiedlicher Verteilungen zusammengesetzt denken kann, beispielsweise die aus einer Vielzahl von Geräteeinflüssen resultierende elektro-optische Entfernungsmessung, lassen sich bei vielen praktischen Anwendungen die Zufallsvariablen als normalverteilt annehmen.

Wie der Vergleich von (241.2) und (241.3) zeigt, wird durch die Substitution $y = (x - \mu)/\sigma$ die Normalverteilung $N(\mu, \sigma^2)$ auf die *standardisierte* Normalverteilung $N(0, 1)$ gebracht, die für eine normalverteilte Zufallsvariable Y mit dem Erwartungswert $\mu = 0$ und der Varianz $\sigma^2 = 1$ gilt. Für $Y \sim N(0, 1)$ ergibt sich aus (223.5) die Verteilungsfunktion $F(x; 0, 1)$ zu

$$F(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy \quad \text{für} \quad -\infty < x < \infty. \quad (241.4)$$

Werte für $F(x; 0, 1)$ findet man häufig tabuliert (z.B. FISHER und YATES 1963, S.45; PEARSON und HARTLEY 1976, *Vol.I*, S.110), eine graphische Darstellung ist in Abbildung 241-2 gegeben. Mit einer Genauigkeit von $1 \cdot 10^{-5}$

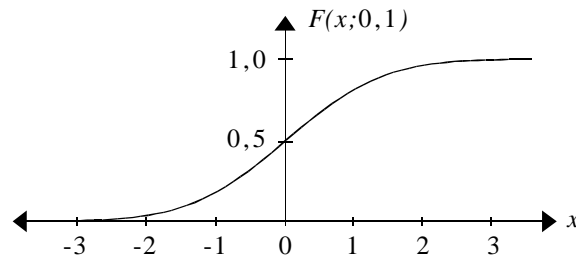


Abb. 241-2: Verteilungsfunktion

gilt für (241.4) die polynomiale Approximation (ABRAMOWITZ und STEGUN 1972, S.932)

$$F(x; 0, 1) = 1 - e^{-x^2/2}(a_1t + a_2t^2 + a_3t^3)/\sqrt{2\pi} \quad \text{für } x \geq 0 \quad (241.5)$$

mit $t = 1/(1 + px)$ und $p = 0,33267, a_1 = 0,4361836, a_2 = -0,1201676, a_3 = 0,9372980$. Ähnliche Approximationen befinden sich in (CARTA 1975).

Beispiel: Für die Zufallsvariable X mit $X \sim N(\mu, \sigma^2)$ sind die Wahrscheinlichkeiten $P(\mu - \sigma < X < \mu + \sigma)$ und $P(\mu - 3\sigma < X < \mu + 3\sigma)$ zu berechnen. Mit der Variablentransformation $y = (x - \mu)/\sigma$ ergibt sich wegen der Symmetrie der Normalverteilung

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= \frac{1}{\sqrt{2\pi}} \int_{-1}^1 e^{-y^2/2} dy \\ &= F(1; 0, 1) - (1 - F(1; 0, 1)), \end{aligned}$$

so daß man mit $F(1; 0, 1) = 0,8413$ erhält

$$P(\mu - \sigma < X < \mu + \sigma) = 0,683.$$

Analog berechnet sich mit $F(3; 0, 1) = 0,9987$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0,997.$$

Die Wahrscheinlichkeit ist also näherungsweise gleich Eins, daß die Werte einer normalverteilten Zufallsvariablen innerhalb des Intervalls liegen, das durch ihren Erwartungswert und die dreifache Standardabweichung definiert ist.

Den Wert x_α , für den

$$F(x_\alpha; 0, 1) = \alpha \quad (241.6)$$

gilt, bezeichnet man als das α -Fraktile, α -Quantil oder den unteren α -Prozentpunkt der standardisierten Normalverteilung. Entsprechend ist das $(1 - \alpha)$ -Fraktile oder der obere α -Prozentpunkt durch $F(x_{1-\alpha}; 0, 1) = 1 - \alpha$ definiert. Wegen der Symmetrie der Normalverteilung gilt $F(-x; 0, 1) = 1 - F(x; 0, 1)$ und somit wegen $F(x_{1-\alpha}; 0, 1) = 1 - F(x_\alpha; 0, 1)$

$$x_\alpha = -x_{1-\alpha} . \quad (241.7)$$

Die α -Fraktile lassen sich den angegebenen Tafeln entnehmen. Bei einer numerischen Berechnung gilt für x_α mit einer Genauigkeit von $4,5 \cdot 10^{-4}$ (ABRAMOWITZ und STEGUN 1972, S.933)

$$x_\alpha = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \quad \text{für } 0,5 \leq \alpha < 1 \quad (241.8)$$

mit

$$c_0 = 2,515\,517, \quad c_1 = 0,802\,853, \quad c_2 = 0,010\,328, \quad d_1 = 1,432\,788, \\ d_2 = 0,189\,269, \quad d_3 = 0,001\,308, \quad t = [\ln(1/(1 - \alpha)^2)]^{1/2} .$$

242 Herleitung der Normalverteilung als Verteilung von Beobachtungsfehlern

Wenn auch die Normalverteilung im vorangegangenen Kapitel definiert wurde, so lassen sich doch eine Reihe von Experimenten angeben, deren Ergeb-

Anordnung der δ		Häufigkeit	Beobachtungsfehler $\epsilon_i = \Sigma \delta$
positiv	negativ		
n	0	$\binom{n}{0}$	$\epsilon_0 = (n - 2 \cdot 0)\delta$
$n - 1$	1	$\binom{n}{1}$	$\epsilon_1 = (n - 2 \cdot 1)\delta$
...
$n - i$	i	$\binom{n}{i}$	$\epsilon_i = (n - 2i)\delta$
$n - i - 1$	$i + 1$	$\binom{n}{i+1}$	$\epsilon_{i+1} = (n - 2(i + 1))\delta$
...
0	n	$\binom{n}{n}$	$\epsilon_n = (n - 2n)\delta$

(242.1)

nisse normalverteilt sind (RAO 1973, S.153), wie beispielsweise die Beobachtungsfehler ϵ_i , die sich aus sehr vielen kleinen, voneinander unabhängigen Elementarfehlern zusammensetzen, die den gleichen Absolutbetrag δ besitzen und ebenso leicht positiv, wie negativ sein können (HAGEN 1837, S.34). Bei n Elementarfehlern ergeben sich die in Tabelle (242.1) angegebenen Anordnungen der δ , die Häufigkeit des Vorkommens der Beobachtungsfehler nach (136.2) und die Werte der Beobachtungsfehler ϵ_i . Bezeichnet man mit N die Summe aller Häufigkeiten, ergeben sich die relativen Häufigkeiten bei $n-i-1$ und $n-i$ positiven δ zu $\binom{n}{i+1}/N = \frac{n-i}{i+1} \binom{n}{i}/N$ und zu $\binom{n}{i}/N$. Als Mittel h_i und Differenz Δh_i dieser relativen Häufigkeiten erhält man

$$h_i = \frac{n+1}{2(i+1)N} \binom{n}{i} \quad \text{und} \quad \Delta h_i = \frac{n-2i-1}{(i+1)N} \binom{n}{i}.$$

Weiter ergeben sich das Mittel x_i und die Differenz Δx_i der Beobachtungsfehler ϵ_{i+1} und ϵ_i zu $x_i = (n-2i-1)\delta$ und $\Delta x_i = -2\delta$. Identifiziert man nun die Zufallsvariable X mit der Variablen, die die Werte x_i annimmt, so gilt für ihre Dichte $f(x)$

$$\frac{\Delta f(x)}{f(x)} = \frac{\Delta h_i}{h_i} = \frac{2(n-2i-1)}{n+1} = \frac{2x_i}{(n+1)\delta} = \frac{-x_i \Delta x_i}{(n+1)\delta^2}.$$

Geht man von der diskreten Zufallsvariablen zu einer stetigen Zufallsvariablen über, ergibt sich $df(x)/f(x) = -x dx / ((n+1)\delta^2)$ und durch Integration $\ln f(x) + c = -x^2 / (2(n+1)\delta^2)$. Setzt man $(n+1)\delta^2 = \sigma^2$ und für die Integrationskonstante $c = \ln(\sqrt{2\pi}\sigma)$, ergibt sich die Dichte der Normalverteilung $N(0, \sigma^2)$.

243 Gammaverteilung

Als weitere univariate Verteilung soll die Gammaverteilung behandelt werden, die für die im Kapitel 26 abgeleiteten Testverteilungen benötigt wird.

Definition: Die Zufallsvariable X besitzt die *Gammaverteilung* $G(b, p)$ mit den reellen Parametern b und p , also $X \sim G(b, p)$, falls ihre Dichte gegeben ist durch

$$f(x) = \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} \quad \text{für} \quad b > 0, p > 0, 0 < x < \infty$$

und $f(x) = 0$ für die übrigen Werte von x . (243.1)

Der Wert $\Gamma(p)$ ergibt sich aus der durch

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt \quad \text{für} \quad p > 0 \quad (243.2)$$

definierten *Gamma-Funktion*.

Die Gammaverteilung erfüllt (223.10), denn zum einen ist $f(x) \geq 0$ und zum anderen gilt

$$\int_0^{\infty} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} dx = 1, \quad (243.3)$$

da mit $y = bx$ und $dy = bdx$ sich wegen (243.2) $\int_0^{\infty} y^{p-1} e^{-y} dy / \Gamma(p) = 1$ ergibt.

Zur Auswertung der Gamma-Funktion (243.2) wird partiell integriert, und man erhält mit der Regel von de L'Hospital (SMIRNOW 1975, *Teil I*, S.167)

$$\Gamma(p) = \left[-t^{p-1} e^{-t} \right]_0^{\infty} + \int_0^{\infty} (p-1)t^{p-2} e^{-t} dt = (p-1) \int_0^{\infty} t^{p-2} e^{-t} dt,$$

woraus die Rekursionsformel folgt

$$\Gamma(p) = (p-1)\Gamma(p-1). \quad (243.4)$$

Ist p eine positive ganze Zahl, gilt $\Gamma(p) = (p-1) \dots 2\Gamma(1)$ und mit $\Gamma(1) = \int_0^{\infty} e^{-t} dt = \left[-e^{-t} \right]_0^{\infty} = 1$

$$\Gamma(p) = (p-1)! \quad \text{für } p \in \mathbb{N} \quad \text{und } p > 0. \quad (243.5)$$

Ist p ein Vielfaches von $1/2$, gilt wegen (243.4)

$$\Gamma\left(p + \frac{1}{2}\right) = \frac{(2p-1)(2p-3) \dots 5 \cdot 3 \cdot 1}{2^p} \sqrt{\pi} \quad \text{für } p \in \mathbb{N} \quad \text{und } p > 0 \quad (243.6)$$

mit $\Gamma(1/2) = \sqrt{\pi}$, denn mit der Substitution $t = y^2/2$ und $dt = ydy$ folgt $\Gamma(1/2) = \int_0^{\infty} t^{-1/2} e^{-t} dt = \sqrt{2} \int_0^{\infty} e^{-y^2/2} dy = \sqrt{\pi}$, da $\int_0^{\infty} e^{-y^2/2} dy / \sqrt{2\pi} = 1/2$ wegen (241.3) gilt. Näherungswerte für die Gammafunktion ergeben sich mit der Stirlingschen Näherung (HENRICI 1977, S.43)

$$\Gamma(p) \approx \sqrt{2\pi} e^{-p} p^{p-1/2}. \quad (243.7)$$

Für die momenterzeugende Funktion einer Zufallsvariablen mit der Gamma-Verteilung gilt der

Satz: Es sei $X \sim G(b, p)$, dann ergibt sich die momenterzeugende Funktion $M_x(t)$ von X zu

$$M_x(t) = (1 - t/b)^{-p} \quad \text{für } t < b. \quad (243.8)$$

Beweis: Mit (234.1) erhält man

$$M_x(t) = \int_0^\infty e^{tx} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} dx .$$

Substituiert man $y = bx$, folgt mit $dy = bdx$

$$\begin{aligned} M_x(t) &= \int_0^\infty \frac{1}{\Gamma(p)} e^{(ty/b-y)} y^{p-1} dy \\ &= (1-t/b)^{-p} \int_0^\infty \frac{(1-t/b)^p}{\Gamma(p)} y^{p-1} e^{-(1-t/b)y} dy , \end{aligned}$$

woraus mit (243.3) die Aussage folgt, da nach (243.1) $(1-t/b) > 0$ gelten muß.

Von der *reproduzierenden Eigenschaft* einer Verteilung spricht man, wenn die Verteilung der Summe unabhängiger Zufallsvariablen, deren Verteilungen der gleichen Klasse angehören, ebenfalls zu dieser Klasse von Verteilungen zählt. Die Gammaverteilung besitzt diese reproduzierende Eigenschaft, denn es gilt der

Satz: Die Zufallsvariablen X_i mit $X_i \sim G(b, p_i)$ und $i \in \{1, \dots, k\}$ seien voneinander unabhängig, dann gilt

$$X_1 + \dots + X_k \sim G(b, \sum_{i=1}^k p_i) . \quad (243.9)$$

Beweis: Da die Zufallsvariablen X_1, \dots, X_k voneinander unabhängig sind, zerfällt nach (228.2) ihre gemeinsame Verteilung in das Produkt der Gammaverteilungen $G(b, p_i)$. Die momenterzeugende Funktion $M_{\Sigma X}(t)$ von $X_1 + \dots + X_k$ berechnet sich dann aus (234.1) mit (243.8) zu

$$M_{\Sigma X}(t) = M_{X_1}(t) \dots M_{X_k}(t) = (1-t/b)^{-\Sigma p_i} .$$

Das ist aber die momenterzeugende Funktion der Gammaverteilung $G(b, \Sigma p_i)$, so daß mit (234.3) die Aussage folgt.

Für die Zufallsvariable X mit $X \sim G(b, p)$ ergibt sich aus (223.5) die Verteilungsfunktion $F(G; b, p)$ zu

$$F(G; b, p) = \frac{b^p}{\Gamma(p)} \int_0^G x^{p-1} e^{-bx} dx . \quad (243.10)$$

Durch partielle Integration erhält man

$$\begin{aligned} F(G; b, p) &= \frac{1}{p} \left(\frac{b^p}{\Gamma(p)} [x^p e^{-bx}]_0^G + \frac{b^{p+1}}{\Gamma(p)} \int_0^G x^p e^{-bx} dx \right) \\ &= \frac{b^p G^p e^{-bG}}{\Gamma(p+1)} \left(1 + \sum_{j=1}^{\infty} \frac{(bG)^j}{(p+1)(p+2)\dots(p+j)} \right). \end{aligned} \quad (243.11)$$

Diese Reihe konvergiert nach dem Quotientenkriterium (BLATTER 1974, I, S.93), denn mit wachsendem j gibt es eine Zahl q , für die $bG/(p+j) \leq q < 1$ gilt.

244 Herleitung der Gammaverteilung als Verteilung von Ankunftszeiten

Die im vorangegangenen Kapitel definierte Gammaverteilung läßt sich auch für die Verteilung der Ankunftszeiten von Fahrzeugen ableiten. Hierfür soll zunächst die Verteilung der Anzahl von Fahrzeugen angegeben werden, die in einem festen Zeitintervall von t Sekunden an einem bestimmten Ort eintreffen. Beträgt die Wahrscheinlichkeit p , daß ein Fahrzeug in der Zeiteinheit von einer Sekunde auftaucht, und erscheint nicht mehr als ein Fahrzeug pro Sekunde, dann folgt für die Anzahl x der Fahrzeuge in $n = t$ angenommenen, voneinander unabhängigen Versuchen aus der Binomialverteilung (224.1) die Dichte

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{für } x \in \{0, 1, \dots, n\}.$$

Werden sehr viel kleinere Zeiteinheiten als eine Sekunde für die einzelnen Versuche gewählt, das Zeitintervall t des Experimentes aber unverändert gelassen, folgt mit $n \rightarrow \infty$, $p \rightarrow 0$ und $np = \lambda t$ für die Dichte $f(x)$ die Poisson-Verteilung (224.2)

$$\begin{aligned} f(x) &= \exp(-\lambda t) (\lambda t)^x / x! \\ &\quad \text{für } x \in \{0, 1, 2, \dots\} \quad \text{und } \lambda > 0, t > 0. \end{aligned} \quad (244.1)$$

Aus dieser Verteilung ergibt sich die Wahrscheinlichkeit, daß in einem Zeitintervall t an einem Ort x Fahrzeuge eintreffen. Da nach (224.4) die Größe λt den Erwartungswert der nach (244.1) verteilten Zufallsvariablen angibt, bedeutet λ die durchschnittlich zu erwartende Anzahl der ankommenden Fahrzeuge pro Zeiteinheit, in der t definiert ist.

Mit (244.1) wird jetzt die Verteilung $f_T(t)$ der Zeit T bis zur Ankunft des

ersten Fahrzeuges abgeleitet. Die Wahrscheinlichkeit, daß T das Zeitintervall t , dessen Länge beliebig ist, überschreitet, gleicht der Wahrscheinlichkeit, daß keine Ankunft in dem Zeitintervall t stattgefunden hat. Diese Wahrscheinlichkeit ergibt sich aus (244.1) für $x = 0$, so daß $P(T > t) = f(0) = \exp(-\lambda t)$ gilt. Damit erhält man die Verteilungsfunktion $F(t)$ von T nach (213.4) und (223.6) zu $F(t) = P(T < t) = 1 - P(T > t) = 1 - \exp(-\lambda t)$ und die Dichte $f_T(t)$ von T nach (223.9) zu

$$f_T(t) = dF(t)/dt = \lambda \exp(-\lambda t) \quad \text{für } \lambda > 0, t > 0. \quad (244.2)$$

Dies ist nach (243.1) die Gamma-Verteilung $T \sim G(\lambda, 1)$. Da die Zeit T bis zur Ankunft des ersten Fahrzeuges kein absolutes Zeitmaß, sondern ein Zeitintervall angibt, gilt (244.2) auch für die Zeit zwischen zwei Ankünften, so daß auch die Verteilung für die Zeit $T_s = T_1 + \dots + T_k$ zwischen k Ankünften angebar ist, wobei $T_i \sim G(\lambda, 1)$ mit $i \in \{1, \dots, k\}$ gilt und die einzelnen Ankunftszeiten voneinander unabhängig sind. Mit (243.9) folgt dann $T_s \sim G(\lambda, k)$.

245 Beta-Verteilung

Als letzte univariate Verteilung, die im folgenden benötigt wird, soll die Beta-Verteilung behandelt werden.

Satz: Die Zufallsvariablen Y und Z mit $Y \sim G(b, \alpha)$ und $Z \sim G(b, \beta)$ seien voneinander unabhängig, dann besitzt die Zufallsvariable $X = Y/(Y + Z)$ die Beta-Verteilung $B(\alpha, \beta)$ mit den reellen Parametern α und β , also $X \sim B(\alpha, \beta)$, und der Dichte

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad \text{für } 0 < x < 1$$

und $f(x) = 0$ für die übrigen Werte von x . (245.1)

Beweis: Wegen der Unabhängigkeit ergibt sich die gemeinsame Verteilung $f(y, z)$ von Y und Z nach (228.2) mit (243.1) zu

$$f(y, z) = \frac{b^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-by-bz} y^{\alpha-1} z^{\beta-1}.$$

Für die Transformation $y = r \sin^2 \psi$, $z = r \cos^2 \psi$ mit $0 < r < \infty$, $0 < \psi < \pi/2$ ergibt sich aus (229.1) mit

$$\det \mathbf{J} = \det \begin{vmatrix} \partial y / \partial r & \partial y / \partial \psi \\ \partial z / \partial r & \partial z / \partial \psi \end{vmatrix} = -2r \sin \psi \cos \psi$$

die Verteilung

$$f(r, \psi) = \frac{2b^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-br} r^{\alpha+\beta-1} (\sin \psi)^{2\alpha-1} (\cos \psi)^{2\beta-1} .$$

Hieraus wird nach (226.4) die Randverteilung $g(\psi)$ berechnet. Mit $br = u$ und $bdr = du$ erhält man wegen (243.2)

$$b^{\alpha+\beta} \int_0^\infty e^{-br} r^{\alpha+\beta-1} dr = b^{\alpha+\beta} \int_0^\infty e^{-u} \frac{u^{\alpha+\beta-1}}{b^{\alpha+\beta}} du = \Gamma(\alpha + \beta) ,$$

so daß folgt

$$g(\psi) = \frac{2\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (\sin \psi)^{2\alpha-1} (\cos \psi)^{2\beta-1} \quad \text{für } 0 < \psi < \frac{\pi}{2} .$$

Wegen $y = r \sin^2 \psi$ und $z = r \cos^2 \psi$ erhält man $x = y/(y+z) = \sin^2 \psi$. Mit der Substitution $x = \sin^2 \psi$ und $\det \mathbf{J} = 1/(2 \sin \psi \cos \psi)$ in (229.1) wegen $dx/d\psi = 2 \sin \psi \cos \psi$ folgt dann schließlich aus der Dichte $g(\psi)$ die Dichte der Betaverteilung.

Die momenterzeugende Funktion der Betaverteilung besitzt keine einfache Form (JOHNSON und KOTZ 1970, Vol. 2, S.40), aber die Momente lassen sich bequem angeben.

Satz: Es sei $X \sim B(\alpha, \beta)$, dann gilt für das k -te Moment $\mu_x^{(k)}$ von X

$$\mu_x^{(k)} = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + k)}{\Gamma(\alpha + \beta + k)\Gamma(\alpha)} . \quad (245.2)$$

Beweis: Mit (232.3) und (245.1) folgt

$$\mu_x^{(k)} = E(X^k) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{k+\alpha-1} (1-x)^{\beta-1} dx .$$

Da $f(x)$ in (245.1) eine Dichte angibt, erhält man mit (223.10)

$$\frac{\Gamma(\alpha + \beta + k)}{\Gamma(\alpha + k)\Gamma(\beta)} \int_0^1 x^{k+\alpha-1} (1-x)^{\beta-1} dx = 1 ,$$

woraus die Aussage folgt.

Die Verteilungsfunktion der Betaverteilung bezeichnet man als *unvollständige Betafunktion* $I_x(\alpha, \beta)$

$$I_x(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt . \quad (245.3)$$

Sie läßt sich mit Hilfe der hypergeometrischen Funktion als unendliche Reihe darstellen (HENRICI 1977, S.62), so daß mit Eulers erster Identität für die hypergeometrische Funktion (HENRICI 1977, S.160) die Reihenentwicklung folgt (ABRAMOWITZ und STEGUN 1972, S.944)

$$I_x(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)x^\alpha}{\Gamma(\alpha)\Gamma(\beta)\alpha} (1-x)^\beta \left(1 + \frac{\alpha + \beta}{\alpha + 1}x + \frac{(\alpha + \beta + 1)(\alpha + \beta)}{(\alpha + 2)(\alpha + 1)}x^2 + \dots + \frac{(\alpha + \beta + n - 1) \dots (\alpha + \beta + 1)(\alpha + \beta)}{(\alpha + n) \dots (\alpha + 2)(\alpha + 1)}x^n + \dots \right) \quad \text{für } 0 < x < 1. \quad (245.4)$$

Die Reihe konvergiert besonders rasch für kleine Werte von x . Da

$$I_x(\alpha, \beta) = 1 - I_{1-x}(\beta, \alpha) \quad (245.5)$$

gilt, läßt sich erreichen, daß in (245.4) immer $x \leq 0,5$ ist. Die Beziehung (245.5) ergibt sich aus (245.3) durch die Variablentransformation $u = 1 - t$ mit $du = -dt$, so daß folgt

$$I_x(\alpha, \beta) = -\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_1^{1-x} (1-u)^{\alpha-1} u^{\beta-1} du = 1 - I_{1-x}(\beta, \alpha).$$

Tafeln für die unvollständige Betafunktion befinden sich in (PEARSON und HARTLEY 1976, Vol.I, S.150).

25 Multivariate Normalverteilung

251 Definition und Herleitung

Die univariate Normalverteilung (241.1) für eine Zufallsvariable läßt sich derart verallgemeinern, daß die multivariate Normalverteilung für mehrere, gemeinsam verteilte Zufallsvariablen gewonnen wird. Zur Vereinfachung der Bezeichnung wird im folgenden, wie bereits im Kapitel 225 erwähnt, die Unterscheidung zwischen den Zufallsvariablen X_i und den Werten x_i , die sie annehmen können, aufgegeben und die Zufallsvariablen und ihre Werte mit x_i bezeichnet. Wo eine Unterscheidung notwendig ist, wird sie wieder vorgenommen.

Definition: Der $n \times 1$ Zufallsvektor $\mathbf{x} = |x_1, \dots, x_n|'$ besitzt die *multivariate Normalverteilung* $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit dem $n \times 1$ Vektor $\boldsymbol{\mu}$ und der $n \times n$ positiv definiten Matrix $\boldsymbol{\Sigma}$ als Parameter, also $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, falls die Dichte $f(\mathbf{x})$ von \mathbf{x} gegeben ist durch

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \boldsymbol{\Sigma})^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \quad (251.1)$$

Im folgenden Kapitel wird gezeigt, daß durch $\boldsymbol{\mu}$ der Vektor der Erwartungswerte von \boldsymbol{x} und durch $\boldsymbol{\Sigma}$ die Kovarianzmatrix von \boldsymbol{x} bestimmt werden.

Die Dichte der multivariaten Normalverteilung ist immer positiv, da die Exponentialfunktion keine negativen Werte annehmen kann und da für die positiv definite Matrix $\boldsymbol{\Sigma}$ wegen (136.11) und (143.4) $\det \boldsymbol{\Sigma} > 0$ gilt. Auch die zweite Bedingung in (225.6) ist für die multivariate Normalverteilung erfüllt, denn es gilt

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right) dx_1 \dots dx_n \\ = (2\pi)^{n/2} (\det \boldsymbol{\Sigma})^{1/2} . \quad (251.2)$$

Da $\boldsymbol{\Sigma}$ positiv definit ist, existiert nach (142.6) und (143.2) die orthogonale Matrix \boldsymbol{T} , so daß $\boldsymbol{T}' \boldsymbol{\Sigma} \boldsymbol{T} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ gilt, wobei $\lambda_1, \dots, \lambda_n$ die Eigenwerte von $\boldsymbol{\Sigma}$ bedeuten, die alle positiv sind. Da $\boldsymbol{T}' \boldsymbol{T} = \boldsymbol{I}$ und somit $\boldsymbol{T}' = \boldsymbol{T}^{-1}$ gilt, folgt mit (131.14) $(\boldsymbol{T}' \boldsymbol{\Sigma} \boldsymbol{T})^{-1} = \boldsymbol{T}' \boldsymbol{\Sigma}^{-1} \boldsymbol{T} = \text{diag}(\lambda_1^{-1}, \dots, \lambda_n^{-1})$. Für die Transformation $\boldsymbol{y} = \boldsymbol{T}'(\boldsymbol{x} - \boldsymbol{\mu})$, für die $\boldsymbol{x} - \boldsymbol{\mu} = \boldsymbol{T} \boldsymbol{y}$ folgt, gilt nach (173.4) die Funktionaldeterminante $\det \boldsymbol{J} = \det \boldsymbol{T} = \pm 1$, da mit (136.8), (136.10) und (136.13) $1 = \det(\boldsymbol{T}' \boldsymbol{T}) = (\det \boldsymbol{T})^2$ sich ergibt. Das Integral in (251.2) berechnet sich daher wegen (241.2) zu

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{y_i^2}{\lambda_i}\right) dy_1 \dots dy_n = \prod_{i=1}^n (2\pi \lambda_i)^{1/2} .$$

Weiter gilt mit (136.9) und (136.13) $\prod_{i=1}^n \lambda_i = \det \boldsymbol{\Lambda} = \det(\boldsymbol{T}' \boldsymbol{\Sigma} \boldsymbol{T}) = \det(\boldsymbol{T}' \boldsymbol{T}) \det \boldsymbol{\Sigma} = \det \boldsymbol{\Sigma}$, so daß (251.2) folgt.

Die multivariate Normalverteilung soll nun aus der univariaten Normalverteilung hergeleitet werden, indem man sich den $n \times 1$ Zufallsvektor \boldsymbol{x} aus der Transformation $\boldsymbol{x} = \boldsymbol{A} \boldsymbol{y} + \boldsymbol{\mu}$ entstanden denkt, in der der $n \times 1$ Zufallsvektor \boldsymbol{y} die n voneinander unabhängigen Zufallsvariablen y_i mit $y_i \sim N(0, 1)$ enthalte und in der der $n \times 1$ Vektor $\boldsymbol{\mu}$ und die $n \times n$ Matrix \boldsymbol{A} konstante Elemente besitze. \boldsymbol{A} sei regulär, so daß die inverse Transformation $\boldsymbol{y} = \boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ existiert. Die gemeinsame Verteilung $f(\boldsymbol{y})$ des Zufallsvektors \boldsymbol{y} ergibt sich nach (228.2) und (241.1) zu

$$f(\boldsymbol{y}) = \prod_{i=1}^n \frac{\exp(-y_i^2/2)}{\sqrt{2\pi}} = \frac{1}{(2\pi)^{n/2}} \exp(-\boldsymbol{y}' \boldsymbol{y}/2) . \quad (251.3)$$

Durch die Variablentransformation $\boldsymbol{x} = \boldsymbol{A} \boldsymbol{y} + \boldsymbol{\mu}$ mit $\boldsymbol{y} = \boldsymbol{A}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$, deren Funktionaldeterminante nach (173.4) mit $\det \boldsymbol{J} = \det \boldsymbol{A}^{-1}$ gegeben ist,

erhält man mit $(\mathbf{A}^{-1})' \mathbf{A}^{-1} = \mathbf{\Sigma}^{-1}$ wegen (136.8) und (136.13) $(\det \mathbf{A}^{-1})^2 = \det \mathbf{\Sigma}^{-1}$ und $\det \mathbf{A}^{-1} = \pm(\det \mathbf{\Sigma}^{-1})^{1/2} = \pm(\det \mathbf{\Sigma})^{-1/2}$ wegen (136.14). Mit $|\det \mathbf{A}^{-1}| = (\det \mathbf{\Sigma})^{-1/2}$, da $\det \mathbf{\Sigma} > 0$ wegen (136.11) und (143.4) gilt, ergibt sich dann anstelle von (251.3) die Dichte (251.1) der multivariaten Normalverteilung.

252 Momenterzeugende Funktion der Normalverteilung

Für die momenterzeugende Funktion der Normalverteilung gilt der

Satz: Der $n \times 1$ Zufallsvektor \mathbf{x} sei normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$, dann ergibt sich die momenterzeugende Funktion $M_{\mathbf{x}}(\mathbf{t})$ von \mathbf{x} zu

$$M_{\mathbf{x}}(\mathbf{t}) = \exp(\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}). \quad (252.1)$$

Beweis: Mit (234.1) und (251.1) gilt

$$M_{\mathbf{x}}(\mathbf{t}) = \frac{1}{(2\pi)^{n/2}(\det \mathbf{\Sigma})^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(\mathbf{t}'\mathbf{x} - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) dx_1 \dots dx_n.$$

Beachtet man, daß $\exp(\mathbf{t}'\mathbf{x} - (\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})/2) = \exp(\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\mathbf{\Sigma}\mathbf{t}/2 - (\mathbf{x} - \boldsymbol{\mu} - \mathbf{\Sigma}\mathbf{t})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu} - \mathbf{\Sigma}\mathbf{t})/2)$ gilt, so folgt

$$M_{\mathbf{x}}(\mathbf{t}) = \frac{\exp(\mathbf{t}'\boldsymbol{\mu} + \mathbf{t}'\mathbf{\Sigma}\mathbf{t}/2)}{(2\pi)^{n/2}(\det \mathbf{\Sigma})^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - \mathbf{\Sigma}\mathbf{t})'\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu} - \mathbf{\Sigma}\mathbf{t})\right) dx_1 \dots dx_n.$$

Mit der Beziehung (251.2), die auch gültig bleibt, wenn anstelle von $\boldsymbol{\mu}$ ein beliebiger anderer konstanter Vektor wie $\boldsymbol{\mu} + \mathbf{\Sigma}\mathbf{t}$ gesetzt wird, ergibt sich dann die Aussage.

Bereits im vorangegangenen Kapitel war auf den folgenden Satz verwiesen worden.

Satz: Für den $n \times 1$ Zufallsvektor \mathbf{x} gelte $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$, dann folgt $E(\mathbf{x}) = \boldsymbol{\mu}$ und $D(\mathbf{x}) = \mathbf{\Sigma}$. (252.2)

Beweis: Mit (131.6) und $\mathbf{t} = (t_i)$, $\boldsymbol{\mu} = (\mu_i)$, $\mathbf{\Sigma} = (\sigma_{ij})$ erhält man für (252.1)

$$M_{\mathbf{x}}(\mathbf{t}) = \exp\left(\sum_{l=1}^n t_l \mu_l + \frac{1}{2} \sum_{l=1}^n \sum_{m=1}^n t_l \sigma_{lm} t_m\right) \quad (252.3)$$

und mit (172.2)

$$\partial M_{\mathbf{x}}(\mathbf{t})/\partial t_i = (\mu_i + \sum_{m=1}^n \sigma_{im} t_m) M_{\mathbf{x}}(\mathbf{t}) \quad (252.4)$$

sowie

$$\begin{aligned} \partial^2 M_{\mathbf{x}}(\mathbf{t})/(\partial t_i \partial t_j) &= \sigma_{ij} M_{\mathbf{x}}(\mathbf{t}) \\ &+ (\mu_i + \sum_{m=1}^n \sigma_{im} t_m)(\mu_j + \sum_{m=1}^n \sigma_{jm} t_m) M_{\mathbf{x}}(\mathbf{t}) . \end{aligned} \quad (252.5)$$

Aus (252.4) ergibt sich wegen (234.2) mit $\mathbf{t} = \mathbf{0}$ der Erwartungswert $E(x_i) = \mu_i$ und damit die erste Aussage. Mit $\mathbf{t} = \mathbf{0}$ folgt aus (252.5) $E(x_i x_j) = \sigma_{ij} + \mu_i \mu_j$ oder $\sigma_{ij} = E(x_i x_j) - E(x_i)E(x_j)$, was identisch ist mit (232.8), so daß $D(\mathbf{x}) = \Sigma$ und damit die zweite Aussage folgt.

Beispiel: Mit Hilfe von (252.1) sollen für den $n \times 1$ Zufallsvektor $\mathbf{e} = (e_i)$ mit $\mathbf{e} \sim N(\mathbf{0}, \Sigma)$ die Erwartungswerte $E(e_i e_j e_k)$ und $E(e_i e_j e_k e_l)$ berechnet werden. Mit $\mathbf{t} = (t_i)$ und $\Sigma = (\sigma_{ij})$ folgt wegen $\boldsymbol{\mu} = \mathbf{0}$ aus (252.3) $M_{\mathbf{e}}(\mathbf{t}) = \exp(\frac{1}{2} \sum_{l=1}^n \sum_{m=1}^n t_l \sigma_{lm} t_m)$ und aus (252.5)

$$\begin{aligned} \partial^3 M_{\mathbf{e}}(\mathbf{t})/\partial t_i \partial t_j \partial t_k &= \sigma_{ij} (\sum_{m=1}^n \sigma_{km} t_m) M_{\mathbf{e}}(\mathbf{t}) \\ &+ \sigma_{ik} (\sum_{m=1}^n \sigma_{jm} t_m) M_{\mathbf{e}}(\mathbf{t}) + \sigma_{jk} (\sum_{m=1}^n \sigma_{im} t_m) M_{\mathbf{e}}(\mathbf{t}) \\ &+ (\sum_{m=1}^n \sigma_{im} t_m \sum_{m=1}^n \sigma_{jm} t_m \sum_{m=1}^n \sigma_{km} t_m) M_{\mathbf{e}}(\mathbf{t}) \end{aligned} \quad (252.6)$$

und schließlich

$$\begin{aligned} \partial^4 M_{\mathbf{e}}(\mathbf{t})/\partial t_i \partial t_j \partial t_k \partial t_l &= \\ &\sigma_{ij} \sigma_{kl} M_{\mathbf{e}}(\mathbf{t}) + \sigma_{ik} \sigma_{jl} M_{\mathbf{e}}(\mathbf{t}) + \sigma_{jk} \sigma_{il} M_{\mathbf{e}}(\mathbf{t}) + h(t_m M_{\mathbf{e}}(\mathbf{t})) , \end{aligned} \quad (252.7)$$

worin $h(t_m M_{\mathbf{e}}(\mathbf{t}))$ Summanden bedeuten, die die Faktoren $t_m M_{\mathbf{e}}(\mathbf{t})$ enthalten. Mit $\mathbf{t} = \mathbf{0}$ ergibt sich dann aufgrund von (234.2) aus (252.6) und (252.7)

$$E(e_i e_j e_k) = 0 \quad \text{und} \quad E(e_i e_j e_k e_l) = \sigma_{ij} \sigma_{kl} + \sigma_{ik} \sigma_{jl} + \sigma_{jk} \sigma_{il} . \quad (252.8)$$

253 Randverteilung und bedingte Verteilung

Im Zusammenhang mit der Parameterschätzung ist die bedingte Normalverteilung von Interesse. Um sie anzugeben, wird zunächst die Randverteilung

abgeleitet.

Satz: Der $n \times 1$ Zufallsvektor \mathbf{x} mit $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ werde mit $\mathbf{x} = |\mathbf{x}'_1, \mathbf{x}'_2|'$ in die $k \times 1$ und $(n - k) \times 1$ Zufallsvektoren \mathbf{x}_1 und \mathbf{x}_2 zerlegt. Mit der entsprechenden Aufteilung des Erwartungswertvektors $\boldsymbol{\mu}$ und der Kovarianzmatrix $\boldsymbol{\Sigma}$ in

$$\boldsymbol{\mu} = |\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2|' \quad \text{und} \quad \boldsymbol{\Sigma} = \begin{vmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{vmatrix} \quad \text{mit} \quad \boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}'_{12}$$

ergibt sich die Randverteilung $g(\mathbf{x}_1)$ von \mathbf{x}_1 zu

$$g(\mathbf{x}_1) = \frac{1}{(2\pi)^{k/2}(\det \boldsymbol{\Sigma}_{11})^{1/2}} e^{-\frac{1}{2}(\mathbf{x}_1 - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)}. \quad (253.1)$$

Beweis: Da $\boldsymbol{\Sigma}$ positiv definit ist, sind auch $\boldsymbol{\Sigma}_{11}$ und $\boldsymbol{\Sigma}_{22}$ positiv definit, wie sich durch entsprechende Wahl des Vektors der quadratischen Form in (143.1) zeigen läßt. Damit existieren nach (143.3) $\boldsymbol{\Sigma}_{11}^{-1}$ und $\boldsymbol{\Sigma}_{22}^{-1}$. Mit dem $k \times 1$ Vektor \mathbf{t}_1 erhält man dann nach (234.1) die momenterzeugende Funktion $M_{\mathbf{x}_1}(\mathbf{t}_1)$ der Randverteilung $g(\mathbf{x}_1)$ von \mathbf{x}_1 zu

$$M_{\mathbf{x}_1}(\mathbf{t}_1) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(\mathbf{t}'_1 \mathbf{x}_1) g(\mathbf{x}_1) dx_1 \dots dx_k$$

und durch Substitution von $g(\mathbf{x}_1)$ nach (226.4) mit (251.1)

$$M_{\mathbf{x}_1}(\mathbf{t}_1) = \frac{1}{(2\pi)^{n/2}(\det \boldsymbol{\Sigma})^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp(\mathbf{t}'_1 \mathbf{x}_1 - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})) dx_1 \dots dx_n.$$

Führt man nun den Vektor $\mathbf{t} = |\mathbf{t}'_1, \mathbf{0}'|'$ ein, so erhält man $M_{\mathbf{x}_1}(\mathbf{t}_1) = M_{\mathbf{x}}(\mathbf{t})$ und hierfür wie im Beweis zu (252.1)

$$M_{\mathbf{x}_1}(\mathbf{t}_1) = \exp(\mathbf{t}'_1 \boldsymbol{\mu}_1 + \frac{1}{2} \mathbf{t}'_1 \boldsymbol{\Sigma}_{11} \mathbf{t}_1),$$

woraus durch Vergleich mit (252.1) wegen (234.3) die Aussage folgt.

Mit Hilfe der Randverteilung kann nun die bedingte Normalverteilung abgeleitet werden.

Satz: Bei der Zerlegung des normalverteilten Zufallsvektors \mathbf{x} in zwei Zufallsvektoren wie im Satz (253.1) ergibt sich die Verteilung des Zufallsvektors \mathbf{x}_1 unter der Bedingung, daß der zweite Zufallsvektor die Werte \mathbf{x}_2 annimmt, zu

$$\mathbf{x}_1 | \mathbf{x}_2 \sim N(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}). \quad (253.2)$$

Beweis: Wie bereits im Beweis zu (253.1) gezeigt, existiert Σ_{22}^{-1} . Weiter gilt nach (227.5) mit (253.1)

$$f(\mathbf{x}_1|\mathbf{x}_2) = \frac{\exp\{-\frac{1}{2}[(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) - (\mathbf{x}_2 - \boldsymbol{\mu}_2)' \Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)]\}}{(2\pi)^{k/2}(\det \Sigma)^{1/2}(\det \Sigma_{22})^{-1/2}}.$$

Setzt man $\mathbf{F} = (\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})^{-1}$ ergibt sich mit (134.3) bis (134.5)

$$\Sigma^{-1} = \begin{vmatrix} \mathbf{F} & -\mathbf{F}\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma_{22}^{-1}\Sigma_{21}\mathbf{F} & \Sigma_{22}^{-1} + \Sigma_{22}^{-1}\Sigma_{21}\mathbf{F}\Sigma_{12}\Sigma_{22}^{-1} \end{vmatrix}$$

und somit für den Exponenten in $f(\mathbf{x}_1|\mathbf{x}_2)$

$$\begin{aligned} &(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) - (\mathbf{x}_2 - \boldsymbol{\mu}_2)' \Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) = \\ &[\mathbf{x}_1 - \boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)]' \mathbf{F}[\mathbf{x}_1 - \boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2)]. \end{aligned}$$

Weiter erhält man mit (136.12) $\det \Sigma = \det(\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}) \det \Sigma_{22}$ und daher $\det \Sigma(\det \Sigma_{22})^{-1} = \det \mathbf{F}^{-1}$. Diese beiden Ausdrücke in $f(\mathbf{x}_1|\mathbf{x}_2)$ substituiert ergeben die Aussage.

254 Unabhängigkeit normalverteilter Zufallsvariablen

Sind Zufallsvariable voneinander unabhängig, so sind ihre Kovarianzen nach (232.7) gleich Null. Für normalverteilte Zufallsvariable gilt auch die Umkehrung dieser Aussage.

Satz: Der Zufallsvektor \mathbf{x} mit $\mathbf{x} \sim N(\boldsymbol{\mu}, \Sigma)$ sei in die k Zufallsvektoren \mathbf{x}_i mit $\mathbf{x} = [\mathbf{x}'_1, \dots, \mathbf{x}'_k]'$ zerlegt. Die Zufallsvektoren \mathbf{x}_i sind genau dann voneinander unabhängig, wenn in der entsprechenden Zerlegung der Kovarianzmatrix $\Sigma = (\Sigma_{ij})$ für $i, j \in \{1, \dots, k\}$ die Beziehung $\Sigma_{ij} = \mathbf{0}$ für $i \neq j$ gilt. (254.1)

Beweis: Mit (252.1) ergibt sich die momenterzeugende Funktion $M_{\mathbf{x}}(\mathbf{t})$ von \mathbf{x} bei einer Zerlegung von \mathbf{t} und $\boldsymbol{\mu}$, die der von \mathbf{x} entspricht, zu

$$M_{\mathbf{x}}(\mathbf{t}) = \exp\left(\sum_{i=1}^k \mathbf{t}'_i \boldsymbol{\mu}_i + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \mathbf{t}'_i \Sigma_{ij} \mathbf{t}_j\right).$$

Mit $\Sigma_{ij} = \mathbf{0}$ für $i \neq j$ folgt

$$M_{\mathbf{x}}(\mathbf{t}) = \exp\left(\sum_{i=1}^k (\mathbf{t}'_i \boldsymbol{\mu}_i + \frac{1}{2} \mathbf{t}'_i \Sigma_{ii} \mathbf{t}_i)\right) = \prod_{i=1}^k \exp(\mathbf{t}'_i \boldsymbol{\mu}_i + \frac{1}{2} \mathbf{t}'_i \Sigma_{ii} \mathbf{t}_i).$$

Das ist aber das Produkt der momenterzeugenden Funktionen der Normalverteilungen der einzelnen Vektoren \mathbf{x}_i , so daß nach (234.4) die Zufallsvektoren \mathbf{x}_i voneinander unabhängig sind. Nimmt man andererseits an, daß die k

Vektoren \mathbf{x}_i mit $\mathbf{x}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{ii})$ voneinander unabhängig sind, dann ergibt sich die momenterzeugende Funktion $M_{\mathbf{x}}(\mathbf{t})$ der gemeinsamen Verteilung von $\mathbf{x} = |\mathbf{x}'_1, \dots, \mathbf{x}'_k|'$ aus (234.4) und (252.1) zu

$$M_{\mathbf{x}}(\mathbf{t}) = \exp\left(\sum_{i=1}^k (\mathbf{t}'_i \boldsymbol{\mu}_i + \frac{1}{2} \mathbf{t}'_i \boldsymbol{\Sigma}_{ii} \mathbf{t}_i)\right) = \exp(\mathbf{t}' \boldsymbol{\mu} + \frac{1}{2} \mathbf{t}' \boldsymbol{\Sigma} \mathbf{t}),$$

falls $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_{11}, \dots, \boldsymbol{\Sigma}_{kk})$, also falls $\boldsymbol{\Sigma}_{ij} = \mathbf{0}$ für $i \neq j$ gilt, so daß die Aussage folgt.

255 Lineare Funktionen normalverteilter Zufallsvariablen

Im folgenden wird häufig die Verteilung von Zufallsvariablen benötigt, die durch eine lineare Transformation normalverteilter Zufallsvariablen entstanden sind.

Satz: Der $n \times 1$ Zufallsvektor \mathbf{x} sei normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, dann besitzt der $m \times 1$ Zufallsvektor \mathbf{y} , der durch die lineare Transformation $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{c}$ erhalten werde, in der \mathbf{A} eine $m \times n$ Matrix von Konstanten mit vollem Zeilenrang m und \mathbf{c} ein $m \times 1$ Vektor von Konstanten bedeuten, die Normalverteilung

$$\mathbf{y} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}') . \quad (255.1)$$

Beweis: Die momenterzeugende Funktion $M_{\mathbf{y}}(\mathbf{t})$ von \mathbf{y} ergibt sich aus (234.1) und (251.1) zu

$$M_{\mathbf{y}}(\mathbf{t}) = \frac{1}{(2\pi)^{n/2} (\det \boldsymbol{\Sigma})^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp[\mathbf{t}'(\mathbf{A}\mathbf{x} + \mathbf{c}) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})] dx_1 \dots dx_n .$$

Für den Exponenten gilt

$$\begin{aligned} \mathbf{t}'(\mathbf{A}\mathbf{x} + \mathbf{c}) - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) &= \mathbf{t}'(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}) + \frac{1}{2} \mathbf{t}' \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}' \mathbf{t} \\ &\quad - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu} - \boldsymbol{\Sigma} \mathbf{A}' \mathbf{t})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu} - \boldsymbol{\Sigma} \mathbf{A}' \mathbf{t}), \end{aligned}$$

so daß sich dem Beweis zu (252.1) entsprechend ergibt

$$M_{\mathbf{y}}(\mathbf{t}) = \exp[\mathbf{t}'(\mathbf{A}\boldsymbol{\mu} + \mathbf{c}) + \frac{1}{2} \mathbf{t}' \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}' \mathbf{t}] .$$

Das ist aber die momenterzeugende Funktion eines normalverteilten Zufallsvektors mit dem Erwartungswert $\mathbf{A}\boldsymbol{\mu} + \mathbf{c}$ und der Kovarianzmatrix $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}'$, die wegen $\text{rg}\mathbf{A} = m$ nach (143.7) positiv definit ist, so daß wegen (234.3) die Aussage folgt.

256 Summe normalverteilter Zufallsvariablen

Wie die Gammaverteilung, so besitzt auch die Normalverteilung die reproduzierende Eigenschaft.

Satz: Die $n \times 1$ Vektoren \mathbf{x}_i mit $i \in \{1, \dots, k\}$ seien voneinander unabhängig und normalverteilt wie $\mathbf{x}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, dann ist der $n \times 1$ Zufallsvektor $\mathbf{x} = c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k$, worin die c_i Konstanten bedeuten, normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit $\boldsymbol{\mu} = c_1\boldsymbol{\mu}_1 + \dots + c_k\boldsymbol{\mu}_k$ und $\boldsymbol{\Sigma} = c_1^2\boldsymbol{\Sigma}_1 + \dots + c_k^2\boldsymbol{\Sigma}_k$. (256.1)
Beweis: Der Zufallsvektor $\mathbf{y} = |\mathbf{x}'_1, \dots, \mathbf{x}'_k|'$ besitzt nach (254.1) die Normalverteilung mit dem Erwartungswert $\boldsymbol{\mu} = |\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_k|'$ und der Kovarianzmatrix $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_k)$. Der Zufallsvektor \mathbf{x} mit

$$\mathbf{x} = |c_1\mathbf{I}, \dots, c_k\mathbf{I}| \begin{vmatrix} \mathbf{x}_1 \\ \dots \\ \mathbf{x}_k \end{vmatrix} = c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k$$

ist dann nach (255.1) normalverteilt wie $\mathbf{x} \sim N(c_1\boldsymbol{\mu}_1 + \dots + c_k\boldsymbol{\mu}_k, c_1^2\boldsymbol{\Sigma}_1 + \dots + c_k^2\boldsymbol{\Sigma}_k)$, so daß die Aussage folgt.

Beispiel: Es seien k voneinander unabhängige Zufallsvariablen x_i mit $x_i \sim N(\mu, \sigma^2)$ gegeben, das Mittel $\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i$ ist dann nach (256.1) normalverteilt wie $\bar{x} \sim N(\mu, \sigma^2/k)$.

26 Testverteilungen für univariate Modelle der Parameterschätzung

261 χ^2 -Verteilung

Die χ^2 -Verteilung (Chi-Quadrat-Verteilung) gehört wie die folgenden Verteilungen zu den *Testverteilungen*, die die Verteilungen von Funktionen normalverteilter Zufallsvariablen angeben und die, wie im Abschnitt 4 gezeigt wird, zur Berechnung der Wahrscheinlichkeiten bei Hypothesentests oder bei Bereichsschätzungen in univariaten Modellen der Parameterschätzung dienen.

Satz: Der $n \times 1$ Zufallsvektor $\mathbf{x} = |x_1, \dots, x_n|'$ sei normalverteilt wie $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$, dann besitzt die Quadratsumme $v = \mathbf{x}'\mathbf{x} = \sum_{i=1}^n x_i^2$ die χ^2 -Verteilung $\chi^2(n)$ mit n Freiheitsgraden, also $v \sim \chi^2(n)$, deren Dichte gegeben ist durch

$$f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2} \quad \text{für } 0 < v < \infty$$

und $f(v) = 0$ für die übrigen Werte von v . (261.1)

Beweis: Mit $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$, was nach (254.1) die Unabhängigkeit der Zufallsvariablen x_i bedeutet, ergibt sich die momenterzeugende Funktion $M_v(t)$ von v aus (234.1) und (251.1) zu

$$M_v(t) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(t \sum_{i=1}^n x_i^2 - \frac{1}{2} \sum_{i=1}^n x_i^2\right) dx_1 \dots dx_n .$$

Jedes der n Integrale besitzt, wie ein Vergleich mit (241.2) zeigt, den Wert

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(1-2t)x_i^2\right) dx_i = (1-2t)^{-1/2} ,$$

so daß $M_v(t) = (1-2t)^{-n/2}$ folgt. Das ist aber für $t < 1/2$ die momenterzeugende Funktion (243.8) der Gammaverteilung (243.1) mit $b = 1/2$ und $p = n/2$, so daß wegen (234.3) die Aussage folgt.

Ist anstelle eines Zufallsvektors mit \mathbf{I} als Kovarianzmatrix der $n \times 1$ Zufallsvektor \mathbf{y} mit der positiv definiten Kovarianzmatrix $\mathbf{\Sigma}$ und $\mathbf{y} \sim N(\mathbf{0}, \mathbf{\Sigma})$ gegeben, so läßt sich die Inverse nach (143.5) zerlegen in $\mathbf{\Sigma}^{-1} = \mathbf{G}\mathbf{G}'$, wobei \mathbf{G} eine reguläre untere Dreiecksmatrix bedeutet. Für den Zufallsvektor

$$\mathbf{x} = \mathbf{G}'\mathbf{y} \tag{261.2}$$

folgt dann mit $\mathbf{\Sigma} = (\mathbf{G}')^{-1}\mathbf{G}^{-1}$ wegen (131.14) aus (255.1) $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I})$ und daher aus (261.1)

$$v = \mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{\Sigma}^{-1}\mathbf{y} \sim \chi^2(n) . \tag{261.3}$$

Ist $\mathbf{\Sigma}$ eine Diagonalmatrix mit $\mathbf{\Sigma} = (\text{diag}(\sigma_i^2))$, so gilt $\mathbf{\Sigma}^{-1} = (\text{diag}(1/\sigma_i^2))$, und man erhält anstelle von (261.3)

$$v = \mathbf{x}'\mathbf{x} = \sum_{i=1}^n (y_i/\sigma_i)^2 \sim \chi^2(n) . \tag{261.4}$$

Für $v \sim \chi^2(n)$ ergibt sich aus (223.5) die Verteilungsfunktion $F(\chi^2; n)$ zu

$$F(\chi^2; n) = [2^{n/2}\Gamma(n/2)]^{-1} \int_0^{\chi^2} v^{(n/2)-1} e^{-v/2} dv . \tag{261.5}$$

Werte von $F(\chi^2; n)$ findet man häufig tabuliert (z.B. FISHER und YATES 1963, S.47; PEARSON und HARTLEY 1976, Vol.I, S.136). Für eine numerische Berechnung eignet sich besonders für kleine Werte von χ^2 die unendliche Reihe, die man mit $b = 1/2, p = n/2$ und $G = \chi^2$ aus (243.11) erhält

$$F(\chi^2; n) = \left(\frac{\chi^2}{2}\right)^{\frac{n}{2}} \frac{e^{-\chi^2/2}}{\Gamma(\frac{n+2}{2})} \left(1 + \sum_{j=1}^{\infty} \frac{(\chi^2)^j}{(n+2)(n+4)\dots(n+2j)}\right) . \tag{261.6}$$

Aus der partiellen Integration des Integrals in (261.5) in den Grenzen von χ^2 bis ∞ erhält man die Reihe

$$\begin{aligned} F(\chi^2; n) &= 1 - (2^{\frac{n}{2}} \Gamma(n/2))^{-1} \left([-2e^{-\frac{v}{2}} v^{\frac{n}{2}-1}]_{\chi^2}^{\infty} \right. \\ &\quad \left. 2\left(\frac{n}{2} - 1\right) \int_{\chi^2}^{\infty} v^{\frac{n}{2}-2} e^{-\frac{v}{2}} dv \right) = 1 - \frac{e^{-\chi^2/2} (\chi^2)^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} \\ &\quad \left(2 + \frac{2^2}{(\chi^2)^1} \left(\frac{n}{2} - 1\right) + \frac{2^3}{(\chi^2)^2} \left(\frac{n}{2} - 1\right) \left(\frac{n}{2} - 2\right) + \dots \right). \end{aligned} \quad (261.7)$$

Obwohl die Reihe für ungerade Werte von n divergiert, können aus ihr für große Werte von χ^2 Näherungswerte für $F(\chi^2; n)$ berechnet werden (ABRAMOWITZ und STEGUN 1972, S.941). Für gerade Werte von n folgt die endliche Reihe

$$F(\chi^2; n) = 1 - e^{-\chi^2/2} \sum_{j=0}^{n/2-1} \frac{(\chi^2/2)^j}{j!}. \quad (261.8)$$

Für $n > 100$ erhält man mit (241.4) die Näherung (ABRAMOWITZ und STEGUN 1972, S.941)

$$F(\chi^2; n) \approx F(x; 0, 1) \quad \text{mit} \quad x = \sqrt{2\chi^2} - \sqrt{2n-1}. \quad (261.9)$$

Entsprechend (241.6) ist das α -Fraktile $\chi_{\alpha;n}^2$ der χ^2 -Verteilung definiert durch

$$F(\chi_{\alpha;n}^2; n) = \alpha. \quad (261.10)$$

Die α -Fraktile lassen sich den angegebenen Tafeln entnehmen oder berechnen. Näherungsweise gilt (JOHNSON und KOTZ 1970, Vol. 1, S.176)

$$\chi_{\alpha;n}^2 \approx n \left(x_{\alpha} \left(\frac{2}{9n} \right)^{1/2} + 1 - \frac{2}{9n} \right)^3, \quad (261.11)$$

worin x_{α} das mit (241.6) definierte α -Fraktile der standardisierten Normalverteilung bedeutet.

Um den Näherungswert aus (261.11), der mit $\bar{\chi}_{\alpha;n}^2$ bezeichnet sei, zu verbessern, läßt sich die differentielle Größe $d\chi^2$ berechnen, falls $\bar{\chi}_{\alpha;n}^2 = \chi_{\alpha;n}^2 + d\chi^2$ gilt. Mit (261.5) wird die Wahrscheinlichkeit α_n aus

$$P(v < \chi_{\alpha;n}^2 + d\chi^2) = F(\bar{\chi}_{\alpha;n}^2; n) = \alpha_n \quad (261.12)$$

erhalten. Mit (223.7) folgt dann $P(\chi_{\alpha;n}^2 < v < \chi_{\alpha;n}^2 + d\chi^2) = \alpha_n - \alpha$ und mit (223.8) und (261.1)

$$f(\bar{\chi}_{\alpha;n}^2) d\chi^2 = \alpha_n - \alpha, \quad (261.13)$$

woraus zusammen mit (261.12) $d\chi^2$ und damit $\chi^2_{\alpha;n}$ iterativ mit der gewünschten Genauigkeit zu ermitteln ist.

262 Nichtzentrale χ^2 -Verteilung

Satz: Der $n \times 1$ Zufallsvektor $\mathbf{x} = |x_1, \dots, x_n|'$ sei normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{I})$, dann besitzt die Quadratsumme $v = \mathbf{x}'\mathbf{x} = \sum_{i=1}^n x_i^2$ die *nichtzentrale χ^2 -Verteilung* $\chi'^2(n, \lambda)$ mit n Freiheitsgraden und dem Nichtzentralitätsparameter $\lambda = \boldsymbol{\mu}'\boldsymbol{\mu}$, also $v \sim \chi'^2(n, \lambda)$, deren Dichte gegeben ist durch

$$f(v) = e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{(\frac{\lambda}{2})^j v^{\frac{n}{2}+j-1} e^{-\frac{v}{2}}}{j! 2^{\frac{n}{2}+j} \Gamma(\frac{n}{2}+j)} \quad \text{für } 0 < v < \infty$$

und $f(v) = 0$ für die übrigen Werte von v . (262.1)

Beweis: Die momenterzeugende Funktion $M_v(t)$ der Quadratsumme v ergibt sich mit (234.1) und (251.1) zu

$$\begin{aligned} M_v(t) &= \left(\frac{1}{2\pi}\right)^{n/2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(t \sum_{i=1}^n x_i^2 - \frac{1}{2} \sum_{i=1}^n (x_i - \mu_i)^2\right) dx_1 \dots dx_n \\ &= \prod_{i=1}^n \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\left(tx_i^2 - \frac{1}{2}(x_i - \mu_i)^2\right) dx_i. \end{aligned}$$

Der Exponent wird umgeformt

$$\begin{aligned} &-\frac{1}{2}(-2tx_i^2 + x_i^2 - 2x_i\mu_i + \mu_i^2) \\ &= -\frac{1}{2}\left(x_i^2(1-2t) - 2x_i\mu_i + \mu_i^2(1-2t)^{-1} + \mu_i^2 - \mu_i^2(1-2t)^{-1}\right) \\ &= -\frac{1}{2}\left((x_i - \mu_i(1-2t)^{-1})^2(1-2t) + \mu_i^2(1 - (1-2t)^{-1})\right), \end{aligned}$$

so daß sich mit (241.2) ergibt

$$\begin{aligned} M_v(t) &= \exp\left(-\frac{1}{2}(1 - (1-2t)^{-1}) \sum_{i=1}^n \mu_i^2\right) \\ &\quad \prod_{i=1}^n \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\left(-\frac{(x_i - \mu_i(1-2t)^{-1})^2}{2(1-2t)^{-1}}\right) dx_i \\ &= (1-2t)^{-n/2} \exp\left(-\frac{\lambda}{2}(1 - (1-2t)^{-1})\right). \end{aligned} \quad (262.2)$$

Andererseits erhält man für die momenterzeugende Funktion $M_v(t)$ der Quadratsumme v mit (262.1) aus (234.1)

$$M_v(t) = \sum_{j=0}^{\infty} \frac{e^{-\lambda/2} (\lambda/2)^j}{j!} \int_{-0}^{\infty} e^{vt} \frac{v^{n/2+j-1} e^{-v/2}}{2^{n/2+j} \Gamma(n/2+j)} dv .$$

Das Integral stellt, wie ein Vergleich mit (261.1) zeigt, die momenterzeugende Funktion für $v \sim \chi^2(n+2j)$ dar, die $(1-2t)^{-n/2-j}$ beträgt, wie aus dem Beweis zu (261.1) sich ergibt. Es folgt daher

$$M_v(t) = e^{-\lambda/2} (1-2t)^{-n/2} \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\lambda}{2} (1-2t)^{-1} \right)^j$$

und mit (224.3) die mit (262.2) identische momenterzeugende Funktion, so daß wegen (234.3) die Aussage folgt.

Mit $\lambda = 0$ erhält man anstelle der nichtzentralen χ^2 -Verteilung (262.1) die χ^2 -Verteilung (261.1).

Liegt anstelle eines Zufallsvektors mit der Kovarianzmatrix \mathbf{I} der $n \times 1$ Zufallsvektor \mathbf{y} mit $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ vor, so gelten mit der linearen Transformation (261.2) entsprechend (261.3) und (261.4) die nichtzentralen χ^2 -Verteilungen mit $\lambda = \boldsymbol{\mu}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$.

Für $v \sim \chi'^2(n, \lambda)$ ergibt sich aus (223.5) die Verteilungsfunktion $F(\chi'^2; n, \lambda)$ mit (262.1) und (261.5) zu

$$F(\chi'^2; n, \lambda) = e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} F(\chi'^2; n+2j) , \quad (262.3)$$

so daß die Verteilungsfunktion der nichtzentralen χ^2 -Verteilung mit Hilfe der Verteilungsfunktion der χ^2 -Verteilung nach (261.6) bis (261.9) zu berechnen ist. Tabellierte Werte für $F(\chi'^2; n, \lambda)$ befinden sich in (PEARSON und HARTLEY 1976, Vol.II, S.232). Für große Werte von λ konvergiert die Reihe (262.3) langsam, so daß man in diesen Fällen zweckmäßig die mit der Verteilungsfunktion (261.5) der χ^2 -Verteilung zu berechnende Näherung wählt (PATNAIK 1949)

$$F(\chi'^2; n, \lambda) \approx F(\chi'^2/\rho; \nu) \quad \text{mit} \quad \rho = \frac{n+2\lambda}{n+\lambda} , \quad \nu = \frac{(n+\lambda)^2}{n+2\lambda} . \quad (262.4)$$

263 F -Verteilung

Die F -Verteilung (Fisher-Verteilung) und die nichtzentrale F -Verteilung werden bei Hypothesentests benötigt.

Satz: Die Zufallsvariablen u und v mit $u \sim \chi^2(m)$ und $v \sim \chi^2(n)$ seien voneinander unabhängig, dann besitzt die Zufallsvariable $w = (u/m)/(v/n)$ die F -Verteilung $F(m, n)$ mit m und n Freiheitsgraden, also $w \sim F(m, n)$, mit der Dichte

$$f(w) = \frac{\Gamma(\frac{m}{2} + \frac{n}{2}) m^{\frac{m}{2}} n^{\frac{n}{2}} w^{\frac{m}{2}-1}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})(n + mw)^{\frac{m}{2}+\frac{n}{2}}} \quad \text{für } 0 < w < \infty$$

und $f(w) = 0$ für die übrigen Werte von w . (263.1)

Beweis: Da u und v voneinander unabhängig sind, erhält man ihre gemeinsame Verteilung $f(u, v)$ mit (228.2) und (261.1) zu

$$f(u, v) = \frac{u^{\frac{m}{2}-1} v^{\frac{n}{2}-1} e^{-\frac{u}{2} - \frac{v}{2}}}{2^{\frac{m}{2}+\frac{n}{2}} \Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \quad \text{mit } u > 0 \quad \text{und } v > 0.$$

Durch die Transformation der Variablen u in w mit $u = mvw/n$ und $\det \mathbf{J} = mv/n$ in (229.1) ergibt sich

$$f(w, v) = \frac{\frac{m}{n} \left(\frac{mw}{n}\right)^{\frac{m}{2}-1} v^{\frac{m}{2}+\frac{n}{2}-1}}{2^{\frac{m}{2}+\frac{n}{2}} \Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} e^{-\frac{1}{2}(1+\frac{mw}{n})v} \quad \text{für } w > 0, v > 0.$$

Aus dieser Dichte wird durch Integration über v die Randverteilung $g(w)$ nach (226.4) berechnet. Beachtet man, daß das Integral von Null bis Unendlich über die Dichte der Gammaverteilung $G(\frac{1}{2}(1 + \frac{mw}{n}), \frac{m}{2} + \frac{n}{2})$ nach (243.3) den Wert Eins ergibt, erhält man die Randverteilung

$$g(w) = \frac{\Gamma(\frac{m}{2} + \frac{n}{2}) \left(\frac{m}{n}\right)^{\frac{m}{2}} w^{\frac{m}{2}-1}}{\left(1 + \frac{mw}{n}\right)^{\frac{m}{2}+\frac{n}{2}} \Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \quad \text{für } w > 0,$$

die identisch ist mit der F -Verteilung, so daß die Aussage folgt.

Für $w \sim F(m, n)$ ergibt sich aus (223.5) die Verteilungsfunktion $F(F_0; m, n)$ zu

$$F(F_0; m, n) = \frac{\Gamma(\frac{m}{2} + \frac{n}{2}) m^{\frac{m}{2}} n^{\frac{n}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \int_0^{F_0} \frac{w^{\frac{m}{2}-1} dw}{(n + mw)^{\frac{m}{2}+\frac{n}{2}}}. \quad (263.2)$$

Werte der Verteilungsfunktion der F -Verteilung liegen in vielen Veröffentlichungen tabelliert vor (z.B. FISHER und YATES 1963, S.49; PEARSON und HARTLEY 1976, Vol.I, S.169). Zur numerischen Berechnung benutzt man zweckmäßig die unvollständige Betafunktion (245.3) zusammen mit (245.4), denn es gilt

$$F(F_0; m, n) = 1 - I_x\left(\frac{n}{2}, \frac{m}{2}\right) \quad \text{mit } x = \frac{n}{n + mF_0} \quad (263.3)$$

oder wegen (245.5)

$$F(F_0; m, n) = I_{1-x} \left(\frac{m}{2}, \frac{n}{2} \right) \quad \text{mit} \quad 1-x = \frac{mF_0}{n+mF_0}. \quad (263.4)$$

Führt man nämlich in (263.2) die Variablentransformation $t = n/(n+mw)$ mit $dt = -nmdw/(n+mw)^2$ durch, ergibt sich

$$F(F_0; m, n) = -\frac{\Gamma(\frac{m}{2} + \frac{n}{2})m^{\frac{m}{2}}n^{\frac{n}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})nm} \int_1^x \frac{(\frac{n}{mt})^{\frac{m}{2}-1}(1-t)^{\frac{n}{2}-1}}{(\frac{n}{t})^{\frac{m}{2}+\frac{n}{2}-2}} dt$$

und daraus (263.3) mit (245.3).

Entsprechend (241.6) ist das α -Fraktile $F_{\alpha; m, n}$ der F -Verteilung definiert durch

$$F(F_{\alpha; m, n}; m, n) = \alpha. \quad (263.5)$$

Die α -Fraktile lassen sich den angegebenen Tafeln entnehmen oder berechnen. Näherungsweise gilt (ABRAMOWITZ und STEGUN 1972, S.947)

$$F_{\alpha; m, n} \approx e^{2w} \quad (263.6)$$

mit

$$w = \frac{x_\alpha(h+\lambda)^{1/2}}{h} - \left(\frac{1}{m-1} - \frac{1}{n-1} \right) \left(\lambda + \frac{5}{6} - \frac{2}{3h} \right)$$

$$h = 2 \left(\frac{1}{m-1} + \frac{1}{n-1} \right)^{-1}, \quad \lambda = \frac{x_\alpha^2 - 3}{6},$$

worin x_α das mit (241.6) definierte α -Fraktile der standardisierten Normalverteilung bedeutet. Der aus (263.6) berechnete Näherungswert läßt sich (261.12) und (261.13) entsprechend verbessern.

Wie im Kapitel 432 gezeigt wird, benötigt man zur Berechnung von Konfidenz-ellipsen das $(1-\alpha)$ -Fraktile $F_{1-\alpha; 2, n}$, das explizit angegeben werden kann. Aus (263.2) folgt

$$1-\alpha = \frac{\Gamma(\frac{n}{2}+1)2n^{\frac{n}{2}}}{\Gamma(\frac{n}{2})\Gamma(1)} \int_0^{F_{1-\alpha; 2, n}} \frac{dw}{(n+2w)^{\frac{n}{2}+1}}$$

$$= n^{\frac{n}{2}+1} \left[-\frac{1}{n(n+2w)^{\frac{n}{2}}} \right]_0^{F_{1-\alpha; 2, n}}$$

$$= n^{\frac{n}{2}} \left(n^{-\frac{n}{2}} - (n+2F_{1-\alpha; 2, n})^{-\frac{n}{2}} \right)$$

$$= 1 - \left(\frac{n}{n+2F_{1-\alpha; 2, n}} \right)^{\frac{n}{2}}$$

und somit

$$\alpha = \left(\frac{n}{n + 2F_{1-\alpha;2,n}} \right)^{\frac{n}{2}} \quad \text{oder} \quad \alpha^{\frac{2}{n}} = \frac{n}{n + 2F_{1-\alpha;2,n}},$$

so daß sich schließlich ergibt

$$F_{1-\alpha;2,n} = \frac{n}{2} \left(\frac{1}{\alpha^{2/n}} - 1 \right). \quad (263.7)$$

264 Nichtzentrale F -Verteilung

Satz: Die Zufallsvariablen u und v mit $u \sim \chi'^2(m, \lambda)$ und $v \sim \chi^2(n)$ seien voneinander unabhängig, dann besitzt die Zufallsvariable $w = (u/m)/(v/n)$ die nichtzentrale F -Verteilung $F'(m, n, \lambda)$ mit m und n Freiheitsgraden und dem Nichtzentralitätsparameter λ , also $w \sim F'(m, n, \lambda)$, mit der Dichte

$$f(w) = e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^j \Gamma\left(\frac{m}{2} + \frac{n}{2} + j\right) m^{\frac{m}{2}+j} n^{\frac{n}{2}} w^{\frac{m}{2}+j-1}}{j! \Gamma\left(\frac{m}{2} + j\right) \Gamma\left(\frac{n}{2}\right) (n + mw)^{\frac{m}{2} + \frac{n}{2} + j}}$$

für $0 < w < \infty$ und $f(w) = 0$ für die übrigen Werte von w . (264.1)

Beweis: Da u und v voneinander unabhängig sind, erhält man ihre gemeinsame Verteilung $f(u, v)$ mit (228.2), (261.1) und (262.1) zu

$$f(u, v) = \frac{v^{\frac{n}{2}-1} e^{-\frac{v}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \sum_{j=0}^{\infty} \frac{e^{-\frac{\lambda}{2}} \left(\frac{\lambda}{2}\right)^j u^{\frac{m}{2}+j-1} e^{-\frac{u}{2}}}{j! 2^{\frac{m}{2}+j} \Gamma\left(\frac{m}{2} + j\right)} \quad \text{für } u > 0, v > 0.$$

Transformiert man die Variable v in x mit $v = u/x$ und $|\det \mathbf{J}| = u/x^2$ in (229.1), ergibt sich

$$f(u, x) = \sum_{j=0}^{\infty} G_j \left(\frac{u}{x}\right)^{\frac{n}{2}-1} e^{-\frac{u}{2x}} u^{\frac{m}{2}+j-1} e^{-\frac{u}{2}} \frac{u}{x^2} \quad \text{für } u > 0, x > 0$$

mit

$$G_j = \frac{e^{-\lambda/2} (\lambda/2)^j}{j! 2^{\frac{m}{2} + \frac{n}{2} + j} \Gamma\left(\frac{m}{2} + j\right) \Gamma\left(\frac{n}{2}\right)}.$$

Die Randverteilung $g(x)$ von x erhält man hieraus nach (226.4) zu

$$g(x) = \sum_{j=0}^{\infty} G_j x^{-\frac{n}{2}-1} \int_0^{\infty} u^{\frac{m}{2} + \frac{n}{2} + j - 1} e^{-\frac{u}{2}\left(1 + \frac{1}{x}\right)} du.$$

Durch die Substitution $z = \frac{u}{2}(1 + \frac{1}{x})$ mit $dz = \frac{du}{2}(1 + \frac{1}{x})$ folgt weiter

$$g(x) = \sum_{j=0}^{\infty} G_j x^{-\frac{n}{2}-1} \left(\frac{2x}{x+1}\right)^{\frac{m}{2}+\frac{n}{2}+j} \int_0^{\infty} z^{\frac{m}{2}+\frac{n}{2}+j-1} e^{-z} dz$$

und mit (243.2)

$$g(x) = \sum_{j=0}^{\infty} G_j \Gamma\left(\frac{m}{2} + \frac{n}{2} + j\right) \left(\frac{2}{x+1}\right)^{\frac{m}{2}+\frac{n}{2}+j} x^{\frac{m}{2}+j-1}.$$

Eine letzte Variablentransformation $x = mw/n$ mit $\det \mathbf{J} = m/n$ führt schließlich wegen

$$\frac{m\left(\frac{mw}{n}\right)^{\frac{m}{2}+j-1}}{n\left(\frac{mw}{n} + 1\right)^{\frac{m}{2}+\frac{n}{2}+j}} = \frac{m^{\frac{m}{2}+j} n^{\frac{n}{2}} w^{\frac{m}{2}+j-1}}{(n + mw)^{\frac{m}{2}+\frac{n}{2}+j}}$$

auf die Dichte der nichtzentralen F -Verteilung, so daß die Aussage folgt.

Mit $\lambda = 0$ ergibt sich anstelle der nichtzentralen F -Verteilung (264.1) die F -Verteilung (263.1).

Für $w \sim F'(m, n, \lambda)$ erhält man die Verteilungsfunktion $F(F'; m, n, \lambda)$, indem man in dem Integral über die Dichte die gleiche Variablentransformation vornimmt, die von (263.2) auf (263.3) und (263.4) führt, somit

$$F(F'; m, n, \lambda) = e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} I_{1-x}\left(\frac{m}{2} + j, \frac{n}{2}\right)$$

mit $1 - x = \frac{mF'}{n + mF'}$. (264.2)

Bestimmt man für den Wert F'_j die Verteilungsfunktion der F -Verteilung, die der unvollständigen Betafunktion in (264.2) entspricht, erhält man mit (263.4)

$$1 - x = \frac{mF'}{n + mF'} = \frac{(m + 2j)F'_j}{n + (m + 2j)F'_j}$$

und daher

$$F(F'; m, n, \lambda) = e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\lambda/2)^j}{j!} F(F'_j; m + 2j, n)$$

mit $F'_j = \frac{mF'}{m + 2j}$. (264.3)

Für große Werte von λ konvergieren die Reihen (264.2) und (264.3) langsam, so daß man zweckmäßig die mit der Verteilungsfunktion (263.2) der F -Verteilung zu berechnende Näherung wählt (PATNAIK 1949)

$$F(F'; m, n, \lambda) \approx F(F_0; \nu, n)$$

$$\text{mit } F_0 = \frac{mF'}{m + \lambda}, \nu = \frac{(m + \lambda)^2}{m + 2\lambda}. \quad (264.4)$$

Genauere, aber mit höherem Aufwand zu berechnende Näherungen sind bei (MUDHOLKAR et al. 1976) angegeben. Tafeln mit Werten für die Verteilungsfunktion der nichtzentralen F -Verteilung befinden sich bei (TIKU 1967, 1972).

Ausdrücke mit einer endlichen Anzahl von Termen für $F(F'; m, n, \lambda)$ sind bei (PRICE 1964) angegeben, wo auch die *doppelt nichtzentrale F -Verteilung* behandelt wird, die aus dem Verhältnis zweier unabhängiger Zufallsvariablen folgt, die beide nichtzentrale χ^2 -Verteilungen besitzen.

265 t -Verteilung

Satz: Die Zufallsvariablen y und u mit $y \sim N(0, 1)$ und $u \sim \chi^2(k)$ seien voneinander unabhängig, dann besitzt die Zufallsvariable x mit

$$x = y / \sqrt{u/k}$$

die t -Verteilung $t(k)$, auch *Student-Verteilung* genannt, mit k Freiheitsgraden, also $x \sim t(k)$, mit der Dichte

$$f(x) = \frac{\Gamma(\frac{k+1}{2})}{\sqrt{k\pi} \Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}} \quad \text{für } -\infty < x < \infty. \quad (265.1)$$

Beweis: Die gemeinsame Verteilung $f(y, u)$ von y und u ergibt sich aus (228.2) mit (241.1) und (261.1). Substituiert man in ihr die Variable y durch $x = y / \sqrt{u/k}$ mit $\det \mathbf{J} = \sqrt{u/k}$ in (229.1), ergibt sich die Randverteilung $g(x)$ von x zu

$$g(x) = \frac{1}{\sqrt{2\pi} \Gamma(\frac{k}{2}) 2^{\frac{k}{2}}} \int_0^\infty u^{\frac{k}{2}-1} e^{-\frac{ux^2}{2k} - \frac{u}{2}} \left(\frac{u}{k}\right)^{\frac{1}{2}} du \quad \text{für } -\infty < x < \infty.$$

Beachtet man, daß das Integral über die Gammaverteilung $G(\frac{1}{2}(1 + \frac{x^2}{k}), \frac{k+1}{2})$ in den Grenzen von Null bis Unendlich nach (243.3) den Wert Eins ergibt, folgt für $g(x)$ die Dichte der t -Verteilung und damit die Aussage.

Werte der Verteilungsfunktion der t -Verteilung sind tabuliert (z.B. FISHER und YATES 1963, S.46; PEARSON und HARTLEY 1976, *Vol.I*, S.138), doch

brauchen hier Rechenformeln für die Verteilungsfunktion nicht angegeben zu werden, da t -verteilte Zufallsvariable durch F -verteilte Variable ersetzt werden können. Hierzu dient der

Satz: Die Zufallsvariablen y und u mit $y \sim N(0,1)$ und $u \sim \chi^2(k)$ seien voneinander unabhängig, dann gilt

$$x^2 \sim F(1, k) \quad \text{und} \quad x \sim t(k) \quad \text{mit} \quad x = y/\sqrt{u/k}. \quad (265.2)$$

Beweis: Unter den getroffenen Voraussetzungen gilt nach (261.1) $y^2 \sim \chi^2(1)$, so daß $x^2 = y^2/(u/k)$ nach (263.1) verteilt ist wie $x^2 \sim F(1, k)$. Weiter ist $x = y/\sqrt{u/k}$ nach (265.1) verteilt wie $x \sim t(k)$, so daß die Aussage folgt.

Für $x^2 \sim F(1, k)$ mit der Dichte $f(x^2)$ gilt nach (263.5)

$$P(x^2 < F_{\alpha;1,k}) = \int_0^{F_{\alpha;1,k}} f(x^2) dx^2 = \alpha.$$

Mit der entsprechenden Variablentransformation im Integral dieser Gleichung folgt weiter $P(\pm x < (F_{\alpha;1,k})^{1/2}) = \alpha$ und hiermit wegen (265.2) das dem α -Fraktile der F -Verteilung entsprechende Fraktile $t_{\alpha;k}$ der t -Verteilung

$$t_{\alpha;k} = (F_{\alpha;1,k})^{1/2} \quad (265.3)$$

mit

$$P(-t_{\alpha;k} < x < t_{\alpha;k}) = \alpha, \quad (265.4)$$

denn $-x < t_{\alpha;k}$ entspricht $x > -t_{\alpha;k}$. Da die t -Verteilung wegen (265.1) symmetrisch ist, gilt

$$\int_{-\infty}^{t_{\alpha;k}} f(x) dx = 1 - \frac{1-\alpha}{2} = \frac{1+\alpha}{2}. \quad (265.5)$$

27 Quadratische Formen

271 Erwartungswert und Kovarianz

Quadratische Formen von Zufallsvariablen benötigt man bei Hypothesentests in univariaten Modellen, und um Varianzen für die Schätzwerte von Parametern anzugeben.

Satz: Besitzen die $n \times 1$ Zufallsvektoren \mathbf{x} und \mathbf{y} die Erwartungswertvektoren

$E(\mathbf{x}) = \boldsymbol{\mu}_x$ sowie $E(\mathbf{y}) = \boldsymbol{\mu}_y$ und die Kovarianzmatrizen $D(\mathbf{x}) = \boldsymbol{\Sigma}_{xx}$ sowie $C(\mathbf{x}, \mathbf{y}) = \boldsymbol{\Sigma}_{xy}$, dann gilt

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = \text{sp}(\mathbf{A}\boldsymbol{\Sigma}_{xx}) + \boldsymbol{\mu}_x'\mathbf{A}\boldsymbol{\mu}_x$$

und

$$E(\mathbf{x}'\mathbf{A}\mathbf{y}) = \text{sp}(\mathbf{A}\boldsymbol{\Sigma}_{yx}) + \boldsymbol{\mu}_x'\mathbf{A}\boldsymbol{\mu}_y,$$

worin \mathbf{A} die symmetrische $n \times n$ Matrix der quadratischen beziehungsweise der bilinearen Form bedeutet. (271.1)

Beweis: Mit (137.3), (137.4), (231.5) und (232.8) erhält man

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = E(\text{sp}(\mathbf{A}\mathbf{x}\mathbf{x}')) = \text{sp}(\mathbf{A}E(\mathbf{x}\mathbf{x}')) = \text{sp}(\mathbf{A}(\boldsymbol{\Sigma}_{xx} + \boldsymbol{\mu}_x\boldsymbol{\mu}_x'))$$

und damit die erste Aussage sowie die zweite entsprechend.

Nimmt man den Zufallsvektor \mathbf{x} als normalverteilt an, sind wegen (252.1) sämtliche Momente der Verteilung durch den Erwartungswertvektor $\boldsymbol{\mu}$ und die Kovarianzmatrix $\boldsymbol{\Sigma}$ bestimmt. In diesem Fall ergeben sich auch für die Kovarianz zweier quadratischer Formen des Zufallsvektors und für die Kovarianz einer linearen und einer quadratischen Form des Zufallsvektors Ausdrücke mit $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$.

Satz: Der $n \times 1$ Zufallsvektor \mathbf{x} sei normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, dann gilt für die Kovarianz der beiden quadratischen Formen $\mathbf{x}'\mathbf{A}\mathbf{x}$ und $\mathbf{x}'\mathbf{B}\mathbf{x}$

$$C(\mathbf{x}'\mathbf{A}\mathbf{x}, \mathbf{x}'\mathbf{B}\mathbf{x}) = 2\text{sp}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\mu}. \quad (271.2)$$

Beweis: Mit (232.5) und (271.1) erhält man

$$\begin{aligned} C(\mathbf{x}'\mathbf{A}\mathbf{x}, \mathbf{x}'\mathbf{B}\mathbf{x}) &= E\{[\mathbf{x}'\mathbf{A}\mathbf{x} - E(\mathbf{x}'\mathbf{A}\mathbf{x})][\mathbf{x}'\mathbf{B}\mathbf{x} - E(\mathbf{x}'\mathbf{B}\mathbf{x})]\} \\ &= E\{[\mathbf{x}'\mathbf{A}\mathbf{x} - \text{sp}(\mathbf{A}\boldsymbol{\Sigma}) - \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}][\mathbf{x}'\mathbf{B}\mathbf{x} - \text{sp}(\mathbf{B}\boldsymbol{\Sigma}) - \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu}]\}. \end{aligned}$$

Für den Zufallsvektor $\mathbf{e} = \mathbf{x} - \boldsymbol{\mu}$ gilt $E(\mathbf{e}) = \mathbf{0}$ und mit (233.2) $D(\mathbf{e}) = \boldsymbol{\Sigma}$ sowie mit (255.1) $\mathbf{e} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, so daß mit $\mathbf{x} = \mathbf{e} + \boldsymbol{\mu}$ und (231.5) folgt

$$\begin{aligned} C(\mathbf{x}'\mathbf{A}\mathbf{x}, \mathbf{x}'\mathbf{B}\mathbf{x}) &= E\{[\mathbf{e}'\mathbf{A}\mathbf{e} + 2\boldsymbol{\mu}'\mathbf{A}\mathbf{e} - \text{sp}(\mathbf{A}\boldsymbol{\Sigma})] \\ &\quad [\mathbf{e}'\mathbf{B}\mathbf{e} + 2\boldsymbol{\mu}'\mathbf{B}\mathbf{e} - \text{sp}(\mathbf{B}\boldsymbol{\Sigma})]\} \\ &= E(\mathbf{e}'\mathbf{A}\mathbf{e}\mathbf{e}'\mathbf{B}\mathbf{e}) + 2\boldsymbol{\mu}'\mathbf{A}E(\mathbf{e}\mathbf{e}'\mathbf{B}\mathbf{e}) + 2\boldsymbol{\mu}'\mathbf{B}E(\mathbf{e}\mathbf{e}'\mathbf{A}\mathbf{e}) \\ &\quad - E(\mathbf{e}'\mathbf{A}\mathbf{e})\text{sp}(\mathbf{B}\boldsymbol{\Sigma}) - E(\mathbf{e}'\mathbf{B}\mathbf{e})\text{sp}(\mathbf{A}\boldsymbol{\Sigma}) \\ &\quad + 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\mu} + \text{sp}(\mathbf{A}\boldsymbol{\Sigma})\text{sp}(\mathbf{B}\boldsymbol{\Sigma}) \\ &= E(\mathbf{e}'\mathbf{A}\mathbf{e}\mathbf{e}'\mathbf{B}\mathbf{e}) + 2\boldsymbol{\mu}'\mathbf{A}E(\mathbf{e}\mathbf{e}'\mathbf{B}\mathbf{e}) + 2\boldsymbol{\mu}'\mathbf{B}E(\mathbf{e}\mathbf{e}'\mathbf{A}\mathbf{e}) \\ &\quad + 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\mu} - \text{sp}(\mathbf{A}\boldsymbol{\Sigma})\text{sp}(\mathbf{B}\boldsymbol{\Sigma}). \end{aligned}$$

Mit (131.6) und (252.8) erhält man, da die Matrizen \mathbf{A} , \mathbf{B} und $\mathbf{\Sigma}$ symmetrisch sind

$$\begin{aligned} E(\mathbf{e}'\mathbf{A}\mathbf{e}\mathbf{e}'\mathbf{B}\mathbf{e}) &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij}b_{kl}E(e_i e_j e_k e_l) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n (a_{ij}\sigma_{ji}b_{kl}\sigma_{lk} + a_{ji}\sigma_{ik}b_{kl}\sigma_{lj} + a_{ij}\sigma_{jk}b_{kl}\sigma_{li}) \\ &= \text{sp}(\mathbf{A}\mathbf{\Sigma})\text{sp}(\mathbf{B}\mathbf{\Sigma}) + 2\text{sp}(\mathbf{A}\mathbf{\Sigma}\mathbf{B}\mathbf{\Sigma}) \end{aligned}$$

und weiter mit (252.8) für $k \in \{1, \dots, n\}$

$$E(\mathbf{e}\mathbf{e}'\mathbf{A}\mathbf{e}) = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij}E(e_i e_j e_k) \right) = \mathbf{0} \quad (271.3)$$

sowie entsprechend $E(\mathbf{e}\mathbf{e}'\mathbf{B}\mathbf{e}) = \mathbf{0}$. Diese Ergebnisse in die Gleichung für $C(\mathbf{x}'\mathbf{A}\mathbf{x}, \mathbf{x}'\mathbf{B}\mathbf{x})$ eingesetzt ergeben die Aussage.

Satz: Der Zufallsvektor \mathbf{x} sei normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$, dann gilt für die Kovarianz der linearen Form $\mathbf{A}\mathbf{x}$ und der quadratischen Form $\mathbf{x}'\mathbf{B}\mathbf{x}$

$$C(\mathbf{A}\mathbf{x}, \mathbf{x}'\mathbf{B}\mathbf{x}) = 2\mathbf{A}\mathbf{\Sigma}\mathbf{B}\boldsymbol{\mu}. \quad (271.4)$$

Beweis: Mit (233.11) und (271.1) folgt

$$\begin{aligned} C(\mathbf{A}\mathbf{x}, \mathbf{x}'\mathbf{B}\mathbf{x}) &= E\{(\mathbf{A}\mathbf{x} - \mathbf{A}\boldsymbol{\mu})[\mathbf{x}'\mathbf{B}\mathbf{x} - \boldsymbol{\mu}'\mathbf{B}\boldsymbol{\mu} - \text{sp}(\mathbf{B}\mathbf{\Sigma})]'\} \\ &= E\{(\mathbf{A}\mathbf{x} - \mathbf{A}\boldsymbol{\mu})[(\mathbf{x} - \boldsymbol{\mu})'\mathbf{B}(\mathbf{x} - \boldsymbol{\mu}) + 2(\mathbf{x} - \boldsymbol{\mu})'\mathbf{B}\boldsymbol{\mu} - \text{sp}(\mathbf{B}\mathbf{\Sigma})]\} \\ &= 2\mathbf{A}\mathbf{\Sigma}\mathbf{B}\boldsymbol{\mu} \end{aligned}$$

mit $\mathbf{e} = \mathbf{x} - \boldsymbol{\mu}$ wegen (271.3) und wegen $E(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{0}$.

272 Verteilung der quadratischen Form

Satz: Ist der Zufallsvektor \mathbf{x} normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma})$, so gilt genau dann $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi'^2(\text{rg}\mathbf{A}, \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu})$, wenn $\mathbf{A}\mathbf{\Sigma}$ idempotent ist. (272.1)

Beweis: Die Matrix $\mathbf{A}\mathbf{\Sigma}$ sei idempotent, dann gilt nach (152.1) $\mathbf{A}\mathbf{\Sigma} = \mathbf{A}\mathbf{\Sigma}\mathbf{A}\mathbf{\Sigma}$ und weiter, da $\mathbf{\Sigma}$ positiv definit ist und somit $\mathbf{\Sigma}^{-1}$ wegen (143.3) existiert, $\mathbf{A} = \mathbf{A}\mathbf{\Sigma}\mathbf{A}$. Nach (143.7) ist dann \mathbf{A} positiv definit oder positiv semidefinit, so daß mit $r = \text{rg}\mathbf{A}$ nach (143.10) für \mathbf{A} die Zerlegung $\mathbf{A} = \mathbf{H}\mathbf{H}'$ gilt, in der \mathbf{H} den vollen Spaltenrang $r = \text{rg}\mathbf{H}$ besitzt. Mit dem $r \times 1$ Vektor $\mathbf{z} = \mathbf{H}'\mathbf{x}$ folgt dann $\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{z}'\mathbf{z}$ und nach (255.1) $\mathbf{z} \sim N(\mathbf{H}'\boldsymbol{\mu}, \mathbf{I})$, denn wegen (143.8) existiert $(\mathbf{H}'\mathbf{H})^{-1}$, so daß $\mathbf{H}'\mathbf{\Sigma}\mathbf{H} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{H}\mathbf{H}'\mathbf{\Sigma}\mathbf{H}\mathbf{H}'\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1} =$

$(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{A}\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1} = \mathbf{I}$ gilt. Dann ergibt sich mit (262.1) $\mathbf{z}'\mathbf{z} \sim \chi'^2(r, \boldsymbol{\mu}'\mathbf{H}\mathbf{H}'\boldsymbol{\mu})$ und damit die erste Aussage. Nimmt man andererseits an, daß $\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi'^2(\text{rg}\mathbf{A}, \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu})$ gilt, läßt sich mit Hilfe der momenterzeugenden Funktion von $\mathbf{x}'\mathbf{A}\mathbf{x}$ zeigen, daß $\mathbf{A}\boldsymbol{\Sigma}$ idempotent ist (GRAYBILL 1976, S.135; SEARLE 1971, S.57).

273 Unabhängigkeit zweier quadratischer Formen

Satz: Ist der Zufallsvektor \mathbf{x} normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so sind die beiden quadratischen Formen $\mathbf{x}'\mathbf{A}\mathbf{x}$ und $\mathbf{x}'\mathbf{B}\mathbf{x}$ mit den positiv definiten oder positiv semidefiniten Matrizen \mathbf{A} und \mathbf{B} genau dann voneinander unabhängig, falls $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$ oder $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{0}$ gilt. (273.1)

Beweis: Da \mathbf{A} , \mathbf{B} und $\boldsymbol{\Sigma}$ symmetrisch sind, ergibt sich die Bedingung $\mathbf{B}\boldsymbol{\Sigma}\mathbf{A} = \mathbf{0}$ durch Transponierung von $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$. Es sei $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$ und \mathbf{A} sowie \mathbf{B} positiv semidefinit, dann erhält man mit $\mathbf{A} = \mathbf{C}\mathbf{C}'$ und $\mathbf{B} = \mathbf{D}\mathbf{D}'$ wegen (143.10), wobei \mathbf{C} und \mathbf{D} vollen Spaltenrang besitzen, $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{C}\mathbf{C}'\boldsymbol{\Sigma}\mathbf{D}\mathbf{D}' = \mathbf{0}$ sowie $\mathbf{C}'\mathbf{C}\mathbf{C}'\boldsymbol{\Sigma}\mathbf{D}\mathbf{D}'\mathbf{D} = \mathbf{0}$ und daraus $\mathbf{C}'\boldsymbol{\Sigma}\mathbf{D} = \mathbf{0}$, da $\mathbf{C}'\mathbf{C}$ und $\mathbf{D}'\mathbf{D}$ wegen (143.8) regulär sind. Mit (233.15) folgt dann $C(\mathbf{C}'\mathbf{x}, \mathbf{D}'\mathbf{x}) = \mathbf{C}'\boldsymbol{\Sigma}\mathbf{D} = \mathbf{0}$, so daß wegen (254.1) die Zufallsvektoren $\mathbf{C}'\mathbf{x}$ und $\mathbf{D}'\mathbf{x}$ voneinander unabhängig und daher nach (228.1) auch die quadratischen Formen $\mathbf{x}'\mathbf{C}\mathbf{C}'\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x}$ und $\mathbf{x}'\mathbf{D}\mathbf{D}'\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}$ voneinander unabhängig sind. Nimmt man andererseits für $\mathbf{x}'\mathbf{A}\mathbf{x}$ und $\mathbf{x}'\mathbf{B}\mathbf{x}$ Unabhängigkeit an, so gilt nach (232.7) und (271.2) $C(\mathbf{x}'\mathbf{A}\mathbf{x}, \mathbf{x}'\mathbf{B}\mathbf{x}) = 2\text{sp}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\mu} = 0$ für beliebige Werte von $\boldsymbol{\mu}$, woraus $\text{sp}(\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}) = 0$ und $\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\mu} = 0$ für beliebige Werte von $\boldsymbol{\mu}$ folgt. Hieraus ergibt sich $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$. Entsprechend verläuft der Beweis für positiv definite Matrizen \mathbf{A} und \mathbf{B} .

274 Unabhängigkeit einer linearen Form und einer quadratischen Form

Satz: Ist der Zufallsvektor \mathbf{x} normalverteilt wie $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so sind die lineare Form $\mathbf{A}\mathbf{x}$ und die quadratische Form $\mathbf{x}'\mathbf{B}\mathbf{x}$ mit der positiv definiten oder positiv semidefiniten Matrix \mathbf{B} genau dann voneinander unabhängig, falls $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$ gilt. (274.1)

Beweis: Es sei $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$ und \mathbf{B} positiv semidefinit, dann ergibt sich mit $\mathbf{B} = \mathbf{C}\mathbf{C}'$ wegen (143.10), wobei \mathbf{C} vollen Spaltenrang besitzt, $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{C}\mathbf{C}' = \mathbf{0}$ sowie $\mathbf{A}\boldsymbol{\Sigma}\mathbf{C}\mathbf{C}'\mathbf{C} = \mathbf{0}$ und schließlich $\mathbf{A}\boldsymbol{\Sigma}\mathbf{C} = \mathbf{0}$, da $\mathbf{C}'\mathbf{C}$ wegen (143.8) regulär ist. Mit (233.15) folgt dann $C(\mathbf{A}\mathbf{x}, \mathbf{C}'\mathbf{x}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{C} = \mathbf{0}$, so daß wegen (254.1) die Zufallsvektoren $\mathbf{A}\mathbf{x}$ und $\mathbf{C}'\mathbf{x}$ und daher nach (228.1) auch $\mathbf{A}\mathbf{x}$ und $\mathbf{x}'\mathbf{C}\mathbf{C}'\mathbf{x} = \mathbf{x}'\mathbf{B}\mathbf{x}$ voneinander unabhängig sind. Nimmt man andererseits Unabhängigkeit an, so gilt nach (232.7) und (271.4) $C(\mathbf{A}\mathbf{x}, \mathbf{x}'\mathbf{B}\mathbf{x}) =$

$2\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\mu} = \mathbf{0}$ für beliebige Werte von $\boldsymbol{\mu}$. Hieraus folgt $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{0}$. Entsprechend verläuft der Beweis für eine positiv definite Matrix \mathbf{B} .

28 Testverteilungen für multivariate Modelle der Parameterschätzung

281 Wishart–Verteilung

Der χ^2 –, F – und t –Verteilung für Bereichsschätzungen und Hypothesentests in univariaten Modellen der Parameterschätzung entsprechen die Wishart–Verteilung und die aus ihr abgeleiteten Verteilungen für die Hypothesentests in multivariaten Modellen. Die Wishart–Verteilung selbst stellt die multivariate Verallgemeinerung der χ^2 –Verteilung dar, wie in dem folgenden Beispiel gezeigt wird.

Satz: Die $p \times 1$ Zufallsvektoren \mathbf{x}_k mit $k \in \{1, \dots, n\}$ seien voneinander unabhängig, und jeder Vektor sei normalverteilt wie $\mathbf{x}_k \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Weiter gelte $p \leq n$, und die $p \times n$ Matrix \mathbf{X} sei definiert durch $\mathbf{X} = |\mathbf{x}_1, \dots, \mathbf{x}_n|$. Für die Elemente der $p \times p$ Matrix \mathbf{V} mit $\mathbf{V} = \sum_{k=1}^n \mathbf{x}_k \mathbf{x}'_k = \mathbf{X}\mathbf{X}'$ gilt dann die *Wishart–Verteilung* $W(n, \boldsymbol{\Sigma})$, also $\mathbf{V} \sim W(n, \boldsymbol{\Sigma})$, deren Dichte gegeben ist durch

$$f(\mathbf{V}) = \frac{(\det \mathbf{V})^{\frac{1}{2}} (n-p-1) e^{-\frac{1}{2}\text{sp}(\boldsymbol{\Sigma}^{-1}\mathbf{V})}}{2^{\frac{np}{2}} \pi^{\frac{p(p-1)}{4}} (\det \boldsymbol{\Sigma})^{\frac{n}{2}} \prod_{i=1}^p \Gamma(\frac{n+1-i}{2})},$$

falls \mathbf{V} positiv definit ist, und $f(\mathbf{V}) = 0$ für die übrigen $p \times p$ Matrizen \mathbf{V} . Gilt $\mathbf{x}_k \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$, ergibt sich für \mathbf{V} die *nichtzentrale Wishart–Verteilung* $W'(n, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$ mit der $p \times p$ Matrix $\boldsymbol{\Lambda} = \mathbf{M}\mathbf{M}'\boldsymbol{\Sigma}^{-1}/2$ der Nichtzentralitätsparameter, worin $\mathbf{M} = |\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n|$ ist, also $\mathbf{V} \sim W'(n, \boldsymbol{\Sigma}, \boldsymbol{\Lambda})$. (281.1)

Bevor dieser Satz bewiesen wird, soll ein einfaches Beispiel behandelt werden. **Beispiel:** Mit $p = 1, x_k \sim N(0, \sigma^2)$ und $\mathbf{V} = \sum_{k=1}^n x_k^2 = s$ ergibt sich aus (281.1)

$$f(s) = \frac{s^{(n/2)-1} e^{-s/(2\sigma^2)}}{2^{n/2} \sigma^n \Gamma(n/2)}.$$

Die identische Dichte erhält man aber mit Hilfe der χ^2 –Verteilung für die Zufallsvariable s mit $s/\sigma^2 \sim \chi^2(n)$ wegen (261.4), falls $v = s/\sigma^2$ in (261.1) mit $\det \mathbf{J} = 1/\sigma^2$ aus (229.1) substituiert wird. Die Wishart–Verteilung stellt also die multivariate Verallgemeinerung der χ^2 –Verteilung dar.

Im folgenden werden Matrizen, die die Wishart–Verteilung besitzen, abgekürzt als *Wishart–Matrizen* bezeichnet.

282 Herleitung der Wishart–Verteilung

Zur Herleitung der Wishart–Verteilung wird zunächst angenommen, daß $\mathbf{x}_k \sim N(\mathbf{0}, \mathbf{I})$ gilt, um dann später auf die Normalverteilung mit Σ überzugehen. Mit $\mathbf{V} = \mathbf{X}\mathbf{X}'$ in (281.1) ist \mathbf{V} nach (143.8) positiv definit oder positiv semidefinit, wobei zunächst der erste Fall angenommen wird.

Es sei $\mathbf{V} = (v_{ij})$ und $\mathbf{v}'_i = |v_{i,i+1}, v_{i,i+2}, \dots, v_{ip}|$, dann ergibt sich die Untermatrix $\mathbf{V}_{ii} = (v_{ij})$ von \mathbf{V} mit $i, j \in \{i, \dots, p\}$ zu

$$\mathbf{V}_{ii} = \begin{vmatrix} v_{ii} & \mathbf{v}'_i \\ \mathbf{v}_i & \mathbf{V}_{i+1,i+1} \end{vmatrix}. \quad (282.1)$$

Mit \mathbf{V} ist auch \mathbf{V}_{ii} positiv definit, wie sich durch die entsprechende Wahl des Vektors in der quadratischen Form in (143.1) zeigen läßt. Weiter sei $v_{(i)}$ definiert durch

$$v_{(i)} = v_{ii} - \mathbf{v}'_i \mathbf{V}_{i+1,i+1}^{-1} \mathbf{v}_i. \quad (282.2)$$

Die Verteilungen von \mathbf{v}_i und $v_{(i)}$ sind gesucht.

Die Zeilen von $\mathbf{X} = (x_{ij})$ in (281.1) seien durch die Vektoren $\mathbf{z}_i = |x_{i1}, \dots, x_{in}|'$ bezeichnet. Da $\mathbf{x}_k \sim N(\mathbf{0}, \mathbf{I})$ gilt, sind wegen (254.1) die Zufallsvektoren \mathbf{z}_i für $i \neq j$ voneinander unabhängig. Weiter gilt $x_{ij} \sim N(0, 1)$, so daß $\mathbf{z}_i \sim N(\mathbf{0}, \mathbf{I})$ folgt, da die Zufallsvektoren \mathbf{x}_k voneinander unabhängig sind. Mit der Untermatrix \mathbf{X}_i von \mathbf{X} mit $\mathbf{X}_i = |\mathbf{z}_i, \dots, \mathbf{z}_p|'$ ergibt sich $\mathbf{v}'_i = \mathbf{z}'_i \mathbf{X}'_{i+1}$ und $\mathbf{V}_{ii} = \mathbf{X}_i \mathbf{X}'_i$, so daß mit (255.1) für \mathbf{v}_i die Normalverteilung $N(\mathbf{0}, \mathbf{V}_{i+1,i+1})$ unter der Bedingung folgt, daß die Zufallsvektoren \mathbf{z}_j , die die Zeilen von \mathbf{X}_{i+1} bilden, die festen Werte x_{jk} annehmen. Weiter erhält man $v_{(i)}$ aus (282.2) als quadratische Form von \mathbf{z}_i

$$v_{(i)} = \mathbf{z}'_i \mathbf{z}_i - \mathbf{z}'_i \mathbf{X}'_{i+1} (\mathbf{X}_{i+1} \mathbf{X}'_{i+1})^{-1} \mathbf{X}_{i+1} \mathbf{z}_i.$$

Wegen (272.1) gilt wieder unter der Bedingung, daß \mathbf{X}_{i+1} die festen Werte x_{jk} enthält, für $v_{(i)}$ die χ^2 -Verteilung $\chi^2(n - (p - i))$, da die Matrix $\mathbf{X}'_{i+1} (\mathbf{X}_{i+1} \mathbf{X}'_{i+1})^{-1} \mathbf{X}_{i+1}$ idempotent ist und wegen (137.3) und (152.3) den Rang $p - i$ besitzt, so daß nach (152.4) die Matrix der quadratischen Form $v_{(i)}$ idempotent mit dem Rang $n - (p - i)$ ist. Da außerdem $\mathbf{X}_{i+1} (\mathbf{I} - \mathbf{X}'_{i+1} (\mathbf{X}_{i+1} \mathbf{X}'_{i+1})^{-1} \mathbf{X}_{i+1}) = \mathbf{0}$ gilt, sind \mathbf{v}_i als lineare Form von \mathbf{z}_i und $v_{(i)}$ als quadratische Form von \mathbf{z}_i wegen (274.1) voneinander unabhängig.

Die bedingten Verteilungen von $v_{(i)}$ und \mathbf{v}_i hängen lediglich von $\mathbf{V}_{i+1,i+1}$ ab, die gemeinsame Verteilung von $v_{(1)}, \mathbf{v}_1, v_{(2)}, \dots, v_{(p-1)}, \mathbf{v}_{p-1}, v_{pp}$ ergibt sich

daher mit (228.2), (251.1) und (261.1) zu

$$\prod_{i=1}^{p-1} \left(\frac{v_{(i)}^{\frac{n-(p-i)}{2} - 1} e^{-\frac{1}{2}v_{(i)}} e^{-\frac{1}{2}\mathbf{v}'_i \mathbf{V}_{i+1,i+1}^{-1} \mathbf{v}_i}}{2^{\frac{n-(p-i)}{2}} \Gamma\left(\frac{n-(p-i)}{2}\right) (2\pi)^{\frac{p-i}{2}} (\det \mathbf{V}_{i+1,i+1})^{\frac{1}{2}}} \right) \frac{v_{pp}^{\frac{n}{2} - 1} e^{-\frac{1}{2}v_{pp}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}. \quad (282.3)$$

Durch die Transformation der Variablen $v_{(i)}$ in v_{ii} mittels (282.2), für die in (229.1) $\det \mathbf{J} = 1$ gilt, ergibt sich dann aus (282.3) die gewünschte Verteilung für die Elemente v_{ij} von \mathbf{V} .

Der Exponent von e in (282.3) berechnet sich mit (282.2) zu $-\frac{1}{2}(v_{pp} + \sum_{i=1}^{p-1} v_{ii}) = -\frac{1}{2}\text{sp}\mathbf{V}$. Weiter erhält man mit (136.12) und (282.1)

$$\det \mathbf{V}_{ii} = \det(v_{ii} - \mathbf{v}'_i \mathbf{V}_{i+1,i+1}^{-1} \mathbf{v}_i) \det \mathbf{V}_{i+1,i+1}$$

und daher mit (282.2) $v_{(i)} = \det \mathbf{V}_{ii} / \det \mathbf{V}_{i+1,i+1}$ sowie

$$v_{pp} \prod_{i=1}^{p-1} v_{(i)} = v_{pp} \prod_{i=1}^{p-1} \frac{\det \mathbf{V}_{ii}}{\det \mathbf{V}_{i+1,i+1}} = \det \mathbf{V}_{11} = \det \mathbf{V}.$$

Mit den beiden letzten Beziehungen ergibt sich für den folgenden Ausdruck in (282.3)

$$\begin{aligned} v_{pp}^{\frac{n}{2} - 1} \prod_{i=1}^{p-1} \frac{v_{(i)}^{\frac{n-(p-i)}{2} - 1}}{(\det \mathbf{V}_{i+1,i+1})^{\frac{1}{2}}} &= v_{pp}^{\frac{n}{2} - 1} \prod_{i=1}^{p-1} \frac{(\det \mathbf{V}_{ii})^{\frac{n-(p-i)}{2} - 1}}{(\det \mathbf{V}_{i+1,i+1})^{\frac{n-(p-i)-1}{2}}} \\ &= (\det \mathbf{V})^{\frac{n-p-1}{2}} v_{pp}^{\frac{p-1}{2}} \prod_{i=1}^{p-1} \frac{(\det \mathbf{V}_{ii})^{\frac{i-1}{2}}}{(\det \mathbf{V}_{i+1,i+1})^{\frac{i}{2}}} = (\det \mathbf{V})^{\frac{n-p-1}{2}}. \end{aligned}$$

Die Potenzen von π in (282.3) betragen $\frac{1}{2}(1 + \dots + (p-2) + (p-1)) = \frac{1}{4}p(p-1)$. Weiter gilt

$$\begin{aligned} \Gamma\left(\frac{n}{2}\right) \prod_{i=1}^{p-1} \Gamma\left(\frac{n-(p-i)}{2}\right) &= \\ \Gamma\left(\frac{1}{2}(n-1+1)\right) \Gamma\left(\frac{1}{2}(n-2+1)\right) \dots \Gamma\left(\frac{1}{2}(n-p+1)\right) &= \prod_{i=1}^p \Gamma\left(\frac{n+1-i}{2}\right). \end{aligned}$$

Mit diesen Beziehungen erhält man schließlich aus (282.3) die Dichte für $\mathbf{V} \sim W(n, \mathbf{I})$ zu

$$f(\mathbf{V}) = \frac{\det \mathbf{V}^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\text{sp}\mathbf{V}}}{2^{\frac{np}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^p \Gamma\left(\frac{n+1-i}{2}\right)}. \quad (282.4)$$

Mit $\mathbf{V} = \mathbf{X}\mathbf{X}'$ wird \mathbf{V} nach (143.8) positiv semidefinit, falls die Matrix \mathbf{X} nicht vollen Zeilenrang besitzt. Dann beträgt die Wahrscheinlichkeit $P = 1$, daß nach (122.3) beispielsweise der Zeilenvektor \mathbf{z}_i von \mathbf{X} sich als Linearkombination der übrigen Zeilenvektoren \mathbf{z}_j mit $i \neq j$ darstellen läßt. Die gemeinsame Verteilung aller Zeilenvektoren kann dann aber nicht mehr in das Produkt ihrer Randverteilungen zerlegt werden, was nach (228.2) wegen der Unabhängigkeit der Zeilenvektoren erforderlich wäre. Dies führt auf einen Widerspruch, so daß, da $p \leq n$ gilt, die Matrix \mathbf{V} positiv definit ist.

Um die Dichte in (281.1) zu erhalten, soll jetzt $\mathbf{x}_k \sim N(\mathbf{0}, \mathbf{\Sigma})$ mit $\mathbf{X} = |\mathbf{x}_1, \dots, \mathbf{x}_n|$ und $\mathbf{V} = \mathbf{X}\mathbf{X}'$ gelten. Für die positiv definite Matrix $\mathbf{\Sigma}$ erhält man wegen (143.5) $\mathbf{\Sigma} = \mathbf{G}^{-1}(\mathbf{G}')^{-1}$, worin \mathbf{G} eine reguläre untere Dreiecksmatrix bedeutet. Mit $\mathbf{G}\mathbf{\Sigma}\mathbf{G}' = \mathbf{I}$ folgt nach (255.1) für die Zufallsvektoren $\mathbf{G}\mathbf{x}_k \sim N(\mathbf{0}, \mathbf{I})$ und für die Matrix $\mathbf{G}\mathbf{X}\mathbf{X}'\mathbf{G}' = \mathbf{G}\mathbf{V}\mathbf{G}' = \mathbf{V}^*$ die Verteilung $\mathbf{V}^* \sim W(n, \mathbf{I})$ mit der Dichte (282.4), falls dort \mathbf{V} durch \mathbf{V}^* ersetzt wird. Um die Verteilung für \mathbf{V} zu erhalten, wird daher \mathbf{V}^* in (282.4) transformiert in $\mathbf{V} = \mathbf{G}^{-1}\mathbf{V}^*(\mathbf{G}')^{-1}$. Mit (173.5) und (229.1) ergibt sich dann für die Dichte $f(\mathbf{V})$ aus (282.4)

$$f(\mathbf{V}) = \frac{\det(\mathbf{G}\mathbf{V}\mathbf{G}')^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2}\text{sp}(\mathbf{G}\mathbf{V}\mathbf{G}')} (\det \mathbf{G})^{p+1}}{2^{\frac{np}{2}} \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^p \Gamma(\frac{n+1-i}{2})}, \quad (282.5)$$

da $|\det \mathbf{G}| = \det \mathbf{G}$ wegen (143.4) und (143.5) gilt. Mit $\det(\mathbf{G}\mathbf{\Sigma}\mathbf{G}') = \det \mathbf{\Sigma}$ ($\det \mathbf{G}$)² = 1 wegen (136.8) und (136.13) ergibt sich $|\det \mathbf{G}| = \det \mathbf{G} = +(\det \mathbf{\Sigma})^{-1/2}$ und damit $\det(\mathbf{G}\mathbf{V}\mathbf{G}') = \det \mathbf{V}(\det \mathbf{\Sigma})^{-1}$. Weiter gilt $\text{sp}(\mathbf{G}\mathbf{V}\mathbf{G}') = \text{sp}(\mathbf{G}'\mathbf{G}\mathbf{V}) = \text{sp}(\mathbf{\Sigma}^{-1}\mathbf{V})$ wegen (137.3) und (131.14). Mit diesen Beziehungen erhält man schließlich anstelle von (282.5) die Dichte $f(\mathbf{V})$ in (281.1).

Gilt $\mathbf{x}_k \sim N(\boldsymbol{\mu}_k, \mathbf{\Sigma})$, gehen in (282.3) die Dichten der nichtzentralen χ^2 -Verteilungen ein, deren Nichtzentralitätsparameter von den Erwartungswertvektoren $\boldsymbol{\mu}_k$ abhängen. Es ergibt sich $\mathbf{V} \sim W'(n, \mathbf{\Sigma}, \mathbf{\Lambda})$ mit $\mathbf{\Lambda} = \mathbf{M}\mathbf{M}'\mathbf{\Sigma}^{-1}/2$ und $\mathbf{M} = |\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n|$ (CONSTANTINE 1963; JAMES 1964). Aus dieser Dichte folgt mit $\mathbf{\Lambda} = \mathbf{0}$ wieder die Dichte der zentralen Wishart-Verteilung.

283 Verteilung der Summe von Wishart-Matrizen

Die Wishart-Verteilung besitzt wie die Gammaverteilung und die Normalverteilung die reproduzierende Eigenschaft.

Satz: Die beiden $p \times p$ Matrizen \mathbf{V}_1 und \mathbf{V}_2 mit $\mathbf{V}_1 \sim W(n_1, \mathbf{\Sigma})$ und $\mathbf{V}_2 \sim W(n_2, \mathbf{\Sigma})$ seien voneinander unabhängig, dann besitzt die Matrix $\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2$ die Wishart-Verteilung $\mathbf{V} \sim W(n_1 + n_2, \mathbf{\Sigma})$. (283.1)

Beweis: Es seien \mathbf{V}_1 und \mathbf{V}_2 mit $\mathbf{V}_1 \sim W(n_1, \mathbf{\Sigma})$ und $\mathbf{V}_2 \sim W(n_2, \mathbf{\Sigma})$

voneinander unabhängig, dann besitzt nach (281.1) \mathbf{V}_1 dieselbe Verteilung wie $\sum_{k=1}^{n_1} \mathbf{x}_k \mathbf{x}'_k$ und \mathbf{V}_2 dieselbe Verteilung wie $\sum_{k=n_1+1}^{n_1+n_2} \mathbf{x}_k \mathbf{x}'_k$, wobei die Zufallsvektoren \mathbf{x}_k mit $\mathbf{x}_k \sim N(\mathbf{0}, \mathbf{\Sigma})$ und $k \in \{1, \dots, n_1 + n_2\}$ voneinander unabhängig seien, da \mathbf{V}_1 und \mathbf{V}_2 voneinander unabhängig sein sollen. Die Summe $\mathbf{V}_1 + \mathbf{V}_2$ ist dann verteilt wie die Matrix $\sum_{k=1}^{n_1+n_2} \mathbf{x}_k \mathbf{x}'_k$, für die $W(n_1 + n_2, \mathbf{\Sigma})$ gilt, so daß die Aussage folgt.

284 Verteilung der transformierten Wishart–Matrix

Wie im Kapitel 412 gezeigt wird, besitzt die Schätzung einer Kovarianzmatrix die Wishart–Verteilung. Es erhebt sich daher die Frage nach der Verteilung einer nach (233.2) transformierten Wishart–Matrix.

Satz: Die $p \times p$ positiv definite Matrix \mathbf{V} sei verteilt wie $\mathbf{V} \sim W(n, \mathbf{\Sigma})$, dann gilt für die Matrix $\mathbf{U} = \mathbf{A}\mathbf{V}\mathbf{A}'$, worin \mathbf{A} vollen Zeilenrang besitze, die Wishart–Verteilung $\mathbf{U} \sim W(n, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$. (284.1)

Beweis: Es sei $\mathbf{V} \sim W(n, \mathbf{\Sigma})$, dann besitzt nach (281.1) \mathbf{V} dieselbe Verteilung wie $\sum_{k=1}^n \mathbf{x}_k \mathbf{x}'_k$, wobei die $p \times 1$ Zufallsvektoren \mathbf{x}_k mit $\mathbf{x}_k \sim N(\mathbf{0}, \mathbf{\Sigma})$ voneinander unabhängig sind. Ferner gelte $\mathbf{V} = \sum_{k=1}^n \mathbf{x}_k \mathbf{x}'_k = \mathbf{X}\mathbf{X}'$, denn \mathbf{V} soll positiv definit sein. Mit $\mathbf{y}_k = \mathbf{A}\mathbf{x}_k$ ergibt sich aus (255.1) $\mathbf{y}_k \sim N(\mathbf{0}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$, wobei wegen (254.1) die \mathbf{y}_k voneinander unabhängig sind, da mit (233.15) $C(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{A}C(\mathbf{x}_i, \mathbf{x}_j)\mathbf{A}' = \mathbf{0}$ wegen $C(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{0}$ für $i \neq j$ gilt. Mit $\mathbf{Y} = |\mathbf{y}_1, \dots, \mathbf{y}_n|$ folgt dann $\mathbf{U} = \mathbf{Y}\mathbf{Y}' = \mathbf{A}\mathbf{X}\mathbf{X}'\mathbf{A}' = \mathbf{A}\mathbf{V}\mathbf{A}'$ und wegen (281.1) $\mathbf{U} \sim W(n, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$, worin $\mathbf{A}\mathbf{\Sigma}\mathbf{A}'$ nach (143.7) positiv definit ist, da $\mathbf{\Sigma}$ positiv definit ist und \mathbf{A} vollen Zeilenrang besitzt.

285 Verteilung der Matrizen quadratischer Formen und Unabhängigkeit der Wishart–Matrizen

Wie im Beispiel zu (281.1) gezeigt, stellt die Wishart–Verteilung die multivariate Verallgemeinerung der χ^2 –Verteilung dar. Man kann daher mit Hilfe des folgenden Satzes aus der χ^2 –Verteilung quadratischer Formen, die in univariaten Modellen der Parameterschätzung benutzt werden, die Wishart–Verteilung von Matrizen bestimmen, deren Diagonalelemente aus quadratischen Formen gebildet werden und die bei der multivariaten Parameterschätzung auftreten. Ebenso läßt sich aus der Unabhängigkeit quadratischer Formen auf die Unabhängigkeit von Wishart–Matrizen schließen.

Satz: Die $p \times 1$ Zufallsvektoren \mathbf{x}_k mit $k \in \{1, \dots, n\}$ seien voneinander unabhängig und normalverteilt wie $\mathbf{x}_k \sim N(\boldsymbol{\mu}_k, \mathbf{\Sigma})$. Weiter sei die $n \times n$ Matrix \mathbf{A} symmetrisch, und die $p \times n$ Matrizen \mathbf{X} und \mathbf{M} seien definiert durch $\mathbf{X} = |\mathbf{x}_1, \dots, \mathbf{x}_n|$ und $\mathbf{M} = |\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_n|$. Genau dann besitzt die Matrix

\mathbf{XAX}' die nichtzentrale Wishart-Verteilung

$$\mathbf{XAX}' \sim W'(r, \mathbf{\Sigma}, \mathbf{MAM}'\mathbf{\Sigma}^{-1}/2),$$

wenn für alle $p \times 1$ Vektoren \mathbf{b} für die quadratische Form $\mathbf{b}'\mathbf{XAX}'\mathbf{b}/\sigma_b^2 \sim \chi'^2(r, \lambda)$ mit $\sigma_b^2 = \mathbf{b}'\mathbf{\Sigma}\mathbf{b}$ gilt. In diesem Fall ist \mathbf{A} idempotent, $r = \text{rg}\mathbf{A}$ und $\lambda = \mathbf{b}'\mathbf{MAM}'\mathbf{b}/\sigma_b^2$. Weiter gilt $\mathbf{XAX}' \sim W(r, \mathbf{\Sigma})$ genau dann, wenn $\mathbf{b}'\mathbf{XAX}'\mathbf{b}/\sigma_b^2 \sim \chi^2(r)$ ist. (285.1)

Beweis: Es gelte $\mathbf{XAX}' \sim W'(r, \mathbf{\Sigma}, \mathbf{MAM}'\mathbf{\Sigma}^{-1}/2)$, dann besitzt wegen (281.1) \mathbf{XAX}' dieselbe Verteilung wie $\sum_{k=1}^r \mathbf{y}_k \mathbf{y}_k'$, wobei die r Zufallsvektoren \mathbf{y}_k voneinander unabhängig sind und wie $\mathbf{y}_k \sim N(\mathbf{n}_k, \mathbf{\Sigma})$ mit $\sum_{k=1}^r \mathbf{n}_k \mathbf{n}_k' = \mathbf{MAM}'$ verteilt sind. Definiert man $\mathbf{Y} = |\mathbf{y}_1, \dots, \mathbf{y}_r|$ und $\mathbf{N} = |\mathbf{n}_1, \dots, \mathbf{n}_r|$, besitzt also \mathbf{XAX}' dieselbe Verteilung wie \mathbf{YY}' , für die

$$\mathbf{YY}' \sim W'(r, \mathbf{\Sigma}, \mathbf{NN}'\mathbf{\Sigma}^{-1}/2) \quad \text{mit} \quad \mathbf{NN}' = \mathbf{MAM}'$$

gilt. Mit $\sigma_b^2 = \mathbf{b}'\mathbf{\Sigma}\mathbf{b}$ ergibt sich für $\mathbf{b}'\mathbf{y}_k/\sigma_b$ nach (255.1) die Verteilung $\mathbf{b}'\mathbf{y}_k/\sigma_b \sim N(\mathbf{b}'\mathbf{n}_k/\sigma_b, 1)$. Faßt man die r voneinander unabhängigen Skalare $\mathbf{b}'\mathbf{y}_k$ im Vektor $\mathbf{z}' = \mathbf{b}'\mathbf{Y}$ zusammen, erhält man wegen (254.1) $\mathbf{z}/\sigma_b \sim N(\mathbf{N}'\mathbf{b}/\sigma_b, \mathbf{I})$. Dann folgt aus (262.1) $\mathbf{z}'\mathbf{z}/\sigma_b^2 \sim \chi'^2(r, \lambda)$ mit $\lambda = \mathbf{b}'\mathbf{NN}'\mathbf{b}/\sigma_b^2 = \mathbf{b}'\mathbf{MAM}'\mathbf{b}/\sigma_b^2$ und daher die erste Aussage, da $\mathbf{b}'\mathbf{XAX}'\mathbf{b}/\sigma_b^2$ dieselbe Verteilung besitzt wie $\mathbf{z}'\mathbf{z}/\sigma_b^2 = \mathbf{b}'\mathbf{YY}'\mathbf{b}/\sigma_b^2$. Gilt $\mathbf{XAX}' \sim W(r, \mathbf{\Sigma})$, dann sind die Zufallsvektoren \mathbf{y}_k verteilt wie $\mathbf{y}_k \sim N(\mathbf{0}, \mathbf{\Sigma})$, so daß sich $\lambda = 0$, $\mathbf{z}'\mathbf{z} \sim \chi^2(r)$ und die nächste Aussage ergibt. Gilt andererseits $\mathbf{b}'\mathbf{XAX}'\mathbf{b}/\sigma_b^2 \sim \chi'^2(r, \lambda)$, so ist zu beachten, daß $\mathbf{b}'\mathbf{x}_k/\sigma_b \sim N(\mathbf{b}'\boldsymbol{\mu}_k/\sigma_b, 1)$ für $k \in \{1, \dots, n\}$ und wegen der Unabhängigkeit der Skalare $\mathbf{b}'\mathbf{x}_k/\sigma_b$ außerdem $\mathbf{X}'\mathbf{b}/\sigma_b \sim N(\mathbf{M}'\mathbf{b}/\sigma_b, \mathbf{I})$ gilt, so daß aus (272.1) folgt, daß \mathbf{A} idempotent ist und $r = \text{rg}\mathbf{A}$ sowie $\lambda = \mathbf{b}'\mathbf{MAM}'\mathbf{b}/\sigma_b^2$ gilt. Nach (152.7) läßt sich dann \mathbf{A} mit r orthonormalen $n \times 1$ Vektoren \mathbf{p}_i darstellen durch $\mathbf{A} = \mathbf{p}_1\mathbf{p}_1' + \dots + \mathbf{p}_r\mathbf{p}_r'$ und somit

$$\mathbf{XAX}' = \mathbf{X}\mathbf{p}_1\mathbf{p}_1'\mathbf{X}' + \dots + \mathbf{X}\mathbf{p}_r\mathbf{p}_r'\mathbf{X}' = \sum_{k=1}^r \mathbf{v}_k\mathbf{v}_k',$$

wobei $\mathbf{v}_k = \mathbf{X}\mathbf{p}_k$ und $\mathbf{p}_k = |p_{1k}, \dots, p_{nk}|'$ bedeuten. Für die Zufallsvektoren \mathbf{v}_k gilt $\mathbf{v}_k = p_{1k}\mathbf{x}_1 + \dots + p_{nk}\mathbf{x}_n$ und nach (256.1) $\mathbf{v}_k \sim N(\mathbf{M}\mathbf{p}_k, \mathbf{\Sigma})$ wegen $\sum_{j=1}^n p_{jk}\boldsymbol{\mu}_j = \mathbf{M}\mathbf{p}_k$ und wegen $\sum_{j=1}^n p_{jk}^2 = 1$, da die Vektoren \mathbf{p}_k orthonormal sind. Außerdem sind die Vektoren \mathbf{v}_k voneinander unabhängig, denn man erhält mit (233.15), da $C(\mathbf{x}_k, \mathbf{x}_l) = \mathbf{0}$ für $k \neq l$ gilt, $C(\mathbf{v}_k, \mathbf{v}_l) = C(p_{1k}\mathbf{x}_1 + \dots + p_{nk}\mathbf{x}_n, p_{1l}\mathbf{x}_1 + \dots + p_{nl}\mathbf{x}_n) = p_{1k}p_{1l}D(\mathbf{x}_1) + \dots + p_{nk}p_{nl}D(\mathbf{x}_n)$

und hieraus $C(\mathbf{v}_k, \mathbf{v}_l) = \mathbf{0}$, da $\mathbf{p}'_k \mathbf{p}_l = 0$ gilt. Dann ergibt sich nach (281.1)

$$\sum_{k=1}^r \mathbf{v}_k \mathbf{v}'_k = \mathbf{X} \mathbf{A} \mathbf{X}' \sim W'(r, \mathbf{\Sigma}, \frac{1}{2} (\sum_{k=1}^r \mathbf{M} \mathbf{p}_k \mathbf{p}'_k \mathbf{M}') \mathbf{\Sigma}^{-1}) = \\ W'(r, \mathbf{\Sigma}, \mathbf{M} \mathbf{A} \mathbf{M}' \mathbf{\Sigma}^{-1} / 2)$$

und damit die weitere Aussage. Gilt schließlich $\mathbf{b}' \mathbf{X} \mathbf{A} \mathbf{X}' \mathbf{b} / \sigma_b^2 \sim \chi^2(r)$, so folgt $\lambda = \mathbf{b}' \mathbf{M} \mathbf{A} \mathbf{M}' \mathbf{b} / \sigma_b^2 = 0$ und daher $\mathbf{M} \mathbf{A} \mathbf{M}' = \mathbf{0}$, da \mathbf{b} beliebig ist. Weiter ergibt sich $\mathbf{M} \mathbf{A} \mathbf{M}' = \sum_{k=1}^r \mathbf{M} \mathbf{p}_k \mathbf{p}'_k \mathbf{M}' = \mathbf{0}$ und somit wegen (131.11) $\mathbf{M} \mathbf{p}_k = \mathbf{0}$. Dann gilt $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{\Sigma})$ und wegen (281.1) $\sum_{k=1}^r \mathbf{v}_k \mathbf{v}'_k = \mathbf{X} \mathbf{A} \mathbf{X}' \sim W'(r, \mathbf{\Sigma})$, so daß die letzte Aussage erhalten wird.

Die Unabhängigkeit von Wishart–Matrizen wird im folgenden Satz behandelt.

Satz: Unter den Voraussetzungen des Satzes (285.1) sind die Matrizen $\mathbf{X} \mathbf{A}_1 \mathbf{X}'$ und $\mathbf{X} \mathbf{A}_2 \mathbf{X}'$ mit Wishart–Verteilungen genau dann voneinander unabhängig, wenn für alle Vektoren \mathbf{b} die quadratischen Formen $\mathbf{b}' \mathbf{X} \mathbf{A}_1 \mathbf{X}' \mathbf{b}$ und $\mathbf{b}' \mathbf{X} \mathbf{A}_2 \mathbf{X}' \mathbf{b}$ mit χ^2 –Verteilungen voneinander unabhängig sind. (285.2)

Beweis: Die Matrizen $\mathbf{X} \mathbf{A}_1 \mathbf{X}'$ und $\mathbf{X} \mathbf{A}_2 \mathbf{X}'$ mit Wishart–Verteilungen seien voneinander unabhängig, dann läßt sich wie im Beweis zu (285.1) zeigen, daß $\mathbf{X} \mathbf{A}_1 \mathbf{X}'$ dieselbe Verteilung besitzt wie $\mathbf{Y}_1 \mathbf{Y}'_1$ und $\mathbf{X} \mathbf{A}_2 \mathbf{X}'$ dieselbe Verteilung wie $\mathbf{Y}_2 \mathbf{Y}'_2$, wobei die Zufallsvektoren, die \mathbf{Y}_1 bilden, unabhängig von den Zufallsvektoren sind, die \mathbf{Y}_2 bilden. Somit sind nach (254.1) und (255.1) auch die Zufallsvektoren $\mathbf{z}_1 = \mathbf{Y}'_1 \mathbf{b}$ und $\mathbf{z}_2 = \mathbf{Y}'_2 \mathbf{b}$ voneinander unabhängig, so daß sich für $\mathbf{z}'_1 \mathbf{z}_1$ und $\mathbf{z}'_2 \mathbf{z}_2$ und damit für $\mathbf{b}' \mathbf{X} \mathbf{A}_1 \mathbf{X}' \mathbf{b}$ und $\mathbf{b}' \mathbf{X} \mathbf{A}_2 \mathbf{X}' \mathbf{b}$ voneinander unabhängige χ^2 –Verteilungen ergeben. Sind andererseits die beiden quadratischen Formen mit χ^2 –Verteilungen voneinander unabhängig, so sind die im Beweis zu (285.1) definierten Zufallsvektoren $\mathbf{X}' \mathbf{b}$ und \mathbf{v}_k der ersten quadratischen Form von denen der zweiten quadratischen Form unabhängig, so daß voneinander unabhängige Matrizen $\mathbf{X} \mathbf{A}_1 \mathbf{X}'$ und $\mathbf{X} \mathbf{A}_2 \mathbf{X}'$ mit Wishart–Verteilungen folgen.

286 Verteilung des Verhältnisses der Determinanten zweier Wishart–Matrizen

Wie mit (425.3) gezeigt wird, treten bei Hypothesenprüfungen in multivariaten Modellen der Parameterschätzung als Testgrößen die Verhältnisse der Determinanten zweier Matrizen auf, die die Wishart–Verteilung besitzen.

Satz: Die $p \times p$ Matrizen \mathbf{U} mit $\mathbf{U} \sim W(m, \mathbf{\Sigma})$ und \mathbf{V} mit $\mathbf{V} \sim W(n, \mathbf{\Sigma})$ seien voneinander unabhängig, dann ist die Größe $\Lambda_{p,n,m} = \det \mathbf{U} / \det(\mathbf{U} + \mathbf{V})$ verteilt wie das Produkt $\prod_{i=1}^p x_i$ voneinander unabhängiger Zufallsvariablen

x_i , die die Betaverteilungen $x_i \sim B(\frac{1}{2}(m+1-i), \frac{1}{2}n)$ besitzen. (286.1)

Beweis: Wie bereits im Zusammenhang mit (245.2) erwähnt, besitzt die momenterzeugende Funktion der Betaverteilung keine einfache Form, so daß der Beweis mit Hilfe der Momente geführt wird. Das h -te Moment von $\Lambda_{p,n,m}$ ergibt sich mit (228.2), (232.2) und (281.1) sowie mit $\mathbf{U} = (u_{ij})$ und $\mathbf{V} = (v_{ij})$ zu

$$\begin{aligned} E(\Lambda_{p,n,m}^h) &= E((\det \mathbf{U})^h (\det(\mathbf{U} + \mathbf{V}))^{-h}) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\det \mathbf{U})^h (\det(\mathbf{U} + \mathbf{V}))^{-h} K(m, \mathbf{\Sigma}) (\det \mathbf{U})^{\frac{1}{2}(m-p-1)} \\ &\quad \exp(-\frac{1}{2} \text{sp}(\mathbf{\Sigma}^{-1} \mathbf{U})) K(n, \mathbf{\Sigma}) (\det \mathbf{V})^{\frac{1}{2}(n-p-1)} \\ &\quad \exp(-\frac{1}{2} \text{sp}(\mathbf{\Sigma}^{-1} \mathbf{V})) d\mathbf{U} d\mathbf{V} \end{aligned}$$

mit $d\mathbf{U} = du_{11} du_{12} \dots du_{pp}$, $d\mathbf{V} = dv_{11} dv_{12} \dots dv_{pp}$ und

$$K^{-1}(m, \mathbf{\Sigma}) = 2^{\frac{mp}{2}} \pi^{\frac{p(p-1)}{4}} (\det \mathbf{\Sigma})^{\frac{m}{2}} \prod_{i=1}^p \Gamma\left(\frac{m+1-i}{2}\right) \quad (286.2)$$

sowie $K(n, \mathbf{\Sigma})$ entsprechend. Weiter gilt

$$\begin{aligned} E(\Lambda_{p,n,m}^h) &= \frac{K(m, \mathbf{\Sigma})}{K(m+2h, \mathbf{\Sigma})} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\det(\mathbf{U} + \mathbf{V}))^{-h} K(m+2h, \mathbf{\Sigma}) \\ &\quad (\det \mathbf{U})^{\frac{1}{2}(m+2h-p-1)} \exp(-\frac{1}{2} \text{sp}(\mathbf{\Sigma}^{-1} \mathbf{U})) K(n, \mathbf{\Sigma}) (\det \mathbf{V})^{\frac{1}{2}(n-p-1)} \\ &\quad \exp(-\frac{1}{2} \text{sp}(\mathbf{\Sigma}^{-1} \mathbf{V})) d\mathbf{U} d\mathbf{V} . \end{aligned}$$

Der Integralausdruck gleicht $E(\det(\mathbf{U} + \mathbf{V})^{-h})$ mit $\mathbf{U} \sim W(m+2h, \mathbf{\Sigma})$ und $\mathbf{V} \sim W(n, \mathbf{\Sigma})$, da \mathbf{U} und \mathbf{V} voneinander unabhängig sind. Weiter gilt nach (283.1) $\mathbf{U} + \mathbf{V} \sim W(m+2h+n, \mathbf{\Sigma})$, so daß man mit (232.2) und (281.1) erhält

$$\begin{aligned} E(\det(\mathbf{U} + \mathbf{V})^{-h}) &= \frac{K(m+2h+n, \mathbf{\Sigma})}{K(m+n, \mathbf{\Sigma})} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\det(\mathbf{U} + \mathbf{V}))^{-h} \\ &\quad K(m+n, \mathbf{\Sigma}) \det(\mathbf{U} + \mathbf{V})^{\frac{1}{2}(m+2h+n-p-1)} \\ &\quad \exp[-\frac{1}{2} \text{sp}(\mathbf{\Sigma}^{-1}(\mathbf{U} + \mathbf{V}))] d(\mathbf{U} + \mathbf{V}) . \end{aligned}$$

Der Integrand gleicht der Dichte der Wishart-Verteilung $W(m+n, \mathbf{\Sigma})$, so daß das Integral wegen (225.6) den Wert Eins besitzt. Folglich ergibt sich mit

(286.2)

$$\begin{aligned}
E(\Lambda_{p,n,m}^h) &= \frac{K(m, \Sigma)K(m+2h+n, \Sigma)}{K(m+2h, \Sigma)K(m+n, \Sigma)} \\
&= \prod_{i=1}^p \frac{\Gamma(\frac{1}{2}(m+1-i)+h)\Gamma(\frac{1}{2}(m+n+1-i))}{\Gamma(\frac{1}{2}(m+1-i))\Gamma(\frac{1}{2}(m+n+1-i)+h)} = \prod_{i=1}^p E(x_i^h)
\end{aligned}$$

mit $x_i \sim B(\frac{1}{2}(m+1-i), \frac{1}{2}n)$ wegen (245.2), so daß die Aussage folgt, da die Verteilung einer für ein endliches Intervall definierten Zufallsvariablen durch ihre Momente eindeutig bestimmt ist (CRAMÉR 1946, S.177).

Mit $p = 1$ ergibt sich für $\Lambda_{p,n,m}$ die Beta-Verteilung und mit (263.3) die Verbindung zur F -Verteilung, die auch für $p = 2$ besteht. Es gilt (ANDERSON 1958, S.195)

$$\begin{aligned}
\frac{1 - \Lambda_{1,n,m}}{\Lambda_{1,n,m}} \frac{m}{n} &\sim F(n, m) \\
\text{und } \frac{1 - \Lambda_{p,1,m}}{\Lambda_{p,1,m}} \frac{m+1-p}{p} &\sim F(p, m+1-p) \quad (286.3)
\end{aligned}$$

sowie

$$\begin{aligned}
\frac{1 - \Lambda_{2,n,m}^{1/2}}{\Lambda_{2,n,m}^{1/2}} \frac{m-1}{n} &\sim F(2n, 2m-2) \\
\text{und } \frac{1 - \Lambda_{p,2,m}^{1/2}}{\Lambda_{p,2,m}^{1/2}} \frac{m+1-p}{p} &\sim F(2p, 2(m+1-p)) \quad (286.4)
\end{aligned}$$

Eine exakte Integraldarstellung der Verteilung für $\Lambda_{p,n,m}$ wird bei (CONSUL 1969) angegeben. Für die numerische Berechnung von Werten der Verteilungsfunktion von $\Lambda_{p,n,m}$ eignen sich genäherte Verteilungen, beispielsweise mit Hilfe der χ^2 -Verteilung auf der Grundlage einer asymptotischen Entwicklung (ANDERSON 1958, S.208) oder mit Hilfe der F -Verteilung, die auch für univariate Hypothesenprüfungen benötigt wird. Es gilt (RAO 1973, S.556)

$$\frac{1 - \Lambda_{p,n,m}^{1/s}}{\Lambda_{p,n,m}^{1/s}} \frac{os - 2\lambda}{np} \sim F(np, os - 2\lambda) \quad (286.5)$$

mit

$$o = m + \frac{1}{2}(n - p - 1), \quad s = \left(\frac{n^2 p^2 - 4}{n^2 + p^2 - 5} \right)^{1/2}, \quad \lambda = \frac{np - 2}{4}.$$

Der Freiheitsgrad $os - 2\lambda$ der F -Verteilung in (286.5) wird im allgemeinen keine ganze Zahl sein, was aber bei einer Berechnung der Verteilungsfunktion der F -Verteilung nach (263.3) oder (263.4) unerheblich ist.

Werte der Verteilungsfunktion von $\Lambda_{p,n,m}$ ergeben sich aus (286.5) folgendermaßen. Setzt man

$$w = \frac{1 - \Lambda_{p,n,m}^{1/s}}{\Lambda_{p,n,m}^{1/s}} \frac{os - 2\lambda}{np} \quad \text{und} \quad F_0 = \frac{1 - \Lambda_{o,p,n,m}^{1/s}}{\Lambda_{o,p,n,m}^{1/s}} \frac{os - 2\lambda}{np}, \quad (286.6)$$

worin $\Lambda_{o,p,n,m}$ einen Wert von $\Lambda_{p,n,m}$ bedeutet, für den die Verteilungsfunktion ermittelt werden soll, so gilt mit (263.2), falls $f(w)$ die Dichte der F -verteilten Zufallsvariablen w bezeichnet,

$$P(w > F_0) = \int_{F_0}^{\infty} f(w)dw = 1 - F(F_0; np, os - 2\lambda).$$

Durch entsprechende Variablentransformation im Integral dieser Gleichung erhält man

$$\begin{aligned} P(w > F_0) &= P\left(\frac{np}{os - 2\lambda}w > \frac{np}{os - 2\lambda}F_0\right) = \\ P\left(1/(1 + \frac{np}{os - 2\lambda}w) < 1/(1 + \frac{np}{os - 2\lambda}F_0)\right) &= P(\Lambda_{p,n,m} < \Lambda_{o,p,n,m}), \end{aligned}$$

denn es gilt

$$\Lambda_{p,n,m} = \left(1/(1 + \frac{np}{os - 2\lambda}w)\right)^s. \quad (286.7)$$

Somit ergeben sich die Werte der Verteilungsfunktion von $\Lambda_{p,n,m}$ aus

$$P(\Lambda_{p,n,m} < \Lambda_{o,p,n,m}) = 1 - F(F_0; np, os - 2\lambda) \quad (286.8)$$

mit F_0 aus (286.6). Setzt man F_0 gleich dem $(1 - \alpha)$ -Fraktile der F -Verteilung, so kann man das α -Fraktile $\Lambda_{\alpha;p,n,m}$ der Verteilung von $\Lambda_{p,n,m}$ aus (286.7) und (286.8) berechnen

$$\Lambda_{\alpha;p,n,m} = \left(1/(1 + \frac{np}{os - 2\lambda}F_{1-\alpha;np,os-2\lambda})\right)^s. \quad (286.9)$$

Tafeln für Werte von $\Lambda_{\alpha;p,n,m}$ befinden sich in (KRES 1983, S.14).

Gilt für die $p \times p$ Matrix \mathbf{U} die Wishart-Verteilung $\mathbf{U} \sim W(m, \mathbf{\Sigma})$, für die $p \times p$ Matrix \mathbf{V} aber im Gegensatz zu (286.1) die nichtzentrale Wishart-Verteilung $\mathbf{V} \sim W'(n, \mathbf{\Sigma}, \mathbf{\Lambda})$ und sind \mathbf{U} und \mathbf{V} voneinander unabhängig, dann läßt sich

die Verteilungsfunktion von $\Lambda_{p,n,m} = \det \mathbf{U} / \det(\mathbf{U} + \mathbf{V})$ auf der Grundlage einer asymptotischen Entwicklung annähern durch (SUGIURA und FUJIKOSHI 1969)

$$\begin{aligned} P(\Lambda_{p,n,m} < \Lambda_{o,p,n,m}) &= 1 - F(\chi'^2; f, \lambda) \\ &- (2o)^{-1}[(n+p+1)\text{sp}\mathbf{\Lambda}F(\chi'^2; f+2, \lambda) \\ &- ((n+p+1)\text{sp}\mathbf{\Lambda} - 2\text{sp}\mathbf{\Lambda}^2)F(\chi'^2; f+4, \lambda) \\ &- 2\text{sp}\mathbf{\Lambda}^2F(\chi'^2; f+6, \lambda)] \end{aligned} \quad (286.10)$$

mit

$$o = m + (n - p - 1)/2, \quad f = np, \quad \lambda = \text{sp}\mathbf{\Lambda}, \quad \chi'^2 = -o \ln \Lambda_{o,p,n,m},$$

worin $F(\chi'^2; f, \lambda)$ den Wert der Verteilungsfunktion (262.3) der nichtzentralen χ^2 -Verteilung an der Stelle χ'^2 bedeutet. In (SUGIURA und FUJIKOSHI 1969) sind zusätzlich die Glieder mit o^{-2} angegeben. Die exakte Verteilung von $\Lambda_{p,n,m}$ befindet sich in (NAGARSENKER 1977).

287 Verteilung spezieller Funktionen von Wishart-Matrizen

Bei der multivariaten Analyse treten in (425.15) und (426.2) noch zwei Testgrößen auf, deren Verteilungen im folgenden angegeben werden.

Die $p \times p$ Matrizen \mathbf{U} mit $\mathbf{U} \sim W(m, \mathbf{\Sigma})$ und \mathbf{V} mit $\mathbf{V} \sim W(n, \mathbf{\Sigma})$ seien voneinander unabhängig. Gesucht ist die Verteilung der Testgröße

$$T_{p,n,m}^2 = \text{sp}(\mathbf{V}\mathbf{U}^{-1}). \quad (287.1)$$

Die exakte Verteilung für $T_{p,n,m}^2$ ist bei (PILLAI und YOUNG 1971) angegeben. Zur numerischen Berechnung von Werten der Verteilungsfunktion eignen sich genäherte Verteilungen, insbesondere mit Hilfe der F -Verteilung, beispielsweise (LÄUTER 1974)

$$\frac{m-p-1}{np} \frac{g_2}{g_2-2} T_{p,n,m}^2 \sim F(g_1, g_2) \quad \text{für } m \geq p+2 \quad (287.2)$$

mit

$$g_1 = \frac{np(m-p)}{n+m-np-1}, \quad g_2 = m-p+1,$$

falls $n+m-np-1 > 0$, und in den übrigen Fällen

$$g_1 = \infty, \quad g_2 = m-p+1 - \frac{(m-p-1)(m-p-3)(n+m-np-1)}{(m-1)(n+m-p-1)},$$

so daß die Verteilungsfunktion von $T_{p,n,m}^2$ nach (263.3) oder (263.4) berechenbar ist, und zwar gilt den Überlegungen entsprechend, die zu (286.8) führen,

$$P(T_{p,n,m}^2 < T_{o,p,n,m}^2) = F(F_0; g_1, g_2) \quad (287.3)$$

mit

$$F_0 = \frac{m-p-1}{np} \frac{g_2}{g_2-2} T_{o,p,n,m}^2,$$

falls $T_{o,p,n,m}^2$ den Wert bedeutet, für den die Verteilungsfunktion zu ermitteln ist. Entsprechend (286.9) erhält man das α -Fraktil $T_{\alpha;p,n,m}^2$ zu

$$T_{\alpha;p,n,m}^2 = \frac{np}{m-p-1} \frac{g_2-2}{g_2} F_{\alpha;g_1,g_2}. \quad (287.4)$$

Tafeln für $T_{\alpha;p,n,m}^2$ befinden sich in (KRES 1983, S.118).

Besitzt die Matrix \mathbf{V} in (287.1) die nichtzentrale Wishart-Verteilung, wird eine asymptotische Entwicklung der Verteilungsfunktion von $T_{p,n,m}^2$ in (MUIR-HEAD 1972) angegeben.

Es sei $\mathbf{\Sigma}_0$ eine positiv definite $p \times p$ Matrix von Konstanten und $\hat{\mathbf{\Sigma}}$ eine $p \times p$ Matrix mit $m\hat{\mathbf{\Sigma}} \sim W(m, \mathbf{\Sigma})$. Unter der Bedingung, daß $\mathbf{\Sigma} = \mathbf{\Sigma}_0$ gilt, ist dann die Testgröße

$$\lambda_{p,m} = m[\ln(\det \mathbf{\Sigma}_0 / \det \hat{\mathbf{\Sigma}}) - p + \text{sp}(\hat{\mathbf{\Sigma}}\mathbf{\Sigma}_0^{-1})] \quad (287.5)$$

mit $m \rightarrow \infty$ verteilt wie $\chi^2(\frac{1}{2}p(p+1))$ (ANDERSON 1958, S.267). Eine Näherungsverteilung für $\lambda_{p,m}$ läßt sich mit Hilfe der χ^2 -Verteilung und eine etwas günstigere mit der F -Verteilung angeben (KORIN 1968). Für letztere gilt

$$\lambda_{p,m}/b \sim F(q_1, q_2) \quad \text{für } p > 1 \quad (287.6)$$

mit

$$q_1 = p(p+1)/2, \quad q_2 = (q_1+2)/(D_2 - D_1^2), \quad b = q_1/(1 - D_1 - q_1/q_2), \\ D_1 = (2p+1 - 2/(p+1))/6m, \quad D_2 = (p-1)(p+2)/6m^2,$$

so daß die Verteilungsfunktion von $\lambda_{p,m}$ und das α -Fraktil $\lambda_{\alpha;p,m}$ (287.3) und (287.4) entsprechend sich berechnen lassen. Tafeln, aus denen $\lambda_{\alpha;p,m}$ zu entnehmen ist, befinden sich in (KRES 1983, S.263).

3 Parameterschätzung in linearen Modellen

Die linearen Modelle für die Parameterschätzung baut man derart auf, daß die Erwartungswerte der Beobachtungen, die für die Parameterschätzung vorgenommen werden und die Zufallsvariable darstellen, als lineare Funktionen der unbekannt Parameter ausgedrückt werden. Die Koeffizienten der linearen Funktionen setzt man als bekannt voraus, so daß die Parameterschätzung in linearen Modellen im wesentlichen eine Schätzung der Erwartungswerte der Beobachtungen bedeutet.

Die Abhängigkeit der Beobachtungen von den Parametern ergibt sich im allgemeinen aus physikalischen oder mathematischen Gesetzmäßigkeiten oder geometrischen Zusammenhängen und die lineare Beziehung aus einer Linearisierung. Beispielsweise schätzt man die Koordinaten eines Punktes in einer Ebene aus den Entfernungsmessungen zu zwei oder mehr Punkten in der Ebene mit bekannten Koordinaten. Im Gegensatz zu diesen Modellen, die durch quantitative Aussagen entstehen und in denen die Parameterschätzung als Regressionsanalyse bezeichnet wird, kann man bei vielen Experimenten nur qualitative Annahmen über den linearen Zusammenhang zwischen den unbekannt Parametern und den Beobachtungen treffen. Sämtliche Koeffizienten der linearen Funktionen erhalten dann die Werte Null oder Eins. Parameterschätzungen in solchen Modellen bezeichnet man als Varianzanalyse, die beispielsweise vorliegt, wenn Preise von Waren in Abhängigkeit sich addierender Einflußfaktoren wie durchschnittliches Preisniveau, Geschäftslage und Zeitpunkt des Angebots analysiert werden.

Während die Koeffizienten der linearen Funktionen der Parameter im folgenden stets als feste Größen angesehen werden – Modelle mit Koeffizienten als Zufallsvariable befinden sich bei (SCHACH und SCHÄFER 1978, S.152; TOUTENBURG 1975, S.141) – werden die unbekannt Parameter sowohl als feste Größen als auch als Zufallsvariable definiert. Schätzungen in Modellen für den letztgenannten Fall führen auf die Vorhersage und Filterung von Beobachtungen und Signalen.

Bei der Einführung der Modelle wird zunächst angenommen, daß nur ein Merkmal eines Experimentes beobachtet wird, beispielsweise der Ertrag an Weizenkörnern bei einem Versuch, die Auswirkungen verschiedener Dünge-

mittel auf eine bestimmte Weizensorte zu untersuchen. Als weitere Merkmale könnten die Größe der Weizenkörner und ihr Proteingehalt gemessen worden sein. Die Analyse mehrerer Merkmale bei der Parameterschätzung wird in den multivariaten Modellen behandelt. Diese Modelle erlauben auch die Schätzung von Kovarianzmatrizen. Wie gezeigt wird, läßt sich diese Schätzung ebenso als Varianz- und Kovarianzkomponentenschätzung herleiten.

31 Methoden der Parameterschätzung

311 Punktschätzung

Die Beobachtungen, die die Information über die unbekannt Parameter enthalten, seien in dem Vektor \mathbf{y} zusammengefaßt. Er ist ein Zufallsvektor und enthält Werte der stetigen Zufallsvariablen, die nach (225.1) für den Wahrscheinlichkeitsraum (S, Z, P) des Experiments definiert ist, aus dem die Beobachtungen resultieren. Da bei stetigen Zufallsvariablen nicht sämtliche Werte, für die die Zufallsvariablen definiert sind, auch beobachtet werden können, stellt jede Messung \mathbf{y} eine *Stichprobe* aus der Grundmenge der Werte des Wahrscheinlichkeitsraums dar. Diese Grundmenge heißt auch *Grundgesamtheit*. Zu jeder Stichprobe gehört eine Verteilung, und die für sie definierten Parameter wie die Erwartungswerte oder Funktionen der Erwartungswerte bilden die *unbekannten Parameter* der Stichprobe, die in dem Vektor $\boldsymbol{\beta}$ zusammengefaßt seien. Die Parametervektoren $\boldsymbol{\beta}$ beliebiger Stichproben spannen nach (122.4) einen Vektorraum auf, der als *Parameterraum* B bezeichnet wird, so daß $\boldsymbol{\beta} \in B$ gilt.

Die unbekannt Parameter $\boldsymbol{\beta}$ werden mittels der Schätzfunktion $s(\mathbf{y})$ als Funktion der Beobachtungen \mathbf{y} geschätzt. Man nennt dies eine *Punktschätzung*. Wird nach einem Bereich gesucht, in dem die Parameter $\boldsymbol{\beta}$ bei vorgegebener Wahrscheinlichkeit liegen, handelt es sich um eine *Bereichsschätzung*, die im Kapitel 43 behandelt wird.

Besitzt die Schätzfunktion $s(\mathbf{y})$ alle Informationen über $\boldsymbol{\beta}$, die in den Beobachtungen \mathbf{y} enthalten sind, spricht man von einer *suffizienten* oder *erschöpfenden* Schätzfunktion (MÜLLER 1975, S.201; RAO 1973, S.130). Eine Schätzfunktion heißt *konsistent*, wenn mit der Wahrscheinlichkeit von Eins die Folge der Schätzungen bei unbegrenzt anwachsendem Stichprobenumfang sich den zu schätzenden Parametern beliebig nähert (MOOD et al. 1974, S.295). Eine Schätzfunktion bezeichnet man als *robust*, wenn ihre Wahrscheinlichkeitsverteilung gegenüber kleinen Änderungen der Verteilung der Stichprobe unempfindlich ist (DOKSUM 1976, S.246). Auf die Eigenschaften der erwartungstreuen und besten Schätzfunktion, die man auch *wirksamste* oder *effiziente* Schätzfunktion nennt, wird im folgenden ausführlich im Zusammenhang mit den drei

gebräuchlichsten Schätzverfahren, der besten erwartungstreuen Schätzung, der Methode der kleinsten Quadrate und der Maximum-Likelihood-Methode eingegangen. Ebenfalls werden robuste Parameterschätzungen behandelt.

312 Beste erwartungstreue Schätzung

Es sei $h(\boldsymbol{\beta})$ eine zu schätzende Funktion der unbekannt Parameter $\boldsymbol{\beta}$ und die Funktion $s(\mathbf{y})$ der Beobachtungen \mathbf{y} ihre Schätzung. Die Schätzung $s(\mathbf{y})$ sollte der zu schätzenden Größe $h(\boldsymbol{\beta})$ möglichst nahe kommen, das heißt, der Fehler der Schätzung, der sich in der Differenz $s(\mathbf{y}) - h(\boldsymbol{\beta})$ ausdrückt, sollte klein sein. Statt aber den Fehler selbst zu minimieren, ist es zweckmäßiger, das Quadrat $[s(\mathbf{y}) - h(\boldsymbol{\beta})]^2$ des Fehlers klein zu halten. Da \mathbf{y} ein Zufallsvektor ist, wird das Fehlerquadrat für einige Werte von \mathbf{y} klein und für andere Werte groß sein, so daß am zweckmäßigsten der zu erwartende quadratische Fehler $E([s(\mathbf{y}) - h(\boldsymbol{\beta})]^2)$ der Schätzung minimiert wird, was auf eine *beste* Schätzung führt.

Der Erwartungswert $E(s(\mathbf{y}) - h(\boldsymbol{\beta}))$ der Abweichung der Schätzung von der zu schätzenden Größe bezeichnet man als *Verzerrung*. Sie sollte gleich Null sein, also $E(s(\mathbf{y}) - h(\boldsymbol{\beta})) = 0$, so daß mit (231.5) $E(s(\mathbf{y})) = E(h(\boldsymbol{\beta}))$ folgt, denn der für die Schätzung zu erwartende Wert sollte nicht vom Erwartungswert der zu schätzenden Größe abweichen. Stimmen beide Erwartungswerte überein, spricht man von *unverzerrter* oder *erwartungstreuer* Schätzung. Folglich ist für eine *beste erwartungstreue* Schätzung

$$E(s(\mathbf{y}) - h(\boldsymbol{\beta})) = 0 \quad \text{und} \quad E([s(\mathbf{y}) - h(\boldsymbol{\beta})]^2) \quad \text{minimal} . \quad (312.1)$$

Enthält der Parametervektor $\boldsymbol{\beta}$ feste Größen im Gegensatz zu Zufallsvariablen, so ist $h(\boldsymbol{\beta})$ eine Konstante und $E(h(\boldsymbol{\beta})) = h(\boldsymbol{\beta})$ wegen (231.5). Man erhält dann für den Erwartungswert des quadratischen Fehlers der Schätzung $E([s(\mathbf{y}) - h(\boldsymbol{\beta})]^2) = E([s(\mathbf{y}) - E(s(\mathbf{y})) + E(s(\mathbf{y})) - h(\boldsymbol{\beta})]^2)$ und mit (232.5)

$$E([s(\mathbf{y}) - h(\boldsymbol{\beta})]^2) = V(s(\mathbf{y})) + [E(s(\mathbf{y})) - h(\boldsymbol{\beta})]^2 ,$$

denn es gilt, da $E(s(\mathbf{y}))$ und $h(\boldsymbol{\beta})$ Konstanten sind, $E\{[s(\mathbf{y}) - E(s(\mathbf{y}))][E(s(\mathbf{y})) - h(\boldsymbol{\beta})]\} = [E(s(\mathbf{y})) - E(s(\mathbf{y}))][E(s(\mathbf{y})) - h(\boldsymbol{\beta})] = 0$. Für eine erwartungstreue Schätzung folgt $E(s(\mathbf{y})) - h(\boldsymbol{\beta}) = 0$ und damit $E([s(\mathbf{y}) - h(\boldsymbol{\beta})]^2) = V(s(\mathbf{y}))$, so daß, um eine beste erwartungstreue Schätzung zu erhalten, die Varianz $V(s(\mathbf{y}))$ der Schätzung zu minimieren ist. Existiert eine solche erwartungstreue Schätzung mit minimaler Varianz für alle Parameter $\boldsymbol{\beta}$, dann nennt man sie gleichmäßig beste erwartungstreue Schätzung.

Definition: Die Funktion $s(\mathbf{y})$ der Beobachtungen \mathbf{y} sei Schätzung der Funktion $h(\boldsymbol{\beta})$ der unbekannt, festen Parameter $\boldsymbol{\beta}$, dann bezeichnet man $s(\mathbf{y})$ als gleichmäßig *beste erwartungstreue* Schätzung, falls für alle $\boldsymbol{\beta}$ gilt

- 1) $E(s(\mathbf{y})) = h(\boldsymbol{\beta})$, das heißt, $s(\mathbf{y})$ ist eine erwartungstreue Schätzung von $h(\boldsymbol{\beta})$,
- 2) $V(s(\mathbf{y})) \leq V(s^*(\mathbf{y}))$,

wobei $s^*(\mathbf{y})$ eine beliebige erwartungstreue Schätzung von $h(\boldsymbol{\beta})$ bedeutet, so daß $s(\mathbf{y})$ minimale Varianz besitzt. (312.2)

Im folgenden sollen mit Hilfe der Beobachtungen \mathbf{y} nicht nur feste Parameter $\boldsymbol{\beta}$, sondern auch Zufallsparameter $\boldsymbol{\gamma}$ der Stichprobe \mathbf{y} geschätzt werden. Ein solches Problem liegt zum Beispiel vor, wenn Beobachtungen vorhergesagt werden sollen. Ist $h(\boldsymbol{\gamma})$ die zu schätzende Funktion der unbekanntnen Zufallsvariablen $\boldsymbol{\gamma}$ und $s(\mathbf{y})$ ihre Schätzung, ergibt sich mit $s(\mathbf{y}) - h(\boldsymbol{\gamma}) = s(\mathbf{y}) - h(\boldsymbol{\gamma}) - E(s(\mathbf{y}) - h(\boldsymbol{\gamma})) + E(s(\mathbf{y})) - E(h(\boldsymbol{\gamma}))$, da auch $\boldsymbol{\gamma}$ ein Zufallsvektor ist, für den Erwartungswert des Fehlerquadrats der Schätzung aus (312.1) wegen (232.5)

$$E([s(\mathbf{y}) - h(\boldsymbol{\gamma})]^2) = V(s(\mathbf{y}) - h(\boldsymbol{\gamma})) + [E(s(\mathbf{y})) - E(h(\boldsymbol{\gamma}))]^2,$$

denn es gilt $E\{[s(\mathbf{y}) - h(\boldsymbol{\gamma}) - E(s(\mathbf{y}) - h(\boldsymbol{\gamma}))][E(s(\mathbf{y})) - E(h(\boldsymbol{\gamma}))]\} = 0$. Bedeutet $s(\mathbf{y})$ eine erwartungstreue Schätzung von $h(\boldsymbol{\gamma})$, erhält man $E(s(\mathbf{y})) - E(h(\boldsymbol{\gamma})) = 0$ wegen $E(h(\boldsymbol{\gamma})) \neq h(\boldsymbol{\gamma})$ und für den Erwartungswert des quadratischen Fehlers $E([s(\mathbf{y}) - h(\boldsymbol{\gamma})]^2) = V(s(\mathbf{y}) - h(\boldsymbol{\gamma}))$. Entsprechend (312.2) ergibt sich nun die

Definition: Die Funktion $s(\mathbf{y})$ der Beobachtungen \mathbf{y} sei Schätzung der Funktion $h(\boldsymbol{\gamma})$ der unbekanntnen Zufallsparameter $\boldsymbol{\gamma}$, dann bezeichnet man $s(\mathbf{y})$ als gleichmäßig *beste erwartungstreue* Schätzung, falls für alle $\boldsymbol{\gamma}$ gilt

- 1) $E(s(\mathbf{y})) = E(h(\boldsymbol{\gamma}))$, das heißt, $s(\mathbf{y})$ ist eine erwartungstreue Schätzung von $h(\boldsymbol{\gamma})$,
- 2) $V(s(\mathbf{y}) - h(\boldsymbol{\gamma})) \leq V(s^*(\mathbf{y}) - h(\boldsymbol{\gamma}))$,

wobei $s^*(\mathbf{y})$ eine beliebige erwartungstreue Schätzung von $h(\boldsymbol{\gamma})$ bedeutet, so daß $s(\mathbf{y}) - h(\boldsymbol{\gamma})$ minimale Varianz besitzt. (312.3)

Während bei der besten erwartungstreuen Schätzung einer Funktion $h(\boldsymbol{\beta})$ fester Parameter $\boldsymbol{\beta}$ nach (312.2) die Varianz $V(s(\mathbf{y}))$ der Schätzung minimal wird, ergibt sich bei der besten erwartungstreuen Schätzung einer Funktion $h(\boldsymbol{\gamma})$ der Zufallsparameter $\boldsymbol{\gamma}$ minimale Varianz $V(s(\mathbf{y}) - h(\boldsymbol{\gamma}))$ für die Differenz zwischen Schätzung und zu schätzender Größe.

313 Methode der kleinsten Quadrate

Eine weitere sinnvolle Methode, unbekanntne Parameter zu schätzen, besteht darin, die Quadratsumme der Abweichungen der Beobachtungen \mathbf{y} von den

Schätzwerten $s[E(\mathbf{y})]$ ihrer Erwartungswerte $E(\mathbf{y})$ zu minimieren, die, wie sich aus dem Kapitel 32 ergibt, Funktionen der unbekannt Parameter darstellen. Die Quadratsumme $(\mathbf{y} - s[E(\mathbf{y})])'(\mathbf{y} - s[E(\mathbf{y})])$ soll also minimal werden. Mit der positiv definiten Kovarianzmatrix $D(\mathbf{y}) = \mathbf{\Sigma}$ der Beobachtungen läßt sich die Methode verallgemeinern, und man fordert, daß die quadratische Form $(\mathbf{y} - s[E(\mathbf{y})])'\mathbf{\Sigma}^{-1}(\mathbf{y} - s[E(\mathbf{y})])$ minimal wird, da in der inversen Kovarianzmatrix $\mathbf{\Sigma}^{-1}$ kleinen Varianzen große Elemente oder wegen (233.18) große Gewichte entsprechen.

Definition: Der Zufallsvektor \mathbf{y} der Beobachtungen besitze die positiv definite Kovarianzmatrix $D(\mathbf{y}) = \mathbf{\Sigma}$, und der Erwartungswertvektor $E(\mathbf{y})$ sei Funktion der unbekannt Parameter und $s[E(\mathbf{y})]$ eine Schätzung von $E(\mathbf{y})$, dann bezeichnet man als *Methode der kleinsten Quadrate* die Schätzung der Parameter, die die quadratische Form

$$(\mathbf{y} - s[E(\mathbf{y})])'\mathbf{\Sigma}^{-1}(\mathbf{y} - s[E(\mathbf{y})])$$

minimal werden läßt.

(313.1)

Auf der Methode der kleinsten Quadrate zur Parameterschätzung beruht die *Ausgleichsrechnung* (GOTTHARDT 1978; GRAFAREND und SCHAFFRIN 1993; GROSSMANN 1969; HÖPCKE 1980; LINNIK 1961; MIKHAIL und ACKERMANN 1976; REISSMANN 1976; WOLF 1968, 1975, 1979). Analogien zur Elastomechanik sind bei (LINKWITZ 1977) behandelt.

314 Maximum–Likelihood–Methode

Die beste erwartungstreue Schätzung und die Methode der kleinsten Quadrate benötigen keine Angaben über die spezielle Art der Verteilung des Beobachtungsvektors \mathbf{y} . Bei der Parameterschätzung nach der Maximum–Likelihood–Methode dagegen muß die Dichte der Beobachtungen in Abhängigkeit von den unbekannt Parametern gegeben sein.

Definition: Der Zufallsvektor \mathbf{y} der Beobachtungen besitze die von den unbekannt, festen Parametern $\boldsymbol{\beta}$ abhängige Dichte $f(\boldsymbol{\beta})$, dann ist die *Likelihoodfunktion* $L(\mathbf{y}; \boldsymbol{\beta})$ definiert durch $L(\mathbf{y}; \boldsymbol{\beta}) = f(\boldsymbol{\beta})$. (314.1)

In der Likelihoodfunktion $L(\mathbf{y}; \boldsymbol{\beta})$ sind also die festen Parameter $\boldsymbol{\beta}$ unbekannt. Ihre Schätzwerte werden derart bestimmt, daß $L(\mathbf{y}; \boldsymbol{\beta})$ für die Schätzwerte maximal wird, denn dann ergeben sich wegen (314.1) auch Maximalwerte für die Dichte der Beobachtungen \mathbf{y} .

Definition: Der Zufallsvektor \mathbf{y} der Beobachtungen besitze die Likelihoodfunktion $L(\mathbf{y}; \boldsymbol{\beta})$, dann bezeichnet man als *Maximum–Likelihood–Methode* die Schätzung $\hat{\boldsymbol{\beta}}$ der festen Parameter $\boldsymbol{\beta}$, die maximale Werte für $L(\mathbf{y}; \boldsymbol{\beta})$ liefert,

also

$$\bar{\boldsymbol{\beta}} = \arg \sup_{\boldsymbol{\beta}} L(\mathbf{y}; \boldsymbol{\beta}) . \quad (314.2)$$

32 Gauß–Markoff–Modell

321 Definition und Linearisierung

a) Definition

Zunächst soll die Parameterschätzung in dem folgenden Modell behandelt werden.

Definition: Es sei \mathbf{X} eine $n \times u$ Matrix gegebener Koeffizienten, $\boldsymbol{\beta}$ ein $u \times 1$ Vektor unbekannter, fester Parameter, \mathbf{y} ein $n \times 1$ Zufallsvektor von Beobachtungen und $D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}$ die $n \times n$ Kovarianzmatrix von \mathbf{y} , wobei die Gewichtsmatrix \mathbf{P} der Beobachtungen \mathbf{y} bekannt und der positive Faktor σ^2 unbekannt sei. Ferner besitze \mathbf{X} den vollen Spaltenrang $\text{rg}\mathbf{X} = u$ und die Gewichtsmatrix \mathbf{P} sei positiv definit. Dann bezeichnet man

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}$$

als *Gauß–Markoff–Modell* mit vollem Rang. (321.1)

Für dieses Modell wird also vorausgesetzt, daß die Erwartungswerte der Beobachtungen \mathbf{y} sich aus Linearkombinationen gegebener Koeffizienten und unbekannter Parameter darstellen lassen. Es liegt daher ein lineares Modell vor. Die lineare Abhängigkeit ergibt sich, häufig nach einer unter b) erläuterten Linearisierung, aufgrund physikalischer oder mathematischer Gesetzmäßigkeiten, also aufgrund *quantitativer* Aussagen. Man bezeichnet diese Abhängigkeit auch als *Regression* und die Schätzung im Modell (321.1) als *Regressionsanalyse*. Dieses Modell unterscheidet sich aber wesentlich von dem im Kapitel 351 zu behandelnden Regressionsmodell, in dem Zufallsparameter aus Linearkombinationen der Beobachtungen geschätzt werden.

Mit $\text{rg}\mathbf{X} = u$ besitzt das Modell vollen Rang, und es muß wegen (132.2) $n \geq u$ gelten. Man geht aber im allgemeinen davon aus, daß die Anzahl n der Beobachtungen größer ist als die Anzahl u der Unbekannten, um den Einfluß des zufälligen Charakters der Beobachtungen \mathbf{y} auf die Schätzwerte gering zu halten. Für $n > u$ ist das Gleichungssystem $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$ nach (154.2) in der Regel nicht konsistent, da nach (135.5) $\dim R(\mathbf{X}) = \text{rg}\mathbf{X} = u$ und im allgemeinen $\mathbf{y} \in E^n$ gilt, so daß $\mathbf{y} \notin R(\mathbf{X})$ folgt. Durch Addition des $n \times 1$ Zufallsvektors \mathbf{e} der Fehler von \mathbf{y} erhält man das konsistente System

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{y} + \mathbf{e} \quad \text{mit} \quad E(\mathbf{e}) = \mathbf{0} \quad \text{und} \quad D(\mathbf{e}) = D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1} , \quad (321.2)$$

denn mit $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ in (321.1) folgt $E(\mathbf{e}) = \mathbf{0}$ und mit $\mathbf{e} = -\mathbf{y} + \mathbf{X}\boldsymbol{\beta}$ aus (233.2) $D(\mathbf{e}) = \sigma^2 \mathbf{P}^{-1}$. Damit stellt (321.2) eine alternative Formulierung des Modells (321.1) dar. Die Gleichungen $\mathbf{X}\boldsymbol{\beta} = \mathbf{y} + \mathbf{e}$ nennt man auch *Beobachtungsgleichungen* und in der Ausgleichungsrechnung *Fehler-* oder *Verbesserungsgleichungen*.

Das Modell (321.1) oder (321.2) trägt den Namen Gauß–Markoff–Modell, denn GAUSS (1809, S.213) leitete in diesem Modell mit Hilfe der Likelihoodfunktion die Methode der kleinsten Quadrate ab und zeigte, daß sie auf eine beste Schätzung führt (GAUSS 1823, S.21). MARKOFF (1912, S.218) bestimmte die Parameter dieses Modells mit Hilfe der besten erwartungstreuen Schätzung. In der Ausgleichungsrechnung bezeichnet man die Schätzung im Gauß–Markoff–Modell als die *Ausgleichung nach vermittelnden Beobachtungen*.

Die Kovarianzmatrix $D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}$ der Beobachtungen \mathbf{y} setzt man in (321.1) und (321.2) bis auf den Faktor σ^2 als bekannt voraus. Die Matrix \mathbf{P} bezeichnet die Gewichtsmatrix der Beobachtungen \mathbf{y} , denn mit $D(\mathbf{y}) = \boldsymbol{\Sigma}$ und $c = \sigma^2$ folgt aus (233.18) $\sigma^2 \mathbf{P}^{-1} = \boldsymbol{\Sigma}$ und daher $D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}$. In der Ausgleichungsrechnung setzt man auch $\mathbf{P}^{-1} = \mathbf{Q}$, also $D(\mathbf{y}) = \sigma^2 \mathbf{Q}$ und bezeichnet \mathbf{Q} als Matrix der *Kofaktoren* oder *Gewichtskoeffizienten* und die Diagonalelemente von \mathbf{Q} als *Gewichtszeproke*. Da \mathbf{P} als positiv definit vorausgesetzt wird, existiert nach (143.3) die Inverse \mathbf{P}^{-1} und ist nach (143.9) positiv definit. Mit $\mathbf{P} = \mathbf{I}$ folgt aus (321.1) $D(\mathbf{y}) = \sigma^2 \mathbf{I}$, so daß der Faktor σ^2 als *Varianz der Gewichtseinheit* bezeichnet wird.

Ein lineares Modell wie in (321.1) sei durch

$$\bar{\mathbf{X}}\boldsymbol{\beta} = E(\bar{\mathbf{y}}) \quad \text{mit} \quad D(\bar{\mathbf{y}}) = \sigma^2 \mathbf{P}^{-1} \quad (321.3)$$

oder wie in (321.2) durch $\bar{\mathbf{X}}\boldsymbol{\beta} = \bar{\mathbf{y}} + \bar{\mathbf{e}}$ mit $D(\bar{\mathbf{e}}) = D(\bar{\mathbf{y}})$ gegeben. Da \mathbf{P} positiv definit ist, gilt nach (143.5) die Cholesky–Faktorisierung $\mathbf{P} = \mathbf{G}\mathbf{G}'$, in der \mathbf{G} eine reguläre untere Dreiecksmatrix bedeutet. Mit

$$\mathbf{P} = \mathbf{G}\mathbf{G}' \quad , \quad \mathbf{X} = \mathbf{G}'\bar{\mathbf{X}} \quad , \quad \mathbf{y} = \mathbf{G}'\bar{\mathbf{y}} \quad \text{und} \quad \mathbf{e} = \mathbf{G}'\bar{\mathbf{e}} \quad , \quad (321.4)$$

wobei wegen (132.4) $\text{rg}\bar{\mathbf{X}} = \text{rg}\mathbf{X} = u$ gilt, folgt dann

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I} \quad (321.5)$$

oder $\mathbf{X}\boldsymbol{\beta} = \mathbf{y} + \mathbf{e}$ mit $D(\mathbf{e}) = \sigma^2 \mathbf{I}$, denn mit (131.14) und (233.2) erhält man $D(\mathbf{y}) = \sigma^2 \mathbf{G}'((\mathbf{G}')^{-1}\mathbf{G}^{-1})\mathbf{G} = \sigma^2 \mathbf{I}$. Das Modell (321.1) oder (321.2) läßt sich also in das einfachere Modell (321.5) transformieren, in dem die Beobachtungen unkorreliert sind und gleiche Varianzen besitzen, was als *Homoskedastizität* bezeichnet wird. Zur einfacheren Ableitung der Schätzwerte

wird daher im folgenden das Modell (321.5) anstelle von (321.1) verwendet. Die Schätzwerte für (321.1) lassen sich dann aus denen für (321.5) durch die Transformationen (321.4) gewinnen. Sie stimmen mit denen überein, die durch Ableitung der Schätzung im Modell (321.1) erhalten werden.

1. Beispiel: Für voneinander unabhängige Beobachtungen ist nach (233.19) die Gewichtsmatrix \mathbf{P} durch die Diagonalmatrix $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_n)$ definiert. Dann ergibt sich die untere Dreiecksmatrix \mathbf{G} der Cholesky-Faktorisierung der Gewichtsmatrix $\mathbf{P} = \mathbf{G}\mathbf{G}'$ zu

$$\mathbf{G} = \text{diag}(\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_n}).$$

Die Transformation der Beobachtungsgleichungen $\bar{\mathbf{X}}\boldsymbol{\beta} = \bar{\mathbf{y}} + \bar{\mathbf{e}}$ in $\mathbf{G}'\bar{\mathbf{X}}\boldsymbol{\beta} = \mathbf{G}'\bar{\mathbf{y}} + \mathbf{G}'\bar{\mathbf{e}}$ bewirkt, daß jede Beobachtungsgleichung mit der Wurzel des Gewichts ihrer Beobachtung multipliziert wird, was man in der Ausgleichsrechnung als *Homogenisierung* bezeichnet.

2. Beispiel: Um einen Kurvenverlauf in einer Ebene zu erfassen, werden für n gegebene Abszissenwerte x_i die Ordinaten y_i der Kurve gemessen, wobei die Beobachtungen unkorreliert seien und gleiche Varianzen besitzen sollen. Für die Darstellung des durch die Meßwerte repräsentierten Kurvenverlaufs genüge ein Polynom 2. Grades mit den drei unbekanntem Koeffizienten $\beta_0, \beta_1, \beta_2$. Man erhält dann das lineare Modell

$$\beta_0 + x_i\beta_1 + x_i^2\beta_2 = E(y_i) \quad \text{mit } i \in \{1, \dots, n\} \quad \text{und } D(\mathbf{y}) = \sigma^2 \mathbf{I}$$

mit $\mathbf{y} = (y_i)$. Ausführlicher wird dieses sogenannte Polynommodell im Kapitel 341 behandelt.

b) Linearisierung

Im allgemeinen wird keine lineare Abhängigkeit zwischen den Parametern $\boldsymbol{\beta}$ und den Beobachtungen \mathbf{y} bestehen, wie sie in (321.1) vorausgesetzt wurde, vielmehr wird (321.2) entsprechend gelten

$$\begin{aligned} h_1(\beta_1, \dots, \beta_u) &= y_1^* + e_1 \\ h_2(\beta_1, \dots, \beta_u) &= y_2^* + e_2 \\ &\dots\dots\dots \\ h_n(\beta_1, \dots, \beta_u) &= y_n^* + e_n, \end{aligned} \tag{321.6}$$

worin $h_i(\beta_1, \dots, \beta_u)$ reelle differenzierbare Funktionen der unbekanntem Parameter β_1, \dots, β_u , y_i^* die Beobachtungen und e_i die Fehler seien. Sind mit $\beta_1 = \beta_{10} + \Delta\beta_1, \dots, \beta_u = \beta_{u0} + \Delta\beta_u$ Näherungswerte β_{j0} für die Parameter gegeben und die Korrekturen $\Delta\beta_j$ unbekannt, läßt sich (321.6) mit der

Taylor–Entwicklung (171.3) linearisieren. Man erhält mit $\beta_0 = (\beta_{j0})$

$$\begin{aligned} h_i(\beta_1, \dots, \beta_u) &= h_i(\beta_{10} + \Delta\beta_1, \dots, \beta_{u0} + \Delta\beta_u) \\ &= h_i(\beta_{10}, \dots, \beta_{u0}) + \frac{\partial h_i}{\partial \beta_1} \Big|_{\beta_0} \Delta\beta_1 + \dots + \frac{\partial h_i}{\partial \beta_u} \Big|_{\beta_0} \Delta\beta_u. \end{aligned} \quad (321.7)$$

Mit

$$\begin{aligned} \mathbf{y} &= |y_1^* - h_1(\beta_{10}, \dots, \beta_{u0}), \dots, y_n^* - h_n(\beta_{10}, \dots, \beta_{u0})|', \\ \beta &= |\Delta\beta_1, \dots, \Delta\beta_u|' \end{aligned}$$

und

$$\mathbf{X} = \begin{vmatrix} \frac{\partial h_1}{\partial \beta_1} \Big|_{\beta_0} & \dots & \frac{\partial h_1}{\partial \beta_u} \Big|_{\beta_0} \\ \dots & \dots & \dots \\ \frac{\partial h_n}{\partial \beta_1} \Big|_{\beta_0} & \dots & \frac{\partial h_n}{\partial \beta_u} \Big|_{\beta_0} \end{vmatrix} \quad (321.8)$$

ergibt sich dann anstelle von (321.6) das Modell (321.1), (321.2) oder (321.5). Werden die Schätzwerte der Parameter dazu benutzt, um iterativ neue Näherungswerte zu berechnen, da die zuerst benutzten Näherungswerte nicht genau genug waren, um eine Linearisierung nach (321.7) zu ermöglichen, müssen die genäherten Beobachtungen und die Differentialquotienten in (321.8) jeweils mit den neuen Näherungswerten gebildet werden. Es wird dann unter der Voraussetzung, daß das Iterationsverfahren konvergiert, solange iteriert, bis die Elemente des Vektors β genügend klein sind. Dieses Verfahren ist für praktische Rechnungen geeignet (SCHEK und MAIER 1976), aber es muß eventuell numerisch geprüft werden, ob mit unterschiedlichen Näherungswerten identische Ergebnisse für die unbekannt Parameter erhalten werden. Weitere Möglichkeiten, nichtlineare Modelle zu behandeln, sind bei (DRAPER und SMITH 1966, S.267; SPÄTH 1974, S.92; TEUNISSEN 1985) angegeben. Ein Beispiel für eine Linearisierung befindet sich im Kapitel 326.

322 Beste lineare erwartungstreue Schätzung

In dem linearen Modell (321.5) soll die lineare Funktion $\mathbf{a}'\beta$ der unbekannt Parameter β durch die lineare Funktion $\mathbf{c}'\mathbf{y}$ der Beobachtungen \mathbf{y} geschätzt werden, so daß eine *lineare Schätzung* vorliegt, wobei \mathbf{a} ein gegebener $u \times 1$ Vektor und \mathbf{c} ein zu bestimmender $n \times 1$ Vektor bedeuten. Um eine beste lineare erwartungstreue Schätzung zu erhalten, muß nach (312.2) für alle β

$$1) \quad E(\mathbf{c}'\mathbf{y}) = \mathbf{a}'\beta \quad (322.1)$$

oder mit (321.5) $\mathbf{c}'\mathbf{X}\beta = \mathbf{a}'\beta$ erfüllt sein. Somit folgt

$$\mathbf{c}'\mathbf{X} = \mathbf{a}' \quad (322.2)$$

Ferner muß

$$2) \quad V(\mathbf{c}'\mathbf{y})$$

minimal werden. Mit (233.2) und (321.5) erhält man

$$V(\mathbf{c}'\mathbf{y}) = \sigma^2 \mathbf{c}'\mathbf{c}, \quad (322.3)$$

so daß (322.3) in Abhängigkeit von \mathbf{c} unter den Restriktionen (322.2) zu minimieren ist.

Zur Lösung dieses Problems wird nach (171.6) die Lagrangesche Funktion $w(\mathbf{c})$ gebildet, deren Ableitungen nach \mathbf{c} gleich Null zu setzen sind. Man erhält $w(\mathbf{c}) = \sigma^2 \mathbf{c}'\mathbf{c} - 2\mathbf{k}'(\mathbf{X}'\mathbf{c} - \mathbf{a})$, worin $-2\mathbf{k}$ den $u \times 1$ Vektor der Lagrangeschen Multiplikatoren bezeichnet. Aus $\partial w(\mathbf{c})/\partial \mathbf{c} = 2\sigma^2 \mathbf{c} - 2\mathbf{X}\mathbf{k} = \mathbf{0}$ wegen (172.1) und (172.2) folgt $\mathbf{c} = \mathbf{X}\mathbf{k}/\sigma^2$ und mit (322.2) $\mathbf{a}' = \mathbf{k}'\mathbf{X}'\mathbf{X}/\sigma^2$ oder $\mathbf{k}'/\sigma^2 = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}$ und schließlich

$$\mathbf{c}' = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' . \quad (322.4)$$

Die Lösung für \mathbf{c} existiert und ist eindeutig, denn die Matrix $(\mathbf{X}'\mathbf{X})^{-1}$ existiert wegen (133.1), da nach (135.6) $\text{rg}\mathbf{X} = \text{rg}(\mathbf{X}'\mathbf{X}) = u$ gilt, und sie ist eindeutig wegen (131.13).

Mit (322.4) folgt aus (322.3)

$$V(\mathbf{c}'\mathbf{y}) = \sigma^2 \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} . \quad (322.5)$$

Es muß jetzt gezeigt werden, daß diese Varianz minimal ist, daß sie also die zweite Bedingung in (312.2) $V(\mathbf{c}'\mathbf{y}) \leq V(\mathbf{c}^{*\prime}\mathbf{y})$ erfüllt, wobei $\mathbf{c}^{*\prime}\mathbf{y}$ eine beliebige erwartungstreue Schätzung von $\mathbf{a}'\boldsymbol{\beta}$ bedeutet. Aus der Erwartungstreue folgt mit (322.2) $\mathbf{c}^{*\prime}\mathbf{X} = \mathbf{a}'$. Weiter erhält man mit (233.13)

$$V(\mathbf{c}^{*\prime}\mathbf{y}) = V(\mathbf{c}^{*\prime}\mathbf{y} - \mathbf{c}'\mathbf{y} + \mathbf{c}'\mathbf{y}) = V(\mathbf{c}^{*\prime}\mathbf{y} - \mathbf{c}'\mathbf{y}) + V(\mathbf{c}'\mathbf{y}) + 2f \quad (322.6)$$

und mit (233.15)

$$f = (\mathbf{c}^* - \mathbf{c})'D(\mathbf{y})\mathbf{c} . \quad (322.7)$$

Substituiert man (322.4), folgt mit $\mathbf{c}^{*\prime}\mathbf{X} = \mathbf{a}'$ schließlich

$$f = \sigma^2 \mathbf{c}^{*\prime}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} - \sigma^2 \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} = 0 ,$$

so daß $V(\mathbf{c}'\mathbf{y}) \leq V(\mathbf{c}^{*\prime}\mathbf{y})$ sich ergibt, da für die Varianz $V(\mathbf{c}^{*\prime}\mathbf{y} - \mathbf{c}'\mathbf{y}) \geq 0$ gilt.

Bezeichnet man mit $\widehat{\mathbf{a}'\boldsymbol{\beta}}$ die beste lineare erwartungstreue Schätzung von $\mathbf{a}'\boldsymbol{\beta}$, gilt $\widehat{\mathbf{a}'\boldsymbol{\beta}} = \mathbf{c}'\mathbf{y}$ und somit $\widehat{\mathbf{a}'\boldsymbol{\beta}} = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Mit $\mathbf{a}' = |1, 0, 0, \dots|$ und $\boldsymbol{\beta} = (\beta_i)$ erhält man aus $\widehat{\mathbf{a}'\boldsymbol{\beta}}$ den Schätzwert $\hat{\beta}_1$ des Parameters β_1 , mit $\mathbf{a}' = |0, 1, 0, \dots|$ den Schätzwert $\hat{\beta}_2$ von β_2 und so fort, so daß mit $\hat{\boldsymbol{\beta}} = (\hat{\beta}_i)$ sich $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ ergibt. Dann folgt $\widehat{\mathbf{a}'\boldsymbol{\beta}} = \mathbf{a}'\hat{\boldsymbol{\beta}}$, und man erhält den

Satz: Die beste lineare erwartungstreue Schätzung $\widehat{\mathbf{a}'\boldsymbol{\beta}}$ und ihre Varianz $V(\widehat{\mathbf{a}'\boldsymbol{\beta}})$ der linearen Funktion $\mathbf{a}'\boldsymbol{\beta}$ der unbekannt Parameter $\boldsymbol{\beta}$ im Gauß–Markoff–Modell (321.5) ist gegeben durch

$$\widehat{\mathbf{a}'\boldsymbol{\beta}} = \mathbf{a}'\hat{\boldsymbol{\beta}} = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \text{und} \\ V(\widehat{\mathbf{a}'\boldsymbol{\beta}}) = V(\mathbf{a}'\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a} . \quad (322.8)$$

Mit $\mathbf{a}' = |1, 0, 0, \dots|$, $\mathbf{a}' = |0, 1, 0, \dots|$ und so fort ergibt sich aus $V(\widehat{\mathbf{a}'\boldsymbol{\beta}})$ für die Varianz $V(\hat{\beta}_1)$ das erste Diagonalelement von $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, für die Varianz $V(\hat{\beta}_2)$ das zweite Diagonalelement von $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ und so weiter. Für die gesamte Kovarianzmatrix $D(\hat{\boldsymbol{\beta}})$ der Schätzwerte $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ folgt aus (233.2) in Übereinstimmung mit diesen Ergebnissen $D(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'D(\mathbf{y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Da $V(\hat{\beta}_i)$ minimal ist, ist auch $\text{sp}D(\hat{\boldsymbol{\beta}})$ minimal. Mit (321.4) ergeben sich aus $\hat{\boldsymbol{\beta}}$ und $D(\hat{\boldsymbol{\beta}})$ die Schätzwerte und ihre Kovarianzmatrix im Modell (321.1), und man erhält den

Satz: Die beste lineare erwartungstreue Schätzung $\hat{\boldsymbol{\beta}}$ der unbekannt Parameter $\boldsymbol{\beta}$ und ihre Kovarianzmatrix $D(\hat{\boldsymbol{\beta}})$, die minimale Spur besitzt, sind im Gauß–Markoff–Modell (321.5) gegeben durch

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \text{und} \quad D(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

und im Gauß–Markoff–Modell (321.1) durch

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{y} \quad \text{und} \quad D(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1} . \quad (322.9)$$

323 Methode der kleinsten Quadrate

Zur Parameterschätzung im Gauß–Markoff–Modell (321.5) nach der Methode der kleinsten Quadrate sind mit $\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y})$ gemäß (313.1) die Schätzwerte für $\boldsymbol{\beta}$ derart zu bestimmen, daß die Quadratsumme

$$S(\boldsymbol{\beta}) = \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (323.1)$$

minimal wird. Nach (171.4) ergibt sich ein Extremwert mit

$$\partial S(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \partial\left(\frac{1}{\sigma^2}(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta})\right)/\partial\boldsymbol{\beta} = \mathbf{0} ,$$

woraus mit (172.1) und (172.2) die mit (322.9) identische Schätzung $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ folgt.

Mit $\hat{\beta}$ wird (323.1) minimal, denn für einen beliebigen $u \times 1$ Vektor β^* gilt

$$\begin{aligned} & (\mathbf{y} - \mathbf{X}\beta^*)'(\mathbf{y} - \mathbf{X}\beta^*) \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta} + \mathbf{X}(\hat{\beta} - \beta^*))'(\mathbf{y} - \mathbf{X}\hat{\beta} + \mathbf{X}(\hat{\beta} - \beta^*)) \\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \beta^*)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta^*) \\ &\geq (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}), \end{aligned} \quad (323.2)$$

da $(\hat{\beta} - \beta^*)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \beta^*) \geq 0$ gilt, weil $\mathbf{X}'\mathbf{X}$ wegen (143.8) positiv definit ist, und da $2(\hat{\beta} - \beta^*)'\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) = 2(\hat{\beta} - \beta^*)'(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{y}) = 0$ gilt. Für die Quadratsumme $(\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$ im Modell (321.5) folgt mit (321.4) die quadratische Form $(\mathbf{y} - \mathbf{X}\hat{\beta})'\mathbf{P}(\mathbf{y} - \mathbf{X}\hat{\beta})$ im Modell (321.1), die ebenfalls minimal ist. Man erhält daher den

Satz: Die beste lineare erwartungstreue Schätzung und die Methode der kleinsten Quadrate im Gauß–Markoff–Modell (321.1) oder (321.5) liefern identische Schätzwerte $\hat{\beta}$, die in (322.9) angegeben sind, für die Parameter β . (323.3)

Bezeichnet man mit dem $n \times 1$ Vektor $\hat{\mathbf{y}}$ die Schätzwerte der Erwartungswerte $E(\mathbf{y})$ der Beobachtungen, erhält man den $n \times 1$ Vektor $\hat{\mathbf{e}}$ der Schätzwerte des Vektors \mathbf{e} in (321.2) zu

$$\hat{\mathbf{e}} = \hat{\mathbf{y}} - \mathbf{y} \quad \text{mit} \quad \hat{\mathbf{y}} = \mathbf{X}\hat{\beta}. \quad (323.4)$$

Man bezeichnet $\hat{\mathbf{e}}$ als Vektor der *Residuen* und in der Ausgleichsrechnung als Vektor der *Verbesserungen*. Wegen (323.2) gilt der

Satz: Bei der besten linearen erwartungstreuen Schätzung und der Methode der kleinsten Quadrate ist im Gauß–Markoff–Modell (321.1) die quadratische Form $\hat{\mathbf{e}}'\mathbf{P}\hat{\mathbf{e}}$ der Residuen und im Modell (321.5) die Quadratsumme $\hat{\mathbf{e}}'\hat{\mathbf{e}}$ der Residuen minimal. (323.5)

Die Schätzung nach der Methode der kleinsten Quadrate läßt sich auch geometrisch interpretieren. Durch den nach (135.1) mit $R(\mathbf{X}) = \{z | z = \mathbf{X}\beta\}$ definierten Spaltenraum der Matrix \mathbf{X} ist wegen $\text{rg}\mathbf{X} = u$ nach (135.5) ein u -dimensionaler Euklidischer Raum E^u bestimmt, in dem $\mathbf{X}\beta$ zu schätzen ist. Dieser Raum ist in Abbildung 323-1 als Ebene dargestellt. Die Schätzung $\hat{\beta}$ im Modell (321.5) wird derart bestimmt, daß $\mathbf{X}\hat{\beta}$ sich als orthogonale Projektion des Beobachtungsvektors $\mathbf{y} \in E^n$ auf den Spaltenraum $R(\mathbf{X})$ ergibt.

Mit dem orthogonalen Projektionsoperator \mathbf{H} aus (162.3) erhält man mit (153.22)

$$\mathbf{H}\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad \text{mit} \quad \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \quad (323.6)$$

woraus durch linksseitige Multiplikation mit $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ die Schätzung $\hat{\boldsymbol{\beta}}$ in (322.9) folgt.

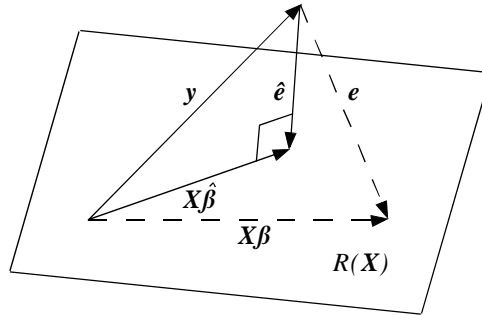


Abb. 323-1: Orthogonale Projektion

Durch die Projektion wird der Beobachtungsvektor \mathbf{y} nach (162.3) eindeutig zerlegt in $\mathbf{y} = \mathbf{H}\mathbf{y} + (\mathbf{I} - \mathbf{H})\mathbf{y}$, so daß wegen (323.4) und (323.6) $\hat{\mathbf{e}} = -(\mathbf{I} - \mathbf{H})\mathbf{y}$ oder

$$\hat{\mathbf{e}} = -(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \quad (323.7)$$

folgt. Weiter ist $\mathbf{X}\hat{\boldsymbol{\beta}} \in R(\mathbf{X})$ und $\hat{\mathbf{e}} \in R(\mathbf{X})^\perp$. Dann gilt nach (162.1) $\hat{\boldsymbol{\beta}}'\mathbf{X}'\hat{\mathbf{e}} = 0$, und nach Substitution von $\hat{\boldsymbol{\beta}}$, da im allgemeinen $\mathbf{y} \neq \mathbf{0}$ ist,

$$\mathbf{X}'\hat{\mathbf{e}} = \mathbf{0}. \quad (323.8)$$

Hieraus folgt mit (323.4) $\mathbf{X}'(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}) = \mathbf{0}$ oder die Gleichungen $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ für die Schätzwerte $\hat{\boldsymbol{\beta}}$, die wegen (323.8) auch als *Normalgleichungen* bezeichnet und im Kapitel 326 ausführlicher behandelt werden. Da $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ und $\hat{\mathbf{e}}$ zueinander orthogonal sind, erhält man aus (323.4) die folgende Zerlegung der Quadratsumme $\mathbf{y}'\mathbf{y}$ der Beobachtungen

$$\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \hat{\mathbf{e}}'\hat{\mathbf{e}}.$$

Die Quadratsumme $\hat{\mathbf{e}}'\hat{\mathbf{e}}$ der Residuen wird bei der Schätzung der Varianz σ^2 der Gewichtseinheit und bei den Hypothesentests eine besondere Rolle spielen.

Den Vektor \hat{e} der Residuen in (323.7) erhält man auch unmittelbar durch Substitution von $\hat{\beta}$ aus (322.9) in (323.4), und durch linksseitige Multiplikation von \mathbf{X}' folgt (323.8). Auf diese Weise oder durch Substitution von $\bar{e} = (\mathbf{G}')^{-1}\mathbf{e}$ aus (321.4) erhält man im Modell (321.1) anstelle von (323.7)

$$\hat{e} = -(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P})\mathbf{y} \quad (323.9)$$

und anstelle von (323.8)

$$\mathbf{X}'\mathbf{P}\hat{e} = \mathbf{0} . \quad (323.10)$$

Da die Matrix $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ gemäß (152.1) idempotent ist, ergibt sich mit (233.2) die Kovarianzmatrix $D(\hat{e})$ des Residuenvektors \hat{e} aus (323.7) zu

$$D(\hat{e}) = \sigma^2(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \quad (323.11)$$

und mit $\text{rg}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{sp}(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) = u$ wegen (137.3) und (152.3) sowie mit (152.4)

$$\text{rg}D(\hat{e}) = n - u \quad \text{und} \quad \text{sp}D(\hat{e}) = \sigma^2(n - u) , \quad (323.12)$$

so daß die Kovarianzmatrix $D(\hat{e})$ nach (133.1) singulär ist. Im Modell (321.1) erhält man mit $\bar{e} = (\mathbf{G}')^{-1}\mathbf{e}$ aus (321.4) sowie mit (233.2) und (323.11)

$$D(\hat{e}) = \sigma^2(\mathbf{P}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}') \quad \text{und} \quad \text{rg}D(\hat{e}) = n - u . \quad (323.13)$$

324 Maximum–Likelihood–Methode

Die beste lineare erwartungstreue Schätzung und die Methode der kleinsten Quadrate benötigen keine Angaben über die Art der Verteilung des Zufallsvektors \mathbf{y} der Beobachtungen, die aber für die Parameterschätzung nach der Maximum–Likelihood–Methode (314.2) erforderlich sind. Aufgrund des im Kapitel 241 erläuterten zentralen Grenzwertsatzes wird \mathbf{y} als normalverteilt angenommen. Mit $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ und $D(\mathbf{y}) = \sigma^2\mathbf{I}$ im Gauß–Markoff–Modell (321.5) folgt dann mit (251.1)

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \quad (324.1)$$

und mit (314.1) die Likelihoodfunktion $L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2) = (2\pi)^{-n/2}(\det \sigma^2\mathbf{I})^{-1/2} \exp(-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/2\sigma^2)$ oder mit (136.9)

$$L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp(-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/2\sigma^2) . \quad (324.2)$$

Nach (314.2) sind die unbekannt Parameter $\boldsymbol{\beta}$ und σ^2 derart zu bestimmen, daß $L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2)$ maximal wird.

Hierzu wird nach (171.4) $L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2)$ nach $\boldsymbol{\beta}$ und σ^2 differenziert und die Ableitungen gleich Null gesetzt. Zur Vereinfachung der Differentiation wird nicht $L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2)$, sondern $\ln L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2)$ abgeleitet, was zulässig ist, da die Likelihoodfunktion wie die Dichte der Normalverteilung positiv ist und mit $\partial \ln L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2) / \partial \boldsymbol{\beta} = \mathbf{0}$ folgt $(1/L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2)) \partial L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2) / \partial \boldsymbol{\beta} = \mathbf{0}$ und daher $\partial L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2) / \partial \boldsymbol{\beta} = \mathbf{0}$. Man erhält

$$\ln L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (324.3)$$

Die von $\boldsymbol{\beta}$ abhängige Quadratsumme in (324.3) ist bis auf den Faktor $1/2$ identisch mit (323.1). Mit $\partial \ln L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2) / \partial \boldsymbol{\beta} = \mathbf{0}$ ergeben sich also die Schätzwerte $\hat{\boldsymbol{\beta}}$ der Methode der kleinsten Quadrate. Mit (323.3) folgt daher der

Satz: Im Gauß–Markoff–Modell (321.1) oder (321.5) liefern die beste lineare erwartungstreue Schätzung, die Methode der kleinsten Quadrate und die Maximum–Likelihood–Methode im Falle normalverteilter Beobachtungen identische Schätzwerte $\hat{\boldsymbol{\beta}}$, die in (322.9) angegeben sind, für die Parameter $\boldsymbol{\beta}$. (324.4)

Aus

$$\frac{\partial \ln L(\mathbf{y}; \boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

folgt für die Schätzung $\bar{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit

$$\bar{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (324.5)$$

Im folgenden Kapitel wird gezeigt, daß $\bar{\sigma}^2$ keine erwartungstreue Schätzung von σ^2 darstellt. Um zu zeigen, daß mit $\hat{\boldsymbol{\beta}}$ und $\bar{\sigma}^2$ die Likelihoodfunktion (324.2) maximal wird, kann man wie bei (GRAYBILL 1976, S.343) vorgehen.

325 Erwartungstreue Schätzung der Varianz der Gewichtseinheit

Die Schätzung $\bar{\sigma}^2$ der unbekanntem Varianz σ^2 der Gewichtseinheit im Gauß–Markoff–Modell (321.5) ergibt sich nach (323.4) und (324.5) mit Hilfe der Quadratsumme $\hat{\mathbf{e}}' \hat{\mathbf{e}}$ der Residuen. Diese sei mit Ω bezeichnet, so daß man in den Modellen (321.5) und (321.1) erhält

$$\Omega = \hat{\mathbf{e}}' \hat{\mathbf{e}} \quad \text{und} \quad \Omega = \hat{\mathbf{e}}' \mathbf{P} \hat{\mathbf{e}}. \quad (325.1)$$

Der Erwartungswert von Ω soll berechnet werden, um daraus eine erwartungstreue Schätzung von σ^2 abzuleiten.

Man erhält Ω mit (323.7) für das Modell (321.5) und mit (323.9) für das Modell (321.1)

$$\begin{aligned}\Omega &= \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \quad \text{und} \\ \Omega &= \mathbf{y}'(\mathbf{P} - \mathbf{P}\mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P})\mathbf{y} .\end{aligned}\quad (325.2)$$

Mit (152.8) folgt aus dem ersten Ausdruck

$$\Omega \geq 0 . \quad (325.3)$$

Nach Substitution von $\hat{\beta}$ aus (322.9) in (325.2) ergeben sich für die beiden Modelle (321.5) und (321.1) die Beziehungen

$$\Omega = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\hat{\beta} \quad \text{und} \quad \Omega = \mathbf{y}'\mathbf{P}\mathbf{y} - \mathbf{y}'\mathbf{P}\mathbf{X}\hat{\beta} , \quad (325.4)$$

die im folgenden Kapitel benötigt werden. Mit (271.1), (321.5) und (323.12) berechnet sich $E(\Omega)$ aus dem ersten Ausdruck in (325.2) zu $E(\Omega) = \sigma^2(n - u) + \beta'\mathbf{X}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X}\beta$ und somit

$$E(\Omega) = \sigma^2(n - u) . \quad (325.5)$$

Die mit der Maximum-Likelihood-Methode abgeleitete Schätzung $\hat{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit ist nicht erwartungstreu, denn man erhält mit (324.5) und (325.5)

$$E(\hat{\sigma}^2) = \frac{1}{n}E(\Omega) = \frac{1}{n}\sigma^2(n - u) \neq \sigma^2 .$$

Die erwartungstreue Schätzung $\hat{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit in den beiden Modellen (321.1) und (321.5) ergibt sich zu

$$\hat{\sigma}^2 = \Omega/(n - u) . \quad (325.6)$$

Diese Schätzung von σ^2 ist unter der Voraussetzung normalverteilter Beobachtungen auch die beste Schätzung im Sinne von (312.2), wie mit (364.5) gezeigt wird.

Aus (322.9) ergibt sich die mit $\hat{\sigma}^2$ berechnete Kovarianzmatrix $\hat{D}(\hat{\beta})$ der Schätzwerte der Parameter zu

$$\hat{D}(\hat{\beta}) = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1} \quad (325.7)$$

und aus (321.5) sowie aus (233.2) mit (323.4) die mit $\hat{\sigma}^2$ berechneten Kovarianzmatrizen $\hat{D}(\mathbf{y})$ der Beobachtungen und $\hat{D}(\hat{\mathbf{y}})$ der geschätzten Erwartungswerte der Beobachtungen zu

$$\hat{D}(\mathbf{y}) = \hat{\sigma}^2\mathbf{I} \quad \text{und} \quad \hat{D}(\hat{\mathbf{y}}) = \hat{\sigma}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' . \quad (325.8)$$

Bildet man mit (137.3) die Spur beider Kovarianzmatrizen, folgt

$$\text{sp}\hat{D}(\mathbf{y}) = \hat{\sigma}^2 n \quad \text{und} \quad \text{sp}\hat{D}(\hat{\mathbf{y}}) = \hat{\sigma}^2 u \quad (325.9)$$

und wegen $n \geq u$

$$\text{sp}\hat{D}(\mathbf{y}) \geq \text{sp}\hat{D}(\hat{\mathbf{y}}) . \quad (325.10)$$

Beispiel: Eine Strecke s sei n -mal gemessen worden, wobei die Beobachtungen y_i voneinander unabhängig seien und gleiche Varianzen besitzen sollen. Gesucht ist der Schätzwert \hat{s} der Strecke s und seine Varianz.

Das Gauß–Markoff–Modell ergibt sich zu

$$\begin{array}{l} s = E(y_1) \\ s = E(y_2) \\ \dots\dots\dots \\ s = E(y_n) \end{array} \quad \text{oder} \quad \begin{array}{l} s = y_1 + e_1 \\ s = y_2 + e_2 \\ \dots\dots\dots \\ s = y_n + e_n \end{array} \quad \text{mit} \quad D\left(\begin{array}{c} y_1 \\ y_2 \\ \dots \\ y_n \end{array}\right) = \sigma^2 \begin{array}{c} \left| \begin{array}{cccc} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{array} \right| \end{array}$$

und abgekürzt geschrieben

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{oder} \quad \mathbf{X}\boldsymbol{\beta} = \mathbf{y} + \mathbf{e} \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I} ,$$

so daß das Modell (321.5) vorliegt mit

$$\mathbf{X} = \begin{array}{c} \left| \begin{array}{c} 1 \\ 1 \\ \dots \\ 1 \end{array} \right| , \quad \boldsymbol{\beta} = s, \quad \mathbf{y} = \begin{array}{c} \left| \begin{array}{c} y_1 \\ y_2 \\ \dots \\ y_n \end{array} \right| \quad \text{und} \quad \mathbf{e} = \begin{array}{c} \left| \begin{array}{c} e_1 \\ e_2 \\ \dots \\ e_n \end{array} \right| . \end{array}$$

Für die Schätzung gilt $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ mit $D(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ oder

$$\hat{s} = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{und} \quad V(\hat{s}) = \frac{\sigma^2}{n} ,$$

so daß als Schätzwert \hat{s} das arithmetische Mittel erhalten wird. Für die Schätzung der Varianz der Gewichtseinheit gilt $\hat{\sigma}^2 = \Omega/(n - u)$ mit $\Omega = \hat{\mathbf{e}}'\hat{\mathbf{e}}$ oder

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (\hat{s} - y_i)^2 \quad \text{und weiter} \quad \hat{V}(\hat{s}) = \frac{\hat{\sigma}^2}{n} , \quad \hat{D}(\mathbf{y}) = \hat{\sigma}^2 \mathbf{I} .$$

Der Schätzwert \hat{s} der Strecke s soll nun aus voneinander unabhängigen Beobachtungen y_i mit unterschiedlichen Gewichten bestimmt werden.

Für das Gewicht p_i einer Beobachtung y_i gilt nach (233.19) $p_i = c/\sigma_i^2$, wobei c eine Konstante und σ_i^2 die Varianz von y_i bedeuten. Dann folgt

$$D\left(\begin{array}{c} y_1 \\ y_2 \\ \dots \\ y_n \end{array}\right) = \begin{array}{c} \left| \begin{array}{cccc} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{array} \right| \\ \dots \\ \left| \begin{array}{cccc} c/p_1 & 0 & \dots & 0 \\ 0 & c/p_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & c/p_n \end{array} \right| \end{array}.$$

Mit $c = \sigma^2$ ergibt sich das Gauß–Markoff–Modell

$$\begin{array}{ll} s = E(y_1) & s = y_1 + e_1 \\ s = E(y_2) & s = y_2 + e_2 \\ \dots & \dots \\ s = E(y_n) & s = y_n + e_n \end{array} \quad \text{oder}$$

$$\text{mit } D\left(\begin{array}{c} y_1 \\ y_2 \\ \dots \\ y_n \end{array}\right) = \sigma^2 \begin{array}{c} \left| \begin{array}{cccc} 1/p_1 & 0 & \dots & 0 \\ 0 & 1/p_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/p_n \end{array} \right| \end{array}$$

oder abgekürzt geschrieben

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{oder} \quad \mathbf{X}\boldsymbol{\beta} = \mathbf{y} + \mathbf{e} \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1},$$

so daß das Modell (321.1) vorliegt mit

$$\mathbf{X} = \begin{array}{c} \left| \begin{array}{c} 1 \\ 1 \\ \dots \\ 1 \end{array} \right|, \quad \boldsymbol{\beta} = s, \quad \mathbf{y} = \begin{array}{c} \left| \begin{array}{c} y_1 \\ y_2 \\ \dots \\ y_n \end{array} \right|, \\ \\ \mathbf{e} = \begin{array}{c} \left| \begin{array}{c} e_1 \\ e_2 \\ \dots \\ e_n \end{array} \right|, \quad \mathbf{P}^{-1} = \begin{array}{c} \left| \begin{array}{cccc} 1/p_1 & 0 & \dots & 0 \\ 0 & 1/p_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/p_n \end{array} \right|. \end{array} \end{array}$$

Für die Schätzung gilt $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{y}$ mit $D(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}$ oder mit $\mathbf{X}'\mathbf{P} = [p_1, p_2, \dots, p_n]$

$$\hat{s} = \frac{1}{\sum_{i=1}^n p_i} \sum_{i=1}^n p_i y_i \quad \text{mit} \quad V(\hat{s}) = \frac{\sigma^2}{\sum_{i=1}^n p_i},$$

so daß als Schätzwert \hat{s} das gewogene arithmetische Mittel erhalten wird. Für die Schätzung der Varianz der Gewichtseinheit gilt $\hat{\sigma}^2 = \Omega/(n - u)$ mit

$\Omega = \hat{\mathbf{e}}' \mathbf{P} \hat{\mathbf{e}}$ oder mit $\hat{\mathbf{e}}' \mathbf{P} = |p_1(\hat{s} - y_1), p_2(\hat{s} - y_2), \dots, p_n(\hat{s} - y_n)|$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n p_i (\hat{s} - y_i)^2$$

und weiter

$$\hat{V}(\hat{s}) = \frac{\hat{\sigma}^2}{\sum_{i=1}^n p_i}, \quad \hat{D}(\mathbf{y}) = \hat{\sigma}^2 \begin{vmatrix} 1/p_1 & 0 & \dots & 0 \\ 0 & 1/p_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/p_n \end{vmatrix}.$$

Wie sich leicht zeigen läßt, führt die Transformation (321.4) des Modells mit $D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}$ als Kovarianzmatrix in das Modell mit $D(\mathbf{y}) = \sigma^2 \mathbf{I}$ zu identischen Schätzwerten.

326 Numerische Berechnung der Schätzwerte und ihrer Kovarianzen

Die Gleichungen

$$\mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}' \mathbf{y} \quad (326.1)$$

zur Berechnung der Schätzwerte $\hat{\boldsymbol{\beta}}$ der Parameter aus (322.9) bezeichnet man, wie bereits im Zusammenhang mit (323.8) erwähnt, als Normalgleichungen und die Matrix $\mathbf{X}' \mathbf{X}$ als *Normalgleichungsmatrix*. Die Normalgleichungen sind eindeutig lösbar, da, wie bei (322.4) erläutert, die Inverse $(\mathbf{X}' \mathbf{X})^{-1}$ existiert und eindeutig ist. Braucht lediglich der Vektor $\hat{\boldsymbol{\beta}}$ berechnet zu werden, genügt die Gaußsche Elimination (133.16) mit anschließender Rückrechnung (133.17). Soll auch die Kovarianzmatrix $D(\hat{\boldsymbol{\beta}})$ in (322.9) angegeben werden, ist die Normalgleichungsmatrix zu invertieren. Für die Numerik setzt man elektronische Datenverarbeitungsanlagen ein, um mit Rechenprogrammen für die im Kapitel 133 erläuterten Verfahren die Normalgleichungen zu lösen oder die Normalgleichungsmatrix zu invertieren.

Dennoch soll im folgenden eine Rechenmethode entwickelt werden, die auf der Gaußschen Elimination basierend sich für Taschenrechner eignet. Will man $\hat{\boldsymbol{\beta}}$ und $D(\hat{\boldsymbol{\beta}})$ mit der Gaußschen Elimination ermitteln, geht man nach (134.9) vor. Soll die Kovarianzmatrix $D(\mathbf{F}\hat{\boldsymbol{\beta}})$ der Schätzwerte für $\mathbf{F}\boldsymbol{\beta}$ berechnet werden, wobei \mathbf{F} eine $m \times u$ Matrix bedeutet, benötigt man nach (322.8) die Matrix $\mathbf{F}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{F}'$. Diese Matrix ergibt sich nach (134.8) durch die Gaußsche Elimination, falls die Normalgleichungsmatrix um die Matrizen \mathbf{F}, \mathbf{F}'

und die $m \times m$ Nullmatrix $\mathbf{0}$ erweitert wird. Faßt man die einzelnen Eliminationsschritte in einer Blockmatrix zusammen, erhält man

$$\left| \begin{array}{cc|cc} \mathbf{I} & \mathbf{0} & \mathbf{X}'\mathbf{X} & \mathbf{F}' \\ -\mathbf{F}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{I} & \mathbf{F} & \mathbf{0} \end{array} \right| = \left| \begin{array}{cc|cc} \mathbf{X}'\mathbf{X} & \mathbf{F}' & & \\ \mathbf{0} & -\mathbf{F}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}' & & \end{array} \right|. \quad (326.2)$$

Ebenfalls mit Hilfe der Gaußschen Elimination läßt sich die Quadratsumme Ω der Residuen ermitteln, die nach (325.6) zur Schätzung der Varianz der Gewichtseinheit benötigt wird. Hierzu erweitert man die Normalgleichungsmatrix um die Absolutglieder $\mathbf{X}'\mathbf{y}$ der Normalgleichungen und die Quadratsumme $\mathbf{y}'\mathbf{y}$ der Beobachtungen. Man erhält nach (134.8) durch die Gaußsche Elimination

$$\left| \begin{array}{cc|cc} \mathbf{I} & \mathbf{0} & \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{y} \\ -\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & 1 & \mathbf{y}'\mathbf{X} & \mathbf{y}'\mathbf{y} \end{array} \right| = \left| \begin{array}{cc|cc} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{y} & & \\ \mathbf{0}' & \Omega & & \end{array} \right|, \quad (326.3)$$

denn es gilt $\Omega = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ wegen (325.4). Sollen $\hat{\beta}$, $D(\hat{\beta})$, $\mathbf{F}\hat{\beta}$, $D(\mathbf{F}\hat{\beta})$ und Ω mit der Gaußschen Elimination berechnet werden, erhält man mit (134.8), (134.9), (326.2) und (326.3)

$$\left| \begin{array}{cccc|cccc} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{y} & \mathbf{I} & \mathbf{F}' \\ -\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & 1 & \mathbf{0}' & \mathbf{0}' & \mathbf{y}'\mathbf{X} & \mathbf{y}'\mathbf{y} & \mathbf{0}' & \mathbf{0}' \\ -(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\mathbf{F}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{F} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{array} \right| = \left| \begin{array}{cccc|cccc} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{y} & \mathbf{I} & \mathbf{F}' & & & & \\ \mathbf{0}' & \Omega & -\hat{\beta}' & -\hat{\beta}'\mathbf{F}' & & & & \\ \mathbf{0} & -\hat{\beta} & -(\mathbf{X}'\mathbf{X})^{-1} & -(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}' & & & & \\ \mathbf{0} & -\mathbf{F}\hat{\beta} & -\mathbf{F}(\mathbf{X}'\mathbf{X})^{-1} & -\mathbf{F}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{F}' & & & & \end{array} \right|. \quad (326.4)$$

Ersetzt man in (326.1) bis (326.4) $\mathbf{X}'\mathbf{X}$ durch $\mathbf{X}'\mathbf{P}\mathbf{X}$, $\mathbf{X}'\mathbf{y}$ durch $\mathbf{X}'\mathbf{P}\mathbf{y}$ und $\mathbf{y}'\mathbf{y}$ durch $\mathbf{y}'\mathbf{P}\mathbf{y}$ folgt das Normalgleichungssystem für das Gauß-Markoff-Modell (321.1). Da die Normalgleichungen in (326.1) und die erweiterten Systeme in (326.2) bis (326.4) symmetrisch sind, brauchen bei der Gaußschen Elimination jeweils nur die Glieder auf und oberhalb der Diagonalen berücksichtigt zu werden. Zur Rechenkontrolle der numerischen Rechnung kann, wie im folgenden Beispiel gezeigt wird, eine Spaltenspalte gebildet und in den Eliminationsprozeß einbezogen werden. Eine weitere Rechenprobe folgt mit der doppelten Berechnung von Ω . Einmal erhält man Ω aus (325.1) als Quadratsumme oder quadratische Form der Residuen, zum anderen ergibt sich Ω aus (325.4) oder bei Gaußscher Elimination unmittelbar aus (326.4).

Eine Rechenkontrolle, die sogenannte Schlußprobe, der gesamten Berechnung der Schätzwerte einschließlich der Linearisierung (321.7) und der eventuell durchgeführten Transformation (321.4) besteht unter der Voraussetzung, daß Rechenfehler sich nicht gegenseitig aufheben, darin, daß im Modell (321.1) die Schätzwerte $y_i^* + \hat{e}_i$ in (321.6) einmal mit Hilfe des Vektors \hat{e} und zum anderen nach (321.7) mittels $h_i(\beta_{10} + \Delta\hat{\beta}_1, \dots, \beta_{u0} + \Delta\hat{\beta}_u)$ mit $\hat{\beta} = (\Delta\hat{\beta}_j)$ und $i \in \{1, \dots, n\}, j \in \{1, \dots, u\}$ berechnet werden. Stimmen die beiden Werte innerhalb der Rechengenauigkeit nicht überein, ohne daß ein Rechenfehler vorliegt, so wird die Differenz durch die Vernachlässigung der Glieder zweiter und höherer Ordnung in der Taylor–Entwicklung (321.7) infolge ungünstig gewählter Näherungswerte verursacht sein. In einem solchen Fall führt meistens, wie bereits im Zusammenhang mit (321.8) erwähnt, eine iterative Vorgehensweise zum Ziel.

Der Rechenablauf der Parameterschätzung besteht also aus den folgenden Schritten:

1. Aufstellen der Beobachtungsgleichungen,
2. Linearisierung oder, falls bei linearen Beobachtungsgleichungen zweckmäßig, ebenfalls Einführung von Näherungswerten,
3. Aufstellen der Normalgleichungen,
4. Berechnung der Schätzwerte, Varianzen und Kovarianzen,
5. Probe der Quadratsumme oder der quadratischen Form der Residuen,
6. Schlußprobe.

Die Parameterschätzung nach (322.9) impliziert, daß mit den Beobachtungsgleichungen (321.2) oder mit den nach (321.4) transformierten Beobachtungsgleichungen die Koeffizientenmatrix \mathbf{X} aufgebaut und dann die Normalgleichungsmatrix $\mathbf{X}'\mathbf{P}\mathbf{X}$ oder $\mathbf{X}'\mathbf{X}$ gebildet wird. Bei sehr vielen, voneinander unabhängigen Beobachtungen kann es rechentechnisch günstiger sein, das Aufstellen der Koeffizientenmatrix \mathbf{X} zu vermeiden und den Beitrag jeder Beobachtungsgleichung zu den Normalgleichungen unmittelbar zu berechnen.

Infolge der Unabhängigkeit der Beobachtungen gelte $\mathbf{P} = \text{diag}(p_1, \dots, p_n)$ und die Koeffizientenmatrix \mathbf{X} werde durch ihre Zeilen \mathbf{x}'_i mit $\mathbf{X} = |\mathbf{x}_1, \dots, \mathbf{x}_n|'$ und $i \in \{1, \dots, n\}$ dargestellt, dann folgt aus (321.2) die Beobachtungsgleichung für y_i mit

$$\mathbf{x}'_i\boldsymbol{\beta} = E(y_i) = y_i + e_i \quad \text{und} \quad V(y_i) = \sigma^2/p_i. \quad (326.5)$$

Die Normalgleichungsmatrix $\mathbf{X}'\mathbf{P}\mathbf{X}$ und der Absolutgliedvektor $\mathbf{X}'\mathbf{P}\mathbf{y}$ aus (322.9) ergeben sich mit $\mathbf{X}'\mathbf{P} = |p_1\mathbf{x}_1, \dots, p_n\mathbf{x}_n|$ zu

$$\begin{aligned}\mathbf{X}'\mathbf{P}\mathbf{X} &= |p_1\mathbf{x}_1\mathbf{x}'_1 + \dots + p_n\mathbf{x}_n\mathbf{x}'_n| \\ \mathbf{X}'\mathbf{P}\mathbf{y} &= |p_1y_1\mathbf{x}_1 + \dots + p_ny_n\mathbf{x}_n|. \end{aligned} \quad (326.6)$$

Die Beiträge der einzelnen Beobachtungsgleichungen fügen sich also additiv zu den Normalgleichungen zusammen.

Beispiel: Gegeben seien drei Punkte P_1, P_2 und P_3 in einer Ebene mit ihren $(x_i; y_i)$ -Koordinaten in der Dimension Meter: $(235, 25; 110, 52)$ für P_1 , $(726, 14; 147, 38)$ für P_2 und $(883, 67; 621, 01)$ für P_3 . Zu einem vierten Punkt P_4 in der Ebene seien, wie in Abbildung 326-1 dargestellt, von P_1, P_2 und P_3

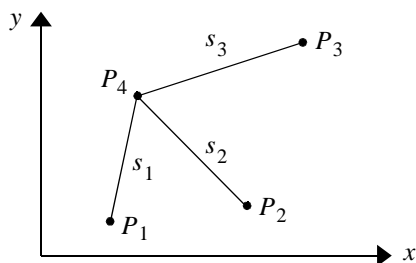


Abb. 326-1: Punkte und Strecken

die Strecken $s_1 = 391,544$ m; $s_2 = 546,146$ m und $s_3 = 591,545$ m gemessen worden. Die Messungen seien unkorreliert und ihre Gewichte mit $\mathbf{P} = \text{diag}(p_1, p_2, p_3)$ in der Dimension $1/\text{cm}^2$ gegeben durch $p_1 = 0,444$; $p_2 = 0,250$ und $p_3 = 0,160$. Gesucht sind die Schätzwerte der Koordinaten $(x_4; y_4)$ von P_4 , ihre Kovarianzmatrix und die Schätzung der Varianz σ^2 der Gewichtseinheit. Damit liegt das Modell (321.1) vor, in dem die Schätzwerte nach den für (321.1) gültigen Formeln berechnet werden sollen. Wie im Beispiel des Kapitels 374 geschehen, könnte man auch numerisch das Modell (321.1) mit Hilfe von (321.4) in das Modell (321.5) überführen und in diesem Modell die Schätzungen berechnen, was auf identische Ergebnisse führen würde.

Mit $s_i = ((x_i - x_4)^2 + (y_i - y_4)^2)^{1/2}$ und $i \in \{1, 2, 3\}$ stellen die Beobachtungen eine nichtlineare Funktion der unbekannt Parameter dar, so daß linearisiert werden muß. Näherungskordinaten $(x_{40}; y_{40})$ für P_4 lassen sich mit Hilfe von s_1 und s_2 und der Koordinaten von P_1 und P_2 berechnen, und zwar wird mit dem Kosinussatz aus s_1, s_2 und der Strecke P_1P_2 der Winkel in P_1 zwischen den Schenkeln P_1P_2 und P_1P_4 ermittelt. Zu die-

sem Winkel ist der Winkel in P_1 zwischen der Parallelen zur x -Achse durch P_1 und der Strecke P_1P_2 zu addieren, um mit dem so gewonnenen Winkel und s_1 die Näherungskoordinaten (305,497; 495,711) für P_4 zu erhalten. Für $s_{i0} = ((x_i - x_{40})^2 + (y_i - y_{40})^2)^{1/2}$ mit $i \in \{1, 2, 3\}$ folgt dann $s_{10} = 391,544$ m; $s_{20} = 546,146$ m und $s_{30} = 591,594$ m, so daß sich mit

$$\left. \frac{\partial s_i}{\partial x_4} \right|_0 = -\frac{x_i - x_{40}}{s_{i0}} \quad \text{und} \quad \left. \frac{\partial s_i}{\partial y_4} \right|_0 = -\frac{y_i - y_{40}}{s_{i0}}$$

und $\beta = |\Delta x_4, \Delta y_4|'$ aus (321.8) mit etwa gleichen Größenordnungen die Matrix \mathbf{X} , der Vektor \mathbf{y} in der Dimension 0,1 m und die Gewichtsmatrix \mathbf{P} in der Dimension $1/(0,1 \text{ m})^2$ ergeben zu

$$\mathbf{X} = \begin{vmatrix} 0,179 & 0,984 \\ -0,770 & 0,638 \\ -0,977 & -0,212 \end{vmatrix}, \quad \mathbf{y} = \begin{vmatrix} 0 \\ 0 \\ -0,49 \end{vmatrix},$$

$$\mathbf{P} = 100 \begin{vmatrix} 0,444 & 0 & 0 \\ & 0,250 & 0 \\ & & 0,160 \end{vmatrix}.$$

Die Normalgleichungen $\mathbf{X}'\mathbf{P}\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{P}\mathbf{y}$ werden mit Hilfe der im Beispiel zu (131.6) angegebenen Probe durch Zeilensummen aufgestellt. Die Lösung soll nach (326.4) erhalten werden, so daß das Schema (326.7) folgt, wenn man die einzelnen Eliminationsschritte untereinander schreibt, wobei die Faktoren und Summanden bei der Elimination in Klammern angegeben sind.

Mit dem Vektor $\hat{\beta}$ läßt sich der Vektor \hat{e} der Residuen aus (323.4) berechnen, so daß man in der Dimension 0,1 m erhält $\hat{e} = |0,080; -0,165; 0,243|'$ und damit $\Omega = \hat{e}'\mathbf{P}\hat{e} = 1,9$ in ausreichender Übereinstimmung mit dem Ergebnis aus der Elimination in (326.7). Als eine weitere mögliche Rechenprobe für die Residuen ergibt sich aus (323.10) $0,01 \cdot \mathbf{X}'\mathbf{P}\hat{e} = |0,000; 0,000|'$. Die geschätzten Erwartungswerte der Beobachtungen $\hat{s}_i = s_i + \hat{e}_i$ ergeben sich zu $\hat{s}_1 = 391,522$ m; $\hat{s}_2 = 546,130$ m und $\hat{s}_3 = 591,569$ m. Bis auf 0,001 m übereinstimmende Werte erhält man, wenn \hat{s}_1, \hat{s}_2 und \hat{s}_3 aus den für P_4 geschätzten Koordinaten $\hat{x}_4 = x_{40} + \Delta\hat{x}_4$ und $\hat{y}_4 = y_{40} + \Delta\hat{y}_4$ mit $\hat{x}_4 = 305,521$ m und $\hat{y}_4 = 495,715$ m und den Koordinaten für P_1 bis P_3 berechnet werden, so daß die gesamte Rechnung einschließlich der Linearisierung geprüft ist. Mit (325.6) folgt weiter $\hat{\sigma}^2 = 1,8/(3-2)$ und mit (325.7) und (325.8) die Kovarianzmatrizen $\hat{D}(\mathbf{y})$ und $\hat{D}(\hat{\beta})$ in der Dimension $(0,01 \text{ m})^2$

$$\hat{D}(\mathbf{y}) = 1,8 \begin{vmatrix} 2,25 & 0 & 0 \\ & 4,00 & 0 \\ & & 6,25 \end{vmatrix}, \quad \hat{D}(\hat{\beta}) = 1,8 \begin{vmatrix} 3,18 & 0,06 \\ & 1,86 \end{vmatrix}.$$

$0,01 \cdot \mathbf{X}'\mathbf{P}\mathbf{X}$		$0,01 \cdot \mathbf{X}'\mathbf{P}\mathbf{y}$	\mathbf{I}		Summe
0,315	-0,011	0,077	1	0	1,381
	0,539	0,017	0	1	1,545
(0,0349)	(0)	(0,003)	(0,035)	(0)	(0,048)
		0,038	0	0	0,132
(-0,244)		(-0,019)	(-0,244)	(0)	(-0,337)
			0	0	1
(-3,175)			(-3,175)	(0)	(-4,385)
				0	1
(0)				(0)	(0)
	0,539	0,020	0,035	1	1,593
		0,019	-0,244	0	-0,205
(-0,0371)		(-0,001)	(-0,001)	(-0,037)	(-0,059)
			-3,175	0	-3,385
(-0,0649)			(-0,002)	(-0,065)	(-0,103)
				0	1
(-1,855)				(-1,855)	(-2,955)
$0,01 \cdot \Omega$ und $-\hat{\beta}'$		0,018	-0,245	-0,037	-0,264
		-100 · $(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}$	-3,177	-0,065	-3,488
				-1,855	-1,955

(326.7)

327 Gauß–Markoff–Modell mit Restriktionen

Die unbekannt Parameter des Gauß–Markoff–Modells sollen jetzt linearen Restriktionen unterworfen werden.

Definition: Gelten im Gauß–Markoff–Modell (321.1) für die Parameter β zusätzlich die Restriktionen $\mathbf{H}\beta = \mathbf{w}$, wobei \mathbf{H} eine $r \times u$ Matrix bekannter

Koeffizienten mit $\text{rg}\mathbf{H} = r$ sowie $r \leq u$ und \mathbf{w} ein bekannter $r \times 1$ Vektor bedeuten, so bezeichnet man

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad \mathbf{H}\boldsymbol{\beta} = \mathbf{w} \quad \text{und} \quad D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}$$

als Gauß–Markoff–Modell mit *Restriktionen*. (327.1)

Bei nichtlinearen Restriktionen ist mit Hilfe von Näherungswerten für die Parameter entsprechend (321.6) bis (321.8) zu linearisieren.

Mit den Transformationen (321.4) läßt sich das Modell (327.1) in

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad \mathbf{H}\boldsymbol{\beta} = \mathbf{w} \quad \text{und} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I} \quad (327.2)$$

überführen, in dem im folgenden die Parameterschätzung abgeleitet wird. Die Ergebnisse lassen sich dann wieder mit (321.4) in die Schätzungen im Modell (327.1) transformieren.

Die Restriktionen werden als Gleichungen formuliert. Müssen Ungleichungen als Restriktionen dienen, ist ein quadratisches Programmierproblem zu lösen, beispielsweise bei Optimierungsaufgaben (FRITSCH 1985; KOCH 1982B; SCHAFFRIN 1981), worauf hier aber nicht eingegangen wird.

Mit $\text{rg}\mathbf{H} = r$ folgt wegen (135.5) $\mathbf{w} \in R(\mathbf{H})$ und damit nach (154.2) die Konsistenz des Gleichungssystems $\mathbf{H}\boldsymbol{\beta} = \mathbf{w}$. Gilt $r < u$, lassen sich also r unbekannte Parameter $\boldsymbol{\beta}$ mit $\mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ aus $\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y})$ eliminieren, um dann für die verbleibenden wie mit (321.5) vorzugehen. Rechentechnisch günstigere Formeln ergeben sich aber, wenn die unbekannt Parameter in (327.1) oder (327.2) nach den Verfahren der Kapitel 312 bis 314 geschätzt werden. Gilt $r = u$, so sind mit $\boldsymbol{\beta} = \mathbf{H}^{-1}\mathbf{w}$ wegen (133.1) die Parameter $\boldsymbol{\beta}$ eindeutig bestimmt.

a) Beste lineare erwartungstreue Schätzung

Die lineare Funktion $\mathbf{a}'\boldsymbol{\beta}$ der unbekannt Parameter $\boldsymbol{\beta}$ soll durch die lineare Funktion $\mathbf{c}'\mathbf{y}$ der Beobachtungen \mathbf{y} geschätzt werden, zu der wegen der Restriktionen noch die lineare Funktion $\mathbf{d}'\mathbf{w}$ zu addieren ist, wobei der $u \times 1$ Vektor \mathbf{a} bekannt und der $n \times 1$ Vektor \mathbf{c} sowie der $r \times 1$ Vektor \mathbf{d} zu bestimmen sind. Für eine beste lineare erwartungstreue Schätzung wird nach (312.2) für alle $\boldsymbol{\beta}$

$$1) \quad E(\mathbf{c}'\mathbf{y} + \mathbf{d}'\mathbf{w}) = \mathbf{a}'\boldsymbol{\beta}$$

oder mit (327.2) $\mathbf{c}'\mathbf{X}\boldsymbol{\beta} + \mathbf{d}'\mathbf{H}\boldsymbol{\beta} = \mathbf{a}'\boldsymbol{\beta}$ gefordert. Somit folgt

$$\mathbf{c}'\mathbf{X} + \mathbf{d}'\mathbf{H} = \mathbf{a}' \quad (327.3)$$

Ferner soll

$$2) \quad V(\mathbf{c}'\mathbf{y} + \mathbf{d}'\mathbf{w}) \quad \text{minimal}$$

werden, wofür man wegen (233.2) und (327.2) erhält

$$V(\mathbf{c}'\mathbf{y} + \mathbf{d}'\mathbf{w}) = V(\mathbf{c}'\mathbf{y}) = \sigma^2 \mathbf{c}'\mathbf{c}. \quad (327.4)$$

Die zur Lösung dieses Minimumproblems nach (171.6) zu bildende Lagrange-sche Funktion lautet für $r < u$ mit dem $u \times 1$ Vektor $-2\mathbf{k}$ der Lagrangeschen Multiplikatoren $w(\mathbf{c}, \mathbf{d}) = \sigma^2 \mathbf{c}'\mathbf{c} - 2\mathbf{k}'(\mathbf{X}'\mathbf{c} + \mathbf{H}'\mathbf{d} - \mathbf{a})$. Aus $\partial w(\mathbf{c}, \mathbf{d})/\partial \mathbf{c} = 2\sigma^2 \mathbf{c} - 2\mathbf{X}\mathbf{k} = \mathbf{0}$ wegen (172.1) und (172.2) folgt $\mathbf{c} = \mathbf{X}\mathbf{k}/\sigma^2$ und mit (327.3) $\mathbf{X}'\mathbf{X}\mathbf{k}/\sigma^2 + \mathbf{H}'\mathbf{d} = \mathbf{a}$ oder $\mathbf{k}/\sigma^2 = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{a} - \mathbf{H}'\mathbf{d})$, denn wie bereits bei (322.4) erwähnt, existiert $(\mathbf{X}'\mathbf{X})^{-1}$. Mit $\partial w(\mathbf{c}, \mathbf{d})/\partial \mathbf{d} = -2\mathbf{H}\mathbf{k} = \mathbf{0}$ folgt $\mathbf{H}\mathbf{k}/\sigma^2 = \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{a} - \mathbf{H}'\mathbf{d}) = \mathbf{0}$ und hieraus

$$\mathbf{d} = (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}, \quad (327.5)$$

denn mit $\text{rg}\mathbf{H} = r$ ist die Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'$ wegen (143.7) positiv definit und daher wegen (143.3) regulär, da auch $(\mathbf{X}'\mathbf{X})^{-1}$ wegen (143.8) und (143.9) positiv definit ist. Weiter erhält man

$$\mathbf{c} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}[\mathbf{I} - \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{a} \quad (327.6)$$

und anstelle von (327.4)

$$V(\mathbf{c}'\mathbf{y}) = \sigma^2 \mathbf{a}'[(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{a}. \quad (327.7)$$

Um zu zeigen, daß (327.7) minimal ist, daß also die zweite Bedingung in (312.2) $V(\mathbf{c}'\mathbf{y} + \mathbf{d}'\mathbf{w}) \leq V(\mathbf{c}^*\mathbf{y} + \mathbf{d}^*\mathbf{w})$ erfüllt ist, wobei $\mathbf{c}^*\mathbf{y} + \mathbf{d}^*\mathbf{w}$ eine beliebige erwartungstreue Schätzung von $\mathbf{a}'\boldsymbol{\beta}$ bedeutet, geht man analog zu (322.6) vor und erhält mit \mathbf{c} aus (327.6) anstelle von (322.7)

$$f = \sigma^2 (\mathbf{c}^* - \mathbf{c})' \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} [\mathbf{I} - \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}] \mathbf{a}$$

und mit $\mathbf{c}^*\mathbf{X} = \mathbf{a}' - \mathbf{d}^*\mathbf{H}$ aus (327.3) schließlich $f = 0$, da $\mathbf{d}^*\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}[\mathbf{I} - \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{a} = 0$ gilt.

Bezeichnet man mit $\widetilde{\mathbf{a}'\boldsymbol{\beta}}$ die beste erwartungstreue Schätzung von $\mathbf{a}'\boldsymbol{\beta}$, so gilt $\widetilde{\mathbf{a}'\boldsymbol{\beta}} = \mathbf{c}'\mathbf{y} + \mathbf{d}'\mathbf{w}$ mit \mathbf{d} und \mathbf{c} aus (327.5) und (327.6) sowie $V(\widetilde{\mathbf{a}'\boldsymbol{\beta}})$ aus (327.7). Durch entsprechende Wahl von \mathbf{a} wie für (322.9) ergibt sich hieraus $\widetilde{\boldsymbol{\beta}}$ und mit (233.2) und (327.2) $D(\widetilde{\boldsymbol{\beta}})$. Wegen $\mathbf{H}\widetilde{\boldsymbol{\beta}} = \mathbf{w}$ erhält man für $r = u$ mit (327.1) oder (327.2) $\widetilde{\boldsymbol{\beta}} = \mathbf{H}^{-1}\mathbf{w} = \boldsymbol{\beta}$. Zusammenfassend gilt der

Satz: Die beste lineare erwartungstreue Schätzung $\tilde{\beta}$ der unbekannt Parameter β im Gauß–Markoff–Modell (327.2) mit Restriktionen ist gegeben durch

$$\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'\mathbf{y} + \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{w} - \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})]$$

und die Kovarianzmatrix $D(\tilde{\beta})$ durch

$$D(\tilde{\beta}) = \sigma^2[(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}]. \quad (327.8)$$

Mit (321.4) ergeben sich die entsprechenden Schätzwerte im Modell (327.1), indem in (327.8) $\mathbf{X}'\mathbf{X}$ durch $\mathbf{X}'\mathbf{P}\mathbf{X}$ und $\mathbf{X}'\mathbf{y}$ durch $\mathbf{X}'\mathbf{P}\mathbf{y}$ ersetzt werden. Für die numerische Berechnung der Schätzwerte sind die im folgenden angegebenen Formeln vorzuziehen.

b) Methode der kleinsten Quadrate und Maximum–Likelihood–Methode

Bei der Parameterschätzung nach der Methode der kleinsten Quadrate im Modell (327.2) ist nach (313.1) die Quadratsumme $(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)/\sigma^2$ unter der Bedingung $\mathbf{H}\beta = \mathbf{w}$ zu minimieren. Wie aus (324.2) ersichtlich, ist im Falle normalverteilter Beobachtungen bei der Maximum–Likelihood–Schätzung im Modell (327.2) die gleiche Quadratsumme, abgesehen von dem Faktor $1/2$, unter den gleichen Bedingungen zu minimieren, so daß beide Methoden identische Schätzungsergebnisse liefern werden.

Nach (171.6) ergibt sich die Lagrangesche Funktion $w(\beta)$ mit dem $r \times 1$ Vektor $2\mathbf{k}/\sigma^2$ der Lagrangeschen Multiplikatoren zu $w(\beta) = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)/\sigma^2 + 2\mathbf{k}'(\mathbf{H}\beta - \mathbf{w})/\sigma^2$ und daraus mit (172.1) und (172.2) $\partial w(\beta)/\partial \beta = -2\mathbf{X}'\mathbf{y}/\sigma^2 + 2\mathbf{X}'\mathbf{X}\beta/\sigma^2 + 2\mathbf{H}'\mathbf{k}/\sigma^2 = \mathbf{0}$. Zusammen mit den Bedingungsgleichungen folgen hieraus die Normalgleichungen für die Schätzwerte $\tilde{\beta}$ und die Multiplikatoren \mathbf{k} zu

$$\begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{H}' \\ \mathbf{H} & \mathbf{0} \end{vmatrix} \begin{vmatrix} \tilde{\beta} \\ \mathbf{k} \end{vmatrix} = \begin{vmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{w} \end{vmatrix}. \quad (327.9)$$

Die Werte für $\tilde{\beta}$ und \mathbf{k} werden hieraus eindeutig bestimmt, da die Normalgleichungsmatrix wegen (136.17) regulär ist, denn mit (136.12) gilt

$$\det \begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{H}' \\ \mathbf{H} & \mathbf{0} \end{vmatrix} = \det(\mathbf{X}'\mathbf{X}) \det(-\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}') \neq 0.$$

Die Schätzwerte $\tilde{\beta}$ aus (327.9) sind identisch mit denen aus (327.8), denn durch Elimination von $\tilde{\beta}$ in (327.9) nach (134.8) ergibt sich

$$-\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'\mathbf{k} = \mathbf{w} - \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (327.10)$$

Substituiert man diese Gleichung in (327.9), erhält man $\tilde{\beta}$ in (327.8). Mit (322.9) erhält man weiter

$$\tilde{\beta} = \hat{\beta} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\hat{\beta} - \mathbf{w}). \quad (327.11)$$

Die Normalgleichungsmatrix in (327.9) ist im Gegensatz zu $\mathbf{X}'\mathbf{X}$ wegen (143.4) nicht positiv definit, da die Gaußsche Elimination von $\tilde{\beta}$ nach (133.2) oder (134.8) auf negative Diagonalelemente führt, wie aus (327.10) zu erkennen ist. Dennoch eignen sich die Normalgleichungen (327.9) zur numerischen Ermittlung der Schätzwerte $\tilde{\beta}$, wobei allerdings die Cholesky-Faktorisierung (133.23) nicht anwendbar ist, falls nicht die Normalgleichungen nach Elimination von $\tilde{\beta}$ mit -1 multipliziert werden.

Die Inverse der Normalgleichungsmatrix in (327.9) ergibt sich aus (134.3) mit $\mathbf{N} = \mathbf{X}'\mathbf{X}$ zu

$$\begin{vmatrix} \mathbf{N} & \mathbf{H}' \\ \mathbf{H} & \mathbf{0} \end{vmatrix}^{-1} = \begin{vmatrix} \mathbf{N}^{-1} - \mathbf{N}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{N}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{N}^{-1} & \mathbf{N}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{N}^{-1}\mathbf{H}')^{-1} \\ (\mathbf{H}\mathbf{N}^{-1}\mathbf{H}')^{-1}\mathbf{H}\mathbf{N}^{-1} & -(\mathbf{H}\mathbf{N}^{-1}\mathbf{H}')^{-1} \end{vmatrix}, \quad (327.12)$$

so daß in der inversen Matrix an der Stelle von $\mathbf{X}'\mathbf{X}$ die zur Berechnung von $D(\tilde{\beta})$ aus (327.8) benötigte Matrix steht. Dies entspricht im Modell (321.5) der Berechnung von $D(\hat{\beta})$ in (322.9) aus $(\mathbf{X}'\mathbf{X})^{-1}$.

Wie im folgenden noch zu zeigen ist, wird mit $\tilde{\beta}$ die Quadratsumme der Residuen minimal. Zusammenfassend gilt daher der

Satz: Im Gauß-Markoff-Modell (327.2) mit Restriktionen liefern die beste lineare erwartungstreue Schätzung, die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode im Falle normalverteilter Beobachtungen identische Schätzwerte $\tilde{\beta}$ für die Parameter, die aus den Normalgleichungen (327.9) berechenbar sind. Ersetzt man dort $\mathbf{X}'\mathbf{X}$ durch $\mathbf{X}'\mathbf{P}\mathbf{X}$ und $\mathbf{X}'\mathbf{y}$ durch $\mathbf{X}'\mathbf{P}\mathbf{y}$ ergeben sich die Schätzwerte im Modell (327.1). (327.13)

Erweitert man also die Normalgleichungsmatrix $\mathbf{X}'\mathbf{X}$ für die Schätzung im Gauß-Markoff-Modell um die Matrix \mathbf{H} der Restriktionen, wie in (327.9) angegeben, so erhält man die Normalgleichungsmatrix für die Schätzung im Gauß-Markoff-Modell mit Restriktionen. Sind zu den Restriktionen weitere hinzuzufügen, etwa in Form von Hypothesen, die im Kapitel 42 behandelt werden, brauchen die Normalgleichungen lediglich um die zusätzlichen Restriktionen erweitert zu werden, um die entsprechenden Schätzwerte berechnen zu können.

c) Erwartungstreue Schätzung der Varianz der Gewichtseinheit

Wird die aus der Likelihoodfunktion (324.2) mit der Bedingung $\mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ sich ergebende Lagrangesche Funktion nach σ^2 differenziert und die Ableitung gleich Null gesetzt, erhält man (324.5) entsprechend als Maximum–Likelihood–Schätzung $\bar{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit

$$\bar{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \quad (327.14)$$

Um eine erwartungstreue Schätzung von σ^2 zu erhalten, wird wie in (325.5) der Erwartungswert der Quadratsumme Ω_H

$$\Omega_H = (\mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{y}) \quad (327.15)$$

ermittelt, in der der Vektor

$$\tilde{\boldsymbol{\epsilon}} = \mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{y} \quad (327.16)$$

wie in (323.4) die Residuen enthält. Ihre Kovarianzmatrix $D(\tilde{\boldsymbol{\epsilon}})$ ergibt sich mit (233.2) und (327.8) zu

$$D(\tilde{\boldsymbol{\epsilon}}) = \sigma^2[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']$$

mit

$$\text{rg}D(\tilde{\boldsymbol{\epsilon}}) = n - u + r \quad \text{und} \quad \text{sp}D(\tilde{\boldsymbol{\epsilon}}) = \sigma^2(n - u + r) \quad (327.17)$$

wegen (137.3) und (152.3), da die Matrix $D(\tilde{\boldsymbol{\epsilon}})/\sigma^2$ idempotent ist und beispielsweise $\text{sp}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \text{sp}(\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1})$ gilt. Nach (133.1) ist $D(\tilde{\boldsymbol{\epsilon}})$ singulär.

Mit $\Omega_H = (\mathbf{X}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})$ folgt wegen (323.8)

$$\Omega_H = (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}) + (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}). \quad (327.18)$$

Der zweite Summand wird mit (327.11) umgeformt, so daß sich mit (325.1) ergibt

$$\Omega_H = \Omega + R \quad \text{mit} \quad R = (\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{w})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{w}). \quad (327.19)$$

Durch die Einführung der Restriktionen $\mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ vergrößert sich im allgemeinen die Quadratsumme Ω , denn es gilt nach (143.1) $R \geq 0$, da die Matrix $\mathbf{H}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'$ und damit auch ihre Inverse positiv definit ist, wie bereits bei

(327.5) gezeigt wurde. Der Erwartungswert von Ω_H ergibt sich mit (271.1), (325.5) und $E(\mathbf{H}\hat{\boldsymbol{\beta}}) = \mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ sowie $D(\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{w}) = \sigma^2 \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'$ zu

$$\begin{aligned} E(\Omega_H) &= E(\Omega) + \sigma^2 \text{sp}[(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'] \\ &= \sigma^2(n - u + r). \end{aligned}$$

Die erwartungstreue Schätzung $\tilde{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit im Gauß-Markoff-Modell mit Restriktionen beträgt daher

$$\tilde{\sigma}^2 = \Omega_H / (n - u + r). \quad (327.20)$$

Mit den Schätzwerten $\tilde{\boldsymbol{\beta}}$ wird die Quadratsumme Ω_H der Residuen minimal, denn bedeutet $\boldsymbol{\beta}^*$ einen beliebigen $u \times 1$ Vektor, für den $\mathbf{H}\boldsymbol{\beta}^* = \mathbf{w}$ gilt, erhält man

$$\begin{aligned} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)' \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) &= (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)' \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \\ &= (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) + (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)' \mathbf{X}'\mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*), \end{aligned} \quad (327.21)$$

da mit (327.11) für $2(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)' \mathbf{X}'\mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = 0$ folgt, denn es gilt $\mathbf{H}\tilde{\boldsymbol{\beta}} = \mathbf{w}$ wegen (327.8) und $\mathbf{H}\boldsymbol{\beta}^* = \mathbf{w}$ nach Voraussetzung. Das Ergebnis (327.21) in (323.2) eingesetzt ergibt mit (327.18), da $\mathbf{X}'\mathbf{X}$ nach (143.8) positiv definit ist,

$$\begin{aligned} (\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y})'(\mathbf{X}\boldsymbol{\beta}^* - \mathbf{y}) &= (\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{y}) + (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})' \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \\ &+ (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)' \mathbf{X}'\mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = \Omega_H + (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)' \mathbf{X}'\mathbf{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \geq \Omega_H. \end{aligned} \quad (327.22)$$

Die Quadratsumme Ω_H der Residuen läßt sich (326.3) entsprechend durch Gaußsche Elimination in den Normalgleichungen (327.9) berechnen. Aus (327.15) ergibt sich $\Omega_H = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\tilde{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$ und mit $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y} - \mathbf{H}'\mathbf{k})$ aus (327.9) sowie $\mathbf{H}\tilde{\boldsymbol{\beta}} = \mathbf{w}$ aus (327.8)

$$\Omega_H = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{w}'\mathbf{k}. \quad (327.23)$$

Erweitert man die Normalgleichungsmatrix in (327.9) um den Absolutgliedvektor $|\mathbf{y}'\mathbf{X}, \mathbf{w}'|'$ sowie um die Quadratsumme $\mathbf{y}'\mathbf{y}$ und führt die Gaußsche Elimination durch, so wird (326.3) entsprechend gebildet

$$\mathbf{y}'\mathbf{y} - |\mathbf{y}'\mathbf{X}, \mathbf{w}'| \left| \begin{array}{cc} \mathbf{X}'\mathbf{X} & \mathbf{H}' \\ \mathbf{H} & \mathbf{0} \end{array} \right|^{-1} \left| \begin{array}{c} \mathbf{X}'\mathbf{y} \\ \mathbf{w} \end{array} \right| = \mathbf{y}'\mathbf{y} - |\mathbf{y}'\mathbf{X}, \mathbf{w}'| \left| \begin{array}{c} \tilde{\boldsymbol{\beta}} \\ \mathbf{k} \end{array} \right|, \quad (327.24)$$

woraus wegen (327.23) Ω_H folgt. Als Rechenprobe läßt sich also Ω_H einmal

aus (327.15) und zum anderen aus (327.23) oder bei Gaußscher Elimination aus (327.24) berechnen. Als eine weitere mögliche Rechenprobe für $\tilde{\mathbf{e}}$ erhält man mit $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y} - \mathbf{H}'\mathbf{k})$ aus (327.9) die Gleichung

$$\mathbf{X}'\tilde{\mathbf{e}} + \mathbf{H}'\mathbf{k} = \mathbf{0}, \quad (327.25)$$

die (323.8) entspricht.

d) Restriktionen als Beobachtungen mit sehr kleinen Varianzen

Das Gauß–Markoff–Modell (327.1) mit Restriktionen läßt sich in das Modell (321.1) ohne Restriktionen überführen, falls die Restriktionen als Beobachtungen mit sehr kleinen Varianzen angesehen werden (KOCH und POPE 1969). Um das zu zeigen, sei das Gauß–Markoff–Modell gegeben

$$\bar{\mathbf{X}}\boldsymbol{\beta} = E(\bar{\mathbf{y}}) \quad \text{mit} \quad D(\bar{\mathbf{y}}) = \sigma^2 \bar{\mathbf{P}}^{-1} \quad (327.26)$$

und

$$\bar{\mathbf{X}} = \begin{vmatrix} \mathbf{X} \\ \mathbf{H} \end{vmatrix}, \quad \bar{\mathbf{y}} = \begin{vmatrix} \mathbf{y} \\ \mathbf{w} \end{vmatrix}, \quad \bar{\mathbf{P}}^{-1} = \begin{vmatrix} \mathbf{P}^{-1} & \mathbf{0} \\ \mathbf{0} & c\mathbf{I} \end{vmatrix}, \quad c \neq 0, \quad c > 0. \quad (327.27)$$

Aus (322.9) folgen die Schätzwerte $\hat{\boldsymbol{\beta}}$ mit

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{P}\mathbf{X} + \mathbf{H}'\mathbf{H}/c)^{-1}(\mathbf{X}'\mathbf{P}\mathbf{y} + \mathbf{H}'\mathbf{w}/c). \quad (327.28)$$

Für einen sehr kleinen Wert für c , der sehr kleine Varianzen für \mathbf{w} bedingt, gleicht $\hat{\boldsymbol{\beta}}$ aus (327.28) näherungsweise $\tilde{\boldsymbol{\beta}}$ aus (327.8), wie sich aus dem folgenden Grenzprozeß ergibt. Mit der Identität (134.6), die für den zweiten Summanden zweimal angewendet wird, und $\mathbf{X}'\mathbf{P}\mathbf{X} = \mathbf{G}\mathbf{G}'$ wegen (143.5) sowie $(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1} = (\mathbf{G}')^{-1}\mathbf{G}^{-1}$ wegen (131.14) erhält man anstelle von (327.28)

$$\hat{\boldsymbol{\beta}} = [(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{H}'(c\mathbf{I} + \mathbf{H}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H} \\ (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{y} + (\mathbf{G}')^{-1}(c\mathbf{I} + \mathbf{G}^{-1}\mathbf{H}'\mathbf{H}(\mathbf{G}')^{-1})^{-1}\mathbf{G}^{-1}\mathbf{H}'\mathbf{w}].$$

Weiter folgt mit (153.21) und (327.8)

$$\lim_{c \rightarrow 0} \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}[\mathbf{X}'\mathbf{P}\mathbf{y} - \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H} \\ (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{y}] + (\mathbf{G}')^{-1}(\mathbf{H}(\mathbf{G}')^{-1})^+\mathbf{w} = \tilde{\boldsymbol{\beta}}. \quad (327.29)$$

wegen $(\mathbf{H}(\mathbf{G}')^{-1})^+ = \mathbf{G}^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{H}')^{-1}$ aus (153.20).

Die Konstante c in (327.28) ist also so klein zu wählen, daß die Differenz von $\hat{\boldsymbol{\beta}}$ aus (327.28) und $\tilde{\boldsymbol{\beta}}$ aus (327.8) vernachlässigbar klein wird, was numerisch

mit Hilfe der Residuen für die Restriktionen überprüft werden kann. Der Wert für c darf aber nicht so klein gesetzt werden, daß wegen der beschränkten Rechengenauigkeit jeder Numerik die Elemente von $\mathbf{X}'\mathbf{P}\mathbf{X}$ und $\mathbf{X}'\mathbf{P}\mathbf{y}$ bei der Addition in (327.28) vernachlässigt werden.

Ein Beispiel zum Gauß-Markoff-Modell mit Restriktionen befindet sich im Kapitel 341.

328 Rekursive Parameterschätzung

Die Beobachtungen im Gauß-Markoff-Modell (321.1) sollen nun in die unkorrelierten Vektoren \mathbf{y}_{m-1} und \mathbf{y}_m mit $m \in \{2, \dots, l\}$ zerfallen, die beispielsweise durch Messungen zu verschiedenen Zeiten gewonnen seien. Man will aber nicht warten, bis sämtliche Beobachtungen vorliegen, um dann die unbekannt Parameter zu schätzen, sondern die Schätzwerte sollen jeweils aufgrund der vorliegenden Beobachtungen berechnet werden. In (321.1) sei

$$\mathbf{X} = \begin{vmatrix} \mathbf{X}_{m-1} \\ \mathbf{X}_m \end{vmatrix}, \quad \mathbf{y} = \begin{vmatrix} \mathbf{y}_{m-1} \\ \mathbf{y}_m \end{vmatrix} \quad \text{und} \quad D(\mathbf{y}) = \sigma^2 \begin{vmatrix} \mathbf{P}_{m-1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_m^{-1} \end{vmatrix},$$

wobei zunächst $\text{rg}\mathbf{X} = \text{rg}\mathbf{X}_{m-1} = u$ angenommen werde und die Kovarianzmatrix $D(\mathbf{y})$ positiv definit sei, so daß \mathbf{P}_{m-1}^{-1} und \mathbf{P}_m^{-1} und wegen (143.9) auch ihre Inversen positiv definit sind, was sich durch die entsprechende Wahl des Vektors der quadratischen Form (143.1) zeigen läßt. Mit (322.9) folgt dann

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'_{m-1}\mathbf{P}_{m-1}\mathbf{X}_{m-1} + \mathbf{X}'_m\mathbf{P}_m\mathbf{X}_m)^{-1} (\mathbf{X}'_{m-1}\mathbf{P}_{m-1}\mathbf{y}_{m-1} + \mathbf{X}'_m\mathbf{P}_m\mathbf{y}_m). \quad (328.1)$$

Bezeichnet man mit $\hat{\boldsymbol{\beta}}_{m-1}$ die Schätzwerte der Parameter aus den Beobachtungen \mathbf{y}_{m-1} und mit $D(\hat{\boldsymbol{\beta}}_{m-1}) = \sigma^2\boldsymbol{\Sigma}_{m-1}$ ihre Kovarianzmatrix, erhält man mit (322.9) anstelle von (328.1), falls $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_m$ gesetzt wird, da $\hat{\boldsymbol{\beta}}_m$ durch Hinzunahme der Beobachtungen \mathbf{y}_m geschätzt wird,

$$\hat{\boldsymbol{\beta}}_m = (\boldsymbol{\Sigma}_{m-1}^{-1} + \mathbf{X}'_m\mathbf{P}_m\mathbf{X}_m)^{-1} (\boldsymbol{\Sigma}_{m-1}^{-1}\hat{\boldsymbol{\beta}}_{m-1} + \mathbf{X}'_m\mathbf{P}_m\mathbf{y}_m). \quad (328.2)$$

Mit $D(\hat{\boldsymbol{\beta}}_m) = \sigma^2\boldsymbol{\Sigma}_m$ gleicht wegen (322.9) die Matrix $\boldsymbol{\Sigma}_m$ der inversen Normalgleichungsmatrix in (328.2). Mit (134.6) folgt für sie

$$\boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}_{m-1} - \mathbf{F}_m\mathbf{X}_m\boldsymbol{\Sigma}_{m-1} \quad (328.3)$$

mit

$$\mathbf{F}_m = \boldsymbol{\Sigma}_{m-1}\mathbf{X}'_m(\mathbf{P}_m^{-1} + \mathbf{X}_m\boldsymbol{\Sigma}_{m-1}\mathbf{X}'_m)^{-1}. \quad (328.4)$$

Setzt man dieses Ergebnis in (328.2) ein, ergibt sich mit (134.7)

$$\hat{\boldsymbol{\beta}}_m = \hat{\boldsymbol{\beta}}_{m-1} + \mathbf{F}_m(\mathbf{y}_m - \mathbf{X}_m \hat{\boldsymbol{\beta}}_{m-1}) . \quad (328.5)$$

Aus (328.5) erhält man durch Hinzunahme der Beobachtungen \mathbf{y}_m die Schätzwerte $\hat{\boldsymbol{\beta}}_m$ rekursiv aus $\hat{\boldsymbol{\beta}}_{m-1}$ und aus (328.3) die rekursive Berechnung von $\boldsymbol{\Sigma}_m$ aus $\boldsymbol{\Sigma}_{m-1}$, wobei $D(\hat{\boldsymbol{\beta}}_m) = \sigma^2 \boldsymbol{\Sigma}_m$ gilt. Wie aus (328.1) ersichtlich, beginnt die Rekursion mit

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}'_1 \mathbf{P}_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{P}_1 \mathbf{y}_1 \quad \text{und} \quad \boldsymbol{\Sigma}_1 = (\mathbf{X}'_1 \mathbf{P}_1 \mathbf{X}_1)^{-1} , \quad (328.6)$$

so daß $\text{rg} \mathbf{X}_1 = u$ gelten muß. Für (328.5) ist wegen (328.4) eine Matrix zu invertieren, deren Dimensionen durch die Anzahl der Komponenten von \mathbf{y}_m bestimmt werden, während die Schätzwerte $\hat{\boldsymbol{\beta}}_m$ bei nicht rekursiver Berechnung mit der $u \times u$ Normalgleichungsmatrix in (328.2) folgen.

Nimmt man an, daß die unbekannt Parameter die Zustandsvariablen eines dynamischen Systems darstellen, die sich vom Zeitpunkt t_{m-1} zum Zeitpunkt t_m linear unter Addition von Störungen transformieren, erhält man aus (328.3) und (328.5) die Gleichungen des *Kalman–Bucy–Filters* (BRAMMER und SIFFLING 1975, S.82; BUCY und JOSEPH 1968, S.140; KOCH 1982A; KOCH 1990, S.92; KOCH 2000, S.108).

329 Abweichungen vom Modell

Die Parameterschätzung im Gauß–Markoff–Modell (321.1) oder (321.5) beruht auf der Voraussetzung, daß sich die Erwartungswerte der Beobachtungen mit der Matrix der gegebenen Koeffizienten als Linearkombinationen der Parameter darstellen lassen und daß die Gewichtsmatrix der Beobachtungen bekannt ist. Im folgenden sollen drei Fälle betrachtet werden, unter denen diese Voraussetzungen verletzt sind, und zwar soll angenommen werden, daß einmal zu wenige und zum anderen zu viele unbekannte Parameter eingeführt werden und daß eine fehlerhafte Gewichtsmatrix der Beobachtungen vorliegt.

Zur Untersuchung der beiden zuerst genannten Fälle sei das Modell gegeben

$$|\mathbf{X}_1, \mathbf{X}_2| \begin{vmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{vmatrix} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I} , \quad (329.1)$$

worin \mathbf{X}_1 eine $n \times k$ Matrix, \mathbf{X}_2 eine $n \times (u-k)$ Matrix, $\boldsymbol{\beta}_1$ ein $k \times 1$ Vektor und $\boldsymbol{\beta}_2$ ein $(u-k) \times 1$ Vektor bedeuten. Es gelte $\text{rg} |\mathbf{X}_1, \mathbf{X}_2| = u$. Die Schätzwerte $\hat{\boldsymbol{\beta}}_1$ und $\hat{\boldsymbol{\beta}}_2$ ergeben sich dann aus (322.9) zu

$$\begin{vmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{vmatrix} = \begin{vmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{X}_2 \\ \mathbf{X}'_2 \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{X}_2 \end{vmatrix}^{-1} \begin{vmatrix} \mathbf{X}'_1 \mathbf{y} \\ \mathbf{X}'_2 \mathbf{y} \end{vmatrix} . \quad (329.2)$$

Mit $N = \mathbf{X}'_1 \mathbf{X}_1$ und $M = \mathbf{X}'_2 \mathbf{X}_2 - \mathbf{X}'_2 \mathbf{X}_1 N^{-1} \mathbf{X}'_1 \mathbf{X}_2$ folgt wegen (134.3)

$$\begin{vmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{X}_2 \\ \mathbf{X}'_2 \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{X}_2 \end{vmatrix}^{-1} = \begin{vmatrix} N^{-1} + N^{-1} \mathbf{X}'_1 \mathbf{X}_2 M^{-1} \mathbf{X}'_2 \mathbf{X}_1 N^{-1} & -N^{-1} \mathbf{X}'_1 \mathbf{X}_2 M^{-1} \\ -M^{-1} \mathbf{X}'_2 \mathbf{X}_1 N^{-1} & M^{-1} \end{vmatrix}. \quad (329.3)$$

Wegen $\text{rg}[\mathbf{X}_1, \mathbf{X}_2] = u$ sind nach (143.8) und (143.9) die Matrizen auf der rechten und linken Seite von (329.3) positiv definit. Dann sind auch, wie sich durch entsprechende Wahl des Vektors in der quadratischen Form (143.1) zeigen läßt, die Matrizen N^{-1} und M^{-1} positiv definit und nach (143.7) die Matrix $N^{-1} \mathbf{X}'_1 \mathbf{X}_2 M^{-1} \mathbf{X}'_2 \mathbf{X}_1 N^{-1}$ positiv definit oder positiv semidefinit.

a) Anzahl der Parameter zu klein

Die Parameterschätzung sei in dem Modell $\mathbf{X}_1 \boldsymbol{\beta}_1 = E(\mathbf{y})$ mit $D(\mathbf{y}) = \sigma^2 \mathbf{I}$ vorgenommen, so daß mit (322.9) sich $\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{y}$ ergibt. Tatsächlich gelte aber das Modell (329.1). Die Schätzwerte $\hat{\boldsymbol{\beta}}_1$ sind dann im allgemeinen verzerrt, also nicht erwartungstreu, denn mit (329.1) folgt $E(\hat{\boldsymbol{\beta}}_1) = (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 (\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2)$ oder, sofern nicht $\mathbf{X}'_1 \mathbf{X}_2 = \mathbf{0}$ gilt,

$$E(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1 \mathbf{X}_2 \boldsymbol{\beta}_2. \quad (329.4)$$

Die Varianzen der Schätzung in $D(\hat{\boldsymbol{\beta}}_1) = \sigma^2 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} = \sigma^2 N^{-1}$ ergeben im allgemeinen zu kleine Werte, da, wie aus (329.3) ersichtlich, die positiv definite oder positiv semidefinite Matrix $N^{-1} \mathbf{X}'_1 \mathbf{X}_2 M^{-1} \mathbf{X}'_2 \mathbf{X}_1 N^{-1}$ addiert werden müßte, um die für $\hat{\boldsymbol{\beta}}_1$ gültige Kovarianzmatrix zu erhalten. Im allgemeinen verzerrt ergibt sich auch die Schätzung der Varianz der Gewichtseinheit, denn mit $\hat{\mathbf{e}} = \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 - \mathbf{y}$ erhält man aus (271.1), (325.5) und (329.1)

$$\begin{aligned} E(\hat{\mathbf{e}}' \hat{\mathbf{e}}) &= \sigma^2 (n - k) + (\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2)' (\mathbf{I} - \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1) \\ &\quad (\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2) \\ &= \sigma^2 (n - k) + \boldsymbol{\beta}'_2 \mathbf{X}'_2 (\mathbf{I} - \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1) \mathbf{X}_2 \boldsymbol{\beta}_2 \\ &\geq \sigma^2 (n - k), \end{aligned} \quad (329.5)$$

da die Matrix $\mathbf{I} - \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{X}_1)^{-1} \mathbf{X}'_1$, wie bereits bei (323.11) erwähnt, idempotent und symmetrisch und daher wegen (152.8) positiv semidefinit ist.

Ist die Kovarianzmatrix der Beobachtungen bekannt und wird $\sigma^2 = 1$ gesetzt, kann wegen (329.5) ein Schätzwert $\hat{\sigma}^2 > 1$ auf zu wenige Parameter im Modell hindeuten.

b) Anzahl der Parameter zu groß

Die unbekannt Parameter seien in dem Modell (329.1) geschätzt, tatsächlich gelte aber das Modell

$$\mathbf{X}_1\boldsymbol{\beta}_1 = E(\mathbf{y}) \quad \text{und} \quad D(\mathbf{y}) = \sigma^2\mathbf{I} . \quad (329.6)$$

Die Schätzwerte $\hat{\boldsymbol{\beta}}_1$ aus (329.2) ergeben sich erwartungstreu, denn mit (329.3) und (329.6) erhält man

$$E(\hat{\boldsymbol{\beta}}_1) = |\mathbf{N}^{-1} + \mathbf{N}^{-1}\mathbf{X}'_1\mathbf{X}_2\mathbf{M}^{-1}\mathbf{X}'_2\mathbf{X}_1\mathbf{N}^{-1}, -\mathbf{N}^{-1}\mathbf{X}'_1\mathbf{X}_2\mathbf{M}^{-1}| \begin{vmatrix} \mathbf{X}'_1\mathbf{X}_1\boldsymbol{\beta}_1 \\ \mathbf{X}'_2\mathbf{X}_1\boldsymbol{\beta}_1 \end{vmatrix} = \boldsymbol{\beta}_1 . \quad (329.7)$$

Die Varianzen in $D(\hat{\boldsymbol{\beta}}_1)$ aus (329.3) ergeben im allgemeinen zu große Werte, denn die positiv definite oder positiv semidefinite Matrix $\mathbf{N}^{-1}\mathbf{X}'_1\mathbf{X}_2\mathbf{M}^{-1}\mathbf{X}'_2\mathbf{X}_1\mathbf{N}^{-1}$ müßte subtrahiert werden, um die für $\hat{\boldsymbol{\beta}}_1$ gültige Kovarianzmatrix $D(\hat{\boldsymbol{\beta}}_1) = \sigma^2\mathbf{N}^{-1}$ zu erhalten. Setzt man $\mathbf{X} = |\mathbf{X}_1, \mathbf{X}_2|$, ergibt sich mit (271.1), (325.5) und (329.6) für den Erwartungswert der Quadratsumme der Residuen

$$E(\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}) = \sigma^2(n - u) + \boldsymbol{\beta}'_1\mathbf{X}'_1(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X}_1\boldsymbol{\beta}_1 = \sigma^2(n - u) , \quad (329.8)$$

da $(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')|\mathbf{X}_1, \mathbf{X}_2| = \mathbf{0}$ gilt. Die Varianz σ^2 der Gewichtseinheit wird also erwartungstreu geschätzt.

c) Fehlerhafte Gewichtsmatrix der Beobachtungen

Für die Parameterschätzung sei das Modell $\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y})$ mit $D(\mathbf{y}) = \sigma^2\mathbf{V}^{-1}$ angenommen, tatsächlich gelte aber das Modell

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2\mathbf{P}^{-1} . \quad (329.9)$$

Mit $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{y}$ aus (322.9) ergibt sich eine erwartungstreue Schätzung von $\boldsymbol{\beta}$, denn es gilt $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. Verzerrt erhält man aber die Schätzung der Varianz der Gewichtseinheit, denn mit (271.1) folgt aus (325.2) und (329.9)

$$E(\hat{\boldsymbol{\epsilon}}'\mathbf{V}\hat{\boldsymbol{\epsilon}}) = \sigma^2\text{sp}[\mathbf{P}^{-1}(\mathbf{V} - \mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V})] \neq \sigma^2(n - u) . \quad (329.10)$$

Um die Auswirkung einer fehlerhaften Gewichtsmatrix der Beobachtungen auf die Parameterschätzung zu berechnen, wird angenommen, daß

$$\mathbf{P} = \mathbf{V} + \boldsymbol{\Delta}\mathbf{V} \quad (329.11)$$

gelte, wobei ΔV wie V positiv definit seien. Für die Schätzung

$$\hat{\beta} = (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}\mathbf{y} = (\mathbf{X}'\mathbf{V}\mathbf{X} + \mathbf{X}'\Delta\mathbf{V}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}\mathbf{y} + \mathbf{X}'\Delta\mathbf{V}\mathbf{y}) \quad (329.12)$$

ergibt sich die inverse Normalgleichungsmatrix mit (134.6) zu

$$\begin{aligned} (\mathbf{X}'\mathbf{V}\mathbf{X} + \mathbf{X}'\Delta\mathbf{V}\mathbf{X})^{-1} &= (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}[\mathbf{I} - \mathbf{X}'(\Delta\mathbf{V}^{-1} + \\ &\quad \mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}')^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}[\mathbf{I} - \mathbf{X}'\Delta\mathbf{V}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X} + \\ &\quad \mathbf{X}'\Delta\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\Delta\mathbf{V})\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}]. \end{aligned} \quad (329.13)$$

Sind die Elemente von ΔV so klein, daß Matrizenprodukte, in denen ΔV zweimal erscheint, vernachlässigt werden können, erhält man

$$(\mathbf{X}'\mathbf{V}\mathbf{X} + \mathbf{X}'\Delta\mathbf{V}\mathbf{X})^{-1} \approx (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}(\mathbf{I} - \mathbf{X}'\Delta\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}) \quad (329.14)$$

und damit anstelle von (329.12) die Schätzung

$$\hat{\beta} \approx (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}\mathbf{y} - \mathbf{X}'\Delta\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{y} + \mathbf{X}'\Delta\mathbf{V}\mathbf{y}). \quad (329.15)$$

33 Gauß–Markoff–Modell mit nicht vollem Rang

331 Methode der kleinsten Quadrate und Maximum–Likelihood–Schätzung

In dem Gauß–Markoff–Modell (321.1) wird eine Koeffizientenmatrix \mathbf{X} mit vollem Spaltenrang vorausgesetzt. Im folgenden soll \mathbf{X} einen Rangdefekt aufweisen, doch wird die Forderung nach einer positiv definiten Kovarianzmatrix der Beobachtungen aufrecht erhalten. Modelle mit singulärer Kovarianzmatrix werden in (RAO 1973, S.297; SCHAFFRIN 1983, S.28; SEARLE 1971, S.221) behandelt.

Definition: Gilt in dem Gauß–Markoff–Modell (321.1) oder (321.5) $\text{rg}\mathbf{X} = q < u$, bezeichnet man es als Gauß–Markoff–Modell mit *nicht vollem Rang*. (331.1)

Zur Vereinfachung der Ableitungen sollen wie im Kapitel 32 die Parameterschätzungen im Modell (321.5) mit nicht vollem Rang vorgenommen werden, das durch die lineare Transformation (321.4) aus dem Modell (321.1) entsteht. Das folgende Modell wird also zugrunde gelegt

$$\mathbf{X}\beta = E(\mathbf{y}) \quad \text{mit} \quad \text{rg}\mathbf{X} = q < u \quad \text{und} \quad D(\mathbf{y}) = \sigma^2\mathbf{I}, \quad (331.2)$$

in dem wieder \mathbf{X} eine $n \times u$ Matrix gegebener Koeffizienten, $\boldsymbol{\beta}$ ein $u \times 1$ Vektor unbekannter Parameter, \mathbf{y} ein $n \times 1$ Zufallsvektor von Beobachtungen und σ^2 die Varianz der Gewichtseinheit bedeuten.

In diesem Modell führt die Methode (313.1) der kleinsten Quadrate nach (323.1) auf die Normalgleichungen $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ und mit $\mathbf{z} = \mathbf{0}$ in (154.9) sowie mit (233.2) auf

$$\bar{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} \quad \text{und} \quad D(\bar{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}((\mathbf{X}'\mathbf{X})^{-})', \quad (331.3)$$

worin $\bar{\boldsymbol{\beta}}$ eine Schätzung von $\boldsymbol{\beta}$, $D(\bar{\boldsymbol{\beta}})$ ihre Kovarianzmatrix und $(\mathbf{X}'\mathbf{X})^{-}$ eine mit (153.1) definierte generalisierte Inverse von $\mathbf{X}'\mathbf{X}$ bedeuten. Weder $\bar{\boldsymbol{\beta}}$ noch $D(\bar{\boldsymbol{\beta}})$ sind eindeutig, da nach (153.2) $(\mathbf{X}'\mathbf{X})^{-}$ nicht eindeutig ist. Die Schätzung ist auch nicht erwartungstreu, denn mit (331.2) folgt $E(\bar{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \neq \boldsymbol{\beta}$, falls $\boldsymbol{\beta} \neq \mathbf{0}$ und kein Eigenvektor der Matrix $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}$ ist, deren Eigenwerte wegen (152.2) und (153.4) Null oder Eins betragen.

Die Schätzwerte $\bar{\boldsymbol{\beta}}$ minimieren die Quadratsumme $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/\sigma^2$, denn für einen beliebigen Vektor $\boldsymbol{\beta}^*$ ergibt sich anstelle von (323.2) $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*) \geq (\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}})$, da $\mathbf{X}'\mathbf{X}$ wegen (143.8) positiv semidefinit ist und da $(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)'\mathbf{X}'(\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}) = (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)'(\mathbf{X}' - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}')\mathbf{y} = 0$ wegen (153.5) gilt. Eine identische Schätzung erhält man mit der Maximum–Likelihood–Methode (314.2) im Falle normalverteilter Beobachtungen wegen (324.2). Somit folgt der

Satz: Im Gauß–Markoff–Modell (331.2) mit nicht vollem Rang liefern die Methode der kleinsten Quadrate und die Maximum–Likelihood–Methode im Falle normalverteilter Beobachtungen identische Schätzwerte $\bar{\boldsymbol{\beta}}$, die in (331.3) angegeben sind, für die Parameter $\boldsymbol{\beta}$. (331.4)

Die Quadratsumme $(\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}})$ gleicht der Quadratsumme Ω der (323.4) entsprechend definierten Residuen $\hat{\mathbf{e}}$ mit

$$\Omega = \hat{\mathbf{e}}'\hat{\mathbf{e}} \quad \text{und} \quad \hat{\mathbf{e}} = \mathbf{X}\bar{\boldsymbol{\beta}} - \mathbf{y}. \quad (331.5)$$

Die Residuen sind eindeutig, da mit $\hat{\mathbf{e}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} - \mathbf{y}$ nach (153.8) die Matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ invariant gegenüber der Wahl von $(\mathbf{X}'\mathbf{X})^{-}$ ist. Da $\hat{\mathbf{e}}$ eindeutig ist, ist auch Ω eindeutig. Zusammenfassend gilt daher der

Satz: Die Quadratsumme Ω der Residuen ist eindeutig bestimmt und minimal. (331.6)

Wie sich mit (153.5) zeigen läßt, ist die Matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ idempotent, so daß mit (152.3) und (153.4) $\text{rg}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}') = \text{sp}((\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}) = q$ folgt. Nach (152.4) ist dann auch $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ idempotent und $\text{rg}(\mathbf{I} -$

$\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = n - q$, so daß man mit (233.2), (331.2) und (331.5) die eindeutige, aber singuläre Kovarianzmatrix $D(\hat{\boldsymbol{\epsilon}})$ der Residuen erhält

$$D(\hat{\boldsymbol{\epsilon}}) = \sigma^2(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ \text{mit } \text{rg}D(\hat{\boldsymbol{\epsilon}}) = n - q \quad \text{und} \quad \text{sp}D(\hat{\boldsymbol{\epsilon}}) = \sigma^2(n - q). \quad (331.7)$$

Entsprechend (324.5) ergibt sich die Maximum-Likelihood-Schätzung $\bar{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit zu

$$\bar{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\bar{\boldsymbol{\beta}}) = \frac{1}{n}\Omega. \quad (331.8)$$

Um die erwartungstreue Schätzung von σ^2 zu erhalten, wird mit (331.5) anstelle von (325.2) gebildet

$$\Omega = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}, \quad (331.9)$$

denn die Matrix $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ ist idempotent und außerdem symmetrisch, so daß mit (152.8) folgt

$$\Omega \geq 0. \quad (331.10)$$

Der Erwartungswert $E(\Omega)$ berechnet sich mit (271.1), (331.2) und (331.7) zu $E(\Omega) = \sigma^2(n - q) + \boldsymbol{\beta}'(\mathbf{X}' - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X}\boldsymbol{\beta}$ und mit (153.5) zu

$$E(\Omega) = \sigma^2(n - q), \quad (331.11)$$

so daß sich als erwartungstreue Schätzung $\hat{\sigma}^2$ von σ^2 im Modell (331.2) mit nicht vollem Rang ergibt

$$\hat{\sigma}^2 = \Omega/(n - q). \quad (331.12)$$

Diese Schätzung ist unter der Voraussetzung normalverteilter Beobachtungen auch beste Schätzung im Sinne von (312.2), wie mit (364.5) gezeigt wird.

332 Schätzbare Funktionen

Da die Parameter $\boldsymbol{\beta}$ im Modell mit nicht vollem Rang wegen (331.3) nicht erwartungstreu schätzbar sind, wird jetzt, da ein lineares Modell vorliegt, nach linearen Funktionen der Parameter gesucht, deren Schätzwerte die Bedingung der Erwartungstreue erfüllen. Entsprechend (322.1) ergibt sich die

Definition: Eine lineare Funktion $\mathbf{a}'\boldsymbol{\beta}$ der Parameter $\boldsymbol{\beta}$ im Modell (331.2),

wobei \mathbf{a} ein gegebener $u \times 1$ Vektor bedeutet, bezeichnet man als *erwartungstreu schätzbar* oder als *schätzbare* Funktion, falls ein $n \times 1$ Vektor \mathbf{c} derart existiert, daß für alle $\boldsymbol{\beta}$ gilt

$$E(\mathbf{c}'\mathbf{y}) = \mathbf{a}'\boldsymbol{\beta} . \quad (332.1)$$

Ist $\mathbf{a}'\boldsymbol{\beta}$ eine schätzbare Funktion, folgt wegen (331.2) $\mathbf{c}'\mathbf{X}\boldsymbol{\beta} = \mathbf{a}'\boldsymbol{\beta}$ für alle $\boldsymbol{\beta}$ und in Übereinstimmung mit (322.2)

$$\mathbf{c}'\mathbf{X} = \mathbf{a}' . \quad (332.2)$$

Zur Prüfung, ob eine schätzbare Funktion vorliegt, dient der

Satz: Die lineare Funktion $\mathbf{a}'\boldsymbol{\beta}$ der Parameter $\boldsymbol{\beta}$ ist genau dann erwartungstreu schätzbar, wenn gilt

$$\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{a}' . \quad (332.3)$$

Beweis: Ist $\mathbf{a}'\boldsymbol{\beta}$ für ein vorgegebenes \mathbf{a} eine schätzbare Funktion, dann folgt mit (332.2) $\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{c}'\mathbf{X} = \mathbf{a}'$ wegen (153.5). Gilt andererseits $\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{a}'$, folgt mit $\mathbf{c}' = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ die Bedingung (332.2), so daß $\mathbf{a}'\boldsymbol{\beta}$ schätzbar ist, und damit die Aussage.

Aus (332.3) ergibt sich unmittelbar, daß im Gauß–Markoff–Modell (321.5) mit vollem Rang, in dem nach (153.22) $(\mathbf{X}'\mathbf{X})^{-} = (\mathbf{X}'\mathbf{X})^{-1}$ gilt, sämtliche lineare Funktionen der Parameter erwartungstreu schätzbar sind. Dieses Ergebnis stimmt mit (322.8) überein. Mit (332.3) läßt sich weiter die Schätzbarkeit der folgenden Funktionen beweisen.

Satz: Schätzbare Funktionen im Modell (331.2) sind

a) der Erwartungswertvektor $E(\mathbf{y})$ der Beobachtungen \mathbf{y} , (332.4)

b) die linearen Funktionen schätzbarer Größen, (332.5)

c) die projizierten Parameter $\boldsymbol{\beta}_b$ mit $\boldsymbol{\beta}_b = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$. (332.6)

Beweis: Mit $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ aus (331.2) und $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{X}$ aus (153.5) ist für jede Zeile von \mathbf{X} (332.3) erfüllt, so daß (332.4) folgt. Stellen $\mathbf{D}\boldsymbol{\beta}$ schätzbare Funktionen dar, so daß $\mathbf{D}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{D}$ gilt, so ist auch $\mathbf{d}'\mathbf{D}\boldsymbol{\beta}$ eine schätzbare Funktion, da $\mathbf{d}'\mathbf{D}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{d}'\mathbf{D}$ gilt, so daß (332.5) folgt. Mit $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}$ wegen (153.1) folgt mit (332.3) schließlich (332.6) und mit (161.2), daß $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}$ ein Projektionsoperator ist.

Für schätzbare Funktionen gilt (322.8) entsprechend der

Satz: Die beste lineare erwartungstreue und eindeutige Schätzung $\hat{\alpha}$ und ihre eindeutige Varianz $V(\hat{\alpha})$ der schätzbaren Funktion $\alpha = \mathbf{a}'\boldsymbol{\beta}$ der unbekannt Parameter $\boldsymbol{\beta}$ im Gauß–Markoff–Modell (331.2) ist gegeben durch

$$\hat{\alpha} = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} \quad \text{und} \quad V(\hat{\alpha}) = \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}. \quad (332.7)$$

Beweis: Erwartungstreue erhält man mit $E(\hat{\alpha}) = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{a}'\boldsymbol{\beta} = \alpha$ wegen (332.3), da $\alpha = \mathbf{a}'\boldsymbol{\beta}$ eine schätzbare Funktion ist. Setzt man $\hat{\alpha} = \mathbf{c}'\mathbf{y}$ mit $\mathbf{c}' = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ ergibt sich (332.2) wegen (332.3). Hiermit folgt $\hat{\alpha} = \mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ und daraus die Eindeutigkeit der Schätzung, da nach (153.8) $\hat{\alpha}$ invariant gegenüber der Wahl der generalisierten Inversen $(\mathbf{X}'\mathbf{X})^{-}$ und da $\mathbf{c}'\mathbf{X}$ eindeutig ist. Ferner ergibt sich die Varianz $V(\hat{\alpha})$ mit (233.2) und (331.2) zu $V(\hat{\alpha}) = \sigma^2\mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}((\mathbf{X}'\mathbf{X})^{-})'\mathbf{X}'\mathbf{c}$ und mit (153.5), (153.7) und (332.2) zu $V(\hat{\alpha}) = \sigma^2\mathbf{c}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{c} = \sigma^2\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{a}$. Wegen (153.8) ist $V(\hat{\alpha})$ invariant gegenüber der Wahl für $(\mathbf{X}'\mathbf{X})^{-}$, so daß $V(\hat{\alpha})$ eindeutig ist. Die Schätzung $\hat{\alpha}$ ist auch beste Schätzung im Sinne von (312.2), denn bezeichnet man mit $\mathbf{c}^{*\prime}\mathbf{y}$ eine beliebige erwartungstreue Schätzung von α , folgt mit $E(\mathbf{c}^{*\prime}\mathbf{y}) = \mathbf{c}^{*\prime}\mathbf{X}\boldsymbol{\beta} = \mathbf{a}'\boldsymbol{\beta}$ für alle $\boldsymbol{\beta}$ die Bedingung $\mathbf{c}^{*\prime}\mathbf{X} = \mathbf{a}'$. Hiermit geht man analog zu (322.6) vor und erhält mit $\mathbf{c}' = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ anstelle von (322.7) $f = \sigma^2(\mathbf{c}^{*\prime} - \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}')\mathbf{X}((\mathbf{X}'\mathbf{X})^{-})'\mathbf{a} = 0$ wegen (332.3), so daß die Aussagen folgen.

333 Projizierte Parameter als schätzbare Funktion

Eine lineare Transformation der Parameter $\boldsymbol{\beta}$ in (331.2) in schätzbare Funktionen bezeichnet man als *Reparameterisierung* (BOCK 1975, S.239; GRAFAREND und SCHAFFRIN 1976; HUMAK 1977, S.51). Die Beziehung der so gewonnenen schätzbaren Funktionen zu den Parametern läßt sich dabei häufig nur schwer interpretieren, so daß hier die schätzbaren Funktionen durch Projektionen der Parameter erhalten werden, wodurch die Beziehung zwischen den ursprünglichen Parametern und den projizierten Parametern mit Hilfe von Restriktionen zu erklären ist (KOCH 1978), wie im folgenden gezeigt wird.

Nach (332.6) stellt der Vektor $\boldsymbol{\beta}_b$

$$\boldsymbol{\beta}_b = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \quad (333.1)$$

eine schätzbare Funktion dar. Da die Matrix $(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}$ nach (153.4) idempotent ist und $\text{rg}((\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}) = q$ gilt, stellt sie nach (161.2) einen Projektionsoperator dar, so daß sich die schätzbare Funktion $\boldsymbol{\beta}_b$ als Projektion des u -dimensionalen Euklidischen Raums E^u mit $\boldsymbol{\beta} \in E^u$ auf den q -dimensionalen Unterraum E^q mit $\boldsymbol{\beta}_b \in E^q$ ergibt. Als beste lineare erwar-

tungstreue Schätzung $\hat{\beta}_b$ von β_b erhält man mit (153.5) aus (332.7)

$$\hat{\beta}_b = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$$

mit der Kovarianzmatrix wegen (233.2)

$$D(\hat{\beta}_b) = \sigma^2(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}((\mathbf{X}'\mathbf{X})^{-})'.$$

Um die Analogie mit der Schätzung (322.9) im Modell mit vollem Rang zu erhalten, bei der die Kovarianzmatrix aus der inversen Normalgleichungsmatrix folgt, wird die Matrix

$$(\mathbf{X}'\mathbf{X})_{rs}^{-} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}((\mathbf{X}'\mathbf{X})^{-})' \quad (333.2)$$

eingeführt. $(\mathbf{X}'\mathbf{X})_{rs}^{-}$ ist eine symmetrische reflexive generalisierte Inverse von $\mathbf{X}'\mathbf{X}$, da (153.11) mit $\mathbf{X}' = \mathbf{X}'\mathbf{X}((\mathbf{X}'\mathbf{X})^{-})'\mathbf{X}'$ wegen (153.5) und (153.7) erfüllt ist. Aufgrund letzterer Beziehung wird die Projektion (333.1) umgeschrieben in

$$\beta_b = (\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{X}'\mathbf{X}\beta \quad (333.3)$$

und die beste lineare erwartungstreue Schätzung $\hat{\beta}_b$ in

$$\hat{\beta}_b = (\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{X}'\mathbf{y} \quad \text{mit} \quad D(\hat{\beta}_b) = \sigma^2(\mathbf{X}'\mathbf{X})_{rs}^{-}, \quad (333.4)$$

so daß nun die Kovarianzmatrix der Schätzung aus jener Inversen $(\mathbf{X}'\mathbf{X})_{rs}^{-}$ der Normalgleichungsmatrix folgt, die auch für die Schätzung und die Projektion benutzt wird.

Da eine generalisierte Inverse nicht eindeutig ist, sind beliebig viele Projektionen zur Erzeugung schätzbarer Funktionen nach (333.1) oder (333.3) möglich. Benutzt man anstelle von $(\mathbf{X}'\mathbf{X})_{rs}^{-}$ die nach (153.17) eindeutige Pseudoinverse $(\mathbf{X}'\mathbf{X})^{+}$ von $\mathbf{X}'\mathbf{X}$, die wegen (153.15) und (153.18) reflexiv und symmetrisch ist, ergibt sich der wegen (155.22) und (162.2) orthogonale Projektionsoperator $(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{X}$. Mit ihm folgen die schätzbaren Funktionen

$$\beta_e = (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{X}\beta \quad (333.5)$$

und analog zu (333.4) die Schätzwerte $\hat{\beta}_e$ sowie die Kovarianzmatrix $D(\hat{\beta}_e)$

$$\hat{\beta}_e = (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y} \quad \text{und} \quad D(\hat{\beta}_e) = \sigma^2(\mathbf{X}'\mathbf{X})^{+}. \quad (333.6)$$

Zweckmäßige Rechenformeln für $(\mathbf{X}'\mathbf{X})_{rs}^{-}$ und $(\mathbf{X}'\mathbf{X})^{+}$ bilden die mit der $(u - q) \times u$ Matrix \mathbf{B} in (155.13) bis (155.19) und die mit der $(u - q) \times u$ Matrix \mathbf{E} in (155.21) bis (155.25) angegebenen Beziehungen. Somit erhält

man die Schätzwerte $\hat{\beta}_b$ und $\hat{\beta}_e$ aus (333.4) und (333.6) mit (155.15) oder (155.17) wegen (155.11) zu

$$\begin{vmatrix} \hat{\beta}_b \\ \mathbf{k} \end{vmatrix} = \begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{B}' \\ \mathbf{B} & \mathbf{0} \end{vmatrix}^{-1} \begin{vmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{0} \end{vmatrix} \quad \text{oder} \quad \hat{\beta}_b = (\mathbf{X}'\mathbf{X} + \mathbf{B}'\mathbf{B})^{-1} \mathbf{X}'\mathbf{y} \quad (333.7)$$

und mit (155.21) oder (155.23) zu

$$\begin{vmatrix} \hat{\beta}_e \\ \mathbf{k} \end{vmatrix} = \begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{E}' \\ \mathbf{E} & \mathbf{0} \end{vmatrix}^{-1} \begin{vmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{0} \end{vmatrix} \quad \text{oder} \quad \hat{\beta}_e = (\mathbf{X}'\mathbf{X} + \mathbf{E}'\mathbf{E})^{-1} \mathbf{X}'\mathbf{y}, \quad (333.8)$$

worin \mathbf{k} ein $(u - q) \times 1$ Vektor bedeutet, der nicht weiter benötigt wird. Da die Matrix $\mathbf{X}'\mathbf{X}$ singulär ist, gilt für die Berechnung der Inversen der um \mathbf{B} oder \mathbf{E} erweiterten Matrix das bei (155.19) und (155.25) Gesagte. Mit (333.7) und (333.8) identische Schätzungsergebnisse erhält man nach (327.9) mit der Methode der kleinsten Quadrate in den Gauß–Markoff–Modellen mit speziellen Restriktionen

$$\mathbf{X}\beta = E(\mathbf{y}) \quad \text{mit} \quad \mathbf{B}\beta = \mathbf{0} \quad \text{und} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I} \quad (333.9)$$

oder

$$\mathbf{X}\beta = E(\mathbf{y}) \quad \text{mit} \quad \mathbf{E}\beta = \mathbf{0} \quad \text{und} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I}. \quad (333.10)$$

Zusammenfassend gilt der

Satz: Ersetzt man in den Schätzungen für das Gauß–Markoff–Modell mit vollem Rang die Inverse des Normalgleichungssystems durch eine symmetrische reflexive generalisierte Inverse oder durch die Pseudoinverse, die beispielsweise aus den Normalgleichungen der Methode der kleinsten Quadrate angewendet auf die Modelle (333.9) oder (333.10) mit speziellen Restriktionen erhalten werden können, ergeben sich die besten linearen erwartungstreuen Schätzungen (333.4) oder (333.6) der nach (333.3) oder (333.5) projizierten Parameter des Gauß–Markoff–Modells mit nicht vollem Rang. (333.11)

Die mit (333.3) und (333.5) vorgenommenen Projektionen lassen sich also als $u - q$ Restriktionen $\mathbf{B}\beta = \mathbf{0}$ oder $\mathbf{E}\beta = \mathbf{0}$ interpretieren, denen die im u -dimensionalen Raum E^u gegebenen Parameter β zu unterwerfen sind, um sie im q -dimensionalen Unterraum E^q erwartungstreu schätzen zu können. Man bezeichnet diese Bedingungen daher auch als Restriktionen zur *Identifizierbarkeit* der Parameter (SCHACH und SCHÄFER 1978, S.31; SEBER 1977, S.74; TOUTENBURG 1975, S.42). Projiziert man mit Hilfe der Pseudoinversen $(\mathbf{X}'\mathbf{X})^+$, werden die β_e derart geschätzt, daß ihre Kovarianzmatrix minimale Spur besitzt, da nach (156.1) $\text{sp}(\mathbf{X}'\mathbf{X})^+ < \text{sp}(\mathbf{X}'\mathbf{X})_{rs}^-$ gilt. Außerdem folgt

nach (156.2) für die Schätzwerte $\hat{\beta}_e$, daß ihre Quadratsumme $\hat{\beta}_e' \hat{\beta}_e$ minimal wird. Diese Eigenschaften sind nach (156.3) und (156.4) für eine Teilmenge der Parameter erfüllt, wenn man mit einer symmetrischen reflexiven generalisierten Inversen $(\mathbf{X}'\mathbf{X})_{rs}^-$ schätzt, die mit einer Matrix \mathbf{B} in (333.7) und (333.9) berechnet wird, in der die Spalten für die Teilmenge der Parameter aus der Matrix \mathbf{E} in (333.8) und (333.10) übernommen werden und die übrigen Spalten Nullelemente enthalten.

Die Matrix \mathbf{E}' , deren Spalten, wie bei (155.11) erwähnt, eine Basis für den Nullraum $N(\mathbf{X})$ der Koeffizientenmatrix \mathbf{X} bilden, läßt sich aus (155.10) berechnen oder auch unmittelbar angeben, denn durch den Vergleich von $\mathbf{X}\mathbf{E}' = \mathbf{0}$ aus (155.11) mit $\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y})$ aus (331.2) erkennt man, daß die Spalten von \mathbf{E}' die Änderungen enthalten, die die Parameter $\boldsymbol{\beta}$ vornehmen können, ohne daß sich die Erwartungswerte der Beobachtungen ändern. Dies soll an dem folgenden Beispiel erläutert werden.

Beispiel: Die Koordinaten von Punkten im dreidimensionalen Euklidischen Raum E^3 mit dem (x, y, z) -Koordinatensystem sollen aus den Beobachtungen der Entfernungen zwischen sämtlichen Punkten mit einem Entfernungsmeßgerät, dessen Maßstab unbekannt ist, bestimmt werden. Bei diesem Problem fehlt die Information über die Position des Netzes, das durch die Streckenmessungen aufgespannt wird, in dem (x, y, z) -Koordinatensystem, in dem die Koordinaten der Netzpunkte zu berechnen sind. Man spricht daher von einem *freien Netz*, dessen *Datum* zu definieren ist, dessen Lage und Orientierung also im Koordinatensystem festzulegen ist.

Bedeutet s_{ij} die vom Punkt i des Netzes mit den Koordinaten x_i, y_i, z_i zum Punkt j gemessene Strecke, e_{ij} den Fehler und m den Maßstabsfaktor der Streckenmessungen, ergibt sich die Beobachtungsgleichung

$$m[(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2]^{1/2} = s_{ij} + e_{ij}.$$

Mit Hilfe von Näherungswerten für die unbekannt Parameter wird linearisiert, so daß die Koeffizientenmatrix \mathbf{X} des Modells der Parameterschätzung aus (321.8) folgt. Bezeichnet man jetzt mit $\mathbf{x}_{i0} = |x_{i0}, y_{i0}, z_{i0}|'$ die im (x, y, z) -Koordinatensystem gegebenen Näherungswerte der Koordinaten des Punktes i , mit $\mathbf{x}_i = |x_i, y_i, z_i|'$ die unbekannt Korrekturen der Näherungskordinaten des Punktes i und mit m die unbekannt Korrektur des Näherungswertes m_0 für den Maßstabsfaktor, dann ist $\boldsymbol{\beta} = |\mathbf{x}'_1, \dots, \mathbf{x}'_i, \dots, m|'$ der Vektor der unbekannt Parameter und $\boldsymbol{\beta}_0 = |\mathbf{x}'_{10}, \dots, \mathbf{x}'_{i0}, \dots, m_0|'$ der Vektor seiner Näherungswerte.

Ohne daß sich der Erwartungswert einer Beobachtung ändert, können die gesuchten Koordinaten einer Translation, einer Rotation und einer Maßstabsän-

derung unterworfen werden, indem das durch die Entfernungsmessungen aufgespannte Netz in Richtung der drei Achsen des (x, y, z) -Koordinatensystems verschoben, um Parallelen zu den drei Achsen gedreht und im Maßstab geändert wird. Unter der Voraussetzung, daß das Netz mindestens drei Punkte umfaßt, ergibt sich daher der Rangdefekt der Koeffizientenmatrix \mathbf{X} zu $u - q = 7$.

Eine Translation des Netzes in Richtung der x -, y - und z -Achse um t_x , t_y und t_z ergibt die Koordinaten des Punktes i zu $\mathbf{x}_{i0} + \mathbf{x}_{ti} = |x_{i0} + x_i + t_x, y_{i0} + y_i + t_y, z_{i0} + z_i + t_z|'$, so daß der gesamte Vektor $\boldsymbol{\beta}_t$ der unbekannt Parameter nach der Translation folgt mit

$$\boldsymbol{\beta}_t = \boldsymbol{\beta} + \begin{vmatrix} \mathbf{I}_3 \\ \dots \\ \mathbf{I}_3 \\ \dots \\ \mathbf{0}' \end{vmatrix} \begin{vmatrix} t_x \\ t_y \\ t_z \end{vmatrix}, \quad (333.12)$$

wobei die Nullzeile deshalb einzuführen ist, weil die Maßstabsunbekannte m durch eine Translation nicht berührt wird. Eine Drehung um endliche Winkel bewirkt nach (141.5) eine nichtlineare Änderung der Koordinaten in Abhängigkeit von den Drehwinkeln. Da aber für die Parameterschätzung linearisiert wurde, brauchen nur die linearen Änderungen infolge differentieller Drehungen nach (141.6) berücksichtigt zu werden. Einer differentiellen Drehung der Koordinatenachsen um die Winkel $-d\alpha$, $-d\beta$ und $-d\gamma$ entspricht eine Drehung des gesamten Netzes um $d\alpha$, $d\beta$ und $d\gamma$. Bezeichnet man mit \mathbf{x}_{di0} und \mathbf{x}_{di} die Näherungskordinaten und die unbekannt Koordinatenkorrekturen des Punktes i nach der Drehung, erhält man mit (141.7) unter Vernachlässigung der Produkte kleiner Größen

$$\mathbf{x}_{di0} + \mathbf{x}_{di} = \begin{vmatrix} 1 & -d\gamma & d\beta \\ d\gamma & 1 & -d\alpha \\ -d\beta & d\alpha & 1 \end{vmatrix} (\mathbf{x}_{i0} + \mathbf{x}_i) = \mathbf{x}_{i0} + \mathbf{x}_i + \begin{vmatrix} 0 & z_{i0} & -y_{i0} \\ -z_{i0} & 0 & x_{i0} \\ y_{i0} & -x_{i0} & 0 \end{vmatrix} \begin{vmatrix} d\alpha \\ d\beta \\ d\gamma \end{vmatrix}. \quad (333.13)$$

Da nur die Änderung von \mathbf{x}_i nach \mathbf{x}_{di} infolge der Rotation interessiert, werden identische Näherungskordinaten vor und nach der Drehung angenommen, also $\mathbf{x}_{di0} = \mathbf{x}_{i0}$. Dann folgt für den gesamten Vektor $\boldsymbol{\beta}_d$ der unbekannt

Parameter nach der Rotation

$$\beta_d = \beta + \begin{vmatrix} \mathbf{S}_1 \\ \dots \\ \mathbf{S}_i \\ \dots \\ \mathbf{0}' \end{vmatrix} \begin{vmatrix} d\alpha \\ d\beta \\ d\gamma \end{vmatrix} \quad \text{mit} \quad \mathbf{S}_i = \begin{vmatrix} 0 & z_{i0} & -y_{i0} \\ -z_{i0} & 0 & x_{i0} \\ y_{i0} & -x_{i0} & 0 \end{vmatrix}. \quad (333.14)$$

Eine Maßstabsänderung durch den Faktor $1 + f$, in dem f eine kleine Größe bedeutet, ergibt für den Punkt i die Koordinaten $\mathbf{x}_{f_{i0}} + \mathbf{x}_{f_i} = (1 + f)(\mathbf{x}_{i0} + \mathbf{x}_i)$ oder $\mathbf{x}_{f_i} = \mathbf{x}_i + f\mathbf{x}_{i0}$ wegen $\mathbf{x}_{f_{i0}} = \mathbf{x}_{i0}$. Zur Maßstabsunbekannten m kann ein beliebiger positiver oder negativer Wert addiert werden, beispielsweise $m_f = m - m_0 f$, ohne daß die Erwartungswerte der Beobachtungen sich ändern. Nach einer Translation, Rotation und Maßstabsänderung folgt dann der Vektor β_a aus β mit

$$\beta_a = \beta + \mathbf{E}' | t_x, t_y, t_z, d\alpha, d\beta, d\gamma, f |'$$

und

$$\mathbf{E}' = \begin{vmatrix} \mathbf{I}_3 & \mathbf{S}_1 & \mathbf{x}_{10} \\ \dots & \dots & \dots \\ \mathbf{I}_3 & \mathbf{S}_i & \mathbf{x}_{i0} \\ \dots & \dots & \dots \\ \mathbf{0}' & \mathbf{0}' & -m_0 \end{vmatrix}. \quad (333.15)$$

Das letzte Element in der letzten Zeile von \mathbf{E}' wird derart gewählt, daß $\mathbf{X}\mathbf{E}' = \mathbf{0}$ sich ergibt, so daß wegen $\text{rg}\mathbf{E} = u - q = 7$ die Spalten der Matrix \mathbf{E}' nach (135.5) eine Basis für den Nullraum $N(\mathbf{X})$ der Matrix \mathbf{X} bilden.

Für Koordinatenbestimmungen mittels Streckenmessungen in der Ebene sind in der Matrix \mathbf{E}' wegen $z_i = 0$ jede dritte Zeile und, da $t_z = d\alpha = d\beta = 0$ gilt, die dritte bis fünfte Spalte zu streichen, so daß man erhält

$$\mathbf{E}' = \begin{vmatrix} 1 & 0 & -y_{10} & x_{10} \\ 0 & 1 & x_{10} & y_{10} \\ \dots & \dots & \dots & \dots \\ 1 & 0 & -y_{i0} & x_{i0} \\ 0 & 1 & x_{i0} & y_{i0} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -m_0 \end{vmatrix}. \quad (333.16)$$

Werden zur Bestimmung des Netzes in der Ebene Richtungen gemessen, so daß neben den unbekanntem Koordinatenkorrekturen x_i, y_i in jedem Punkt i eine Orientierungsunbekannte o_i einzuführen ist (WOLF 1968, S.274), dann

gilt mit $\boldsymbol{\beta} = |x_1, y_1, o_1, \dots, x_i, y_i, o_i, \dots, m|'$ und $\rho = 400/2\pi$ in der Dimension n für die Matrix \mathbf{E}' , falls die Orientierungsunbekannten o_i ebenfalls in der Dimension n definiert werden,

$$\mathbf{E}' = \begin{pmatrix} 1 & 0 & -y_{10} & x_{10} \\ 0 & 1 & x_{10} & y_{10} \\ 0 & 0 & \rho & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & -y_{i0} & x_{i0} \\ 0 & 1 & x_{i0} & y_{i0} \\ 0 & 0 & \rho & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & -m_0 \end{pmatrix}. \quad (333.17)$$

Für das Nivellement folgt schließlich

$$\mathbf{E}' = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix}.$$

Durch die Wahl der Projektion für die unbekannt Parameter, die sich nach (333.9) oder (333.10) als Hinzufügen der Restriktionen $\mathbf{B}\boldsymbol{\beta} = \mathbf{0}$ oder $\mathbf{E}\boldsymbol{\beta} = \mathbf{0}$ interpretieren läßt, wird das Datum eines freien Netzes festgelegt. Benutzt man die Pseudoinverse zur Projektion, so daß (333.5) folgt, entspricht dies $\mathbf{E}\boldsymbol{\beta} = \mathbf{0}$. Mit \mathbf{E} aus (333.15) erhält man für die drei ersten Restriktionen von $\mathbf{E}\boldsymbol{\beta} = \mathbf{0}$, daß für die Korrekturen x_s, y_s, z_s der Näherungskoodinaten des Schwerpunktes des Netzes bei k Punkten mit

$$x_s = \frac{1}{k} \sum_{i=1}^k x_i, \quad y_s = \frac{1}{k} \sum_{i=1}^k y_i, \quad z_s = \frac{1}{k} \sum_{i=1}^k z_i \quad (333.18)$$

der Nullvektor geschätzt wird, für dessen Kovarianzmatrix sich wegen (155.14), (233.2) und (333.6) $\mathbf{E}(\mathbf{X}'\mathbf{X})^+\mathbf{E}' = \mathbf{0}$ ergibt, so daß der Schwerpunkt zu einem festen Punkt wird (MEISSEL 1969; MITTERMAYER 1972; POPE 1971). Zur Interpretation der vier restlichen Restriktionen von $\mathbf{E}\boldsymbol{\beta} = \mathbf{0}$ soll angenommen werden, daß der Vektor $\boldsymbol{\beta}$ zusammen mit einem Fehlervektor \mathbf{e} durch eine Drehung und eine Maßstabsänderung dargestellt werden kann, also

$$\mathbf{E}'_o \mathbf{o} = \boldsymbol{\beta} + \mathbf{e} \quad \text{mit} \quad \mathbf{o} = |d\alpha, d\beta, d\gamma, f|',$$

worin \mathbf{E}'_o aus der vierten bis siebten Spalte von \mathbf{E}' in (333.15) gebildet wird. Berechnet man \mathbf{o} nach der Methode der kleinsten Quadrate, folgt mit (323.3)

$$\hat{\mathbf{o}} = (\mathbf{E}_o \mathbf{E}'_o)^{-1} \mathbf{E}_o \boldsymbol{\beta} = \mathbf{0} \quad (333.19)$$

wegen $\mathbf{E}\boldsymbol{\beta} = \mathbf{0}$. Die vierte bis siebte Bedingung in $\mathbf{E}\boldsymbol{\beta} = \mathbf{0}$ mit \mathbf{E} aus (333.15) bewirken also, daß die unbekannt Parameter derart geschätzt werden, daß die nach der Methode der kleinsten Quadrate bestimmte Drehung und Maßstabsänderung des freien Netzes gegenüber dem durch die Näherungskoodinaten definierten Netz gleich Null ist.

Bei der Benutzung der Pseudoinversen für die Datumsdefinition ergibt sich als weitere Eigenschaft aus (156.1) und (156.2), daß $\text{sp}D(\hat{\boldsymbol{\beta}}_e)$ und $\hat{\boldsymbol{\beta}}_e' \hat{\boldsymbol{\beta}}_e$ minimal werden. Umfaßt der Vektor der projizierten Parameter lediglich Koordinaten, bedeutet dies, daß die Summe der Varianzen der geschätzten Koordinaten minimal und daß die Quadratsumme der Korrekturen der Näherungskoodinaten minimal wird. Man bezeichnet letzteren Effekt als *Auffelderung* der gesuchten Koordinaten auf die Näherungskoodinaten (CASPARY 1978; WOLF 1973).

Bei freien Netzen wünscht man häufig, daß sich die Eigenschaft der minimalen Varianz und der Auffelderung nur auf bestimmte Punkte bezieht, zum Beispiel bei der Punkteinschaltung in freien Netzen, bei der man die Summe minimaler Varianzen lediglich für die Koordinaten der eingeschalteten Punkte und der Anschlußpunkte wünscht (KOCH 1983B) oder bei der Deformationsanalyse, wo zur Definition des Datums nur auf die festen Punkte aufgefördert werden darf (KOCH 1985A). Nach (156.3) und (156.4) erhält man diese Eigenschaft der minimalen Teilspur der Kovarianzmatrix oder der minimalen Teilnorm für die Koordinatenkorrekturen bestimmter Punkte mit einer symmetrischen reflexiven generalisierten Inversen, die mit einer Matrix \mathbf{B} berechnet wird, deren Spalten für die Koordinaten der betreffenden Punkte aus der Matrix \mathbf{E} übernommen werden und deren übrige Spalten Nullen enthalten.

Mit Matrizen \mathbf{B} und \mathbf{C} , die (155.12) erfüllen sollen, seien die beiden symmetrischen reflexiven generalisierten Inversen $[(\mathbf{X}'\mathbf{X})_{rs}^-]_b$ und $[(\mathbf{X}'\mathbf{X})_{rs}^-]_c$ berechnet und jeweils zur Datumsdefinition eines freien Netzes benutzt worden, so daß nach (333.3) und (333.4) die projizierten Parameter

$$\boldsymbol{\beta}_b = [(\mathbf{X}'\mathbf{X})_{rs}^-]_b \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

mit der Schätzung und ihrer Kovarianzmatrix

$$\hat{\boldsymbol{\beta}}_b = [(\mathbf{X}'\mathbf{X})_{rs}^-]_b \mathbf{X}'\mathbf{y} \quad \text{und} \quad D(\hat{\boldsymbol{\beta}}_b) = \sigma^2 [(\mathbf{X}'\mathbf{X})_{rs}^-]_b$$

sowie

$$\boldsymbol{\beta}_c = [(\mathbf{X}'\mathbf{X})_{rs}^-]_c \mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

mit

$$\hat{\boldsymbol{\beta}}_c = [(\mathbf{X}'\mathbf{X})_{rs}^-]_c \mathbf{X}'\mathbf{y} \quad \text{und} \quad D(\hat{\boldsymbol{\beta}}_c) = \sigma^2 [(\mathbf{X}'\mathbf{X})_{rs}^-]_c$$

erhalten werden. Substituiert man $\mathbf{X}' = \mathbf{X}'\mathbf{X}[(\mathbf{X}'\mathbf{X})_{rs}^-]_b\mathbf{X}'$ aus (153.5) in den Gleichungen für β_c und $\hat{\beta}_c$ erhält man

$$\beta_c = [(\mathbf{X}'\mathbf{X})_{rs}^-]_c\mathbf{X}'\mathbf{X}\beta_b \quad (333.20)$$

und mit (155.19)

$$\hat{\beta}_c = [(\mathbf{X}'\mathbf{X})_{rs}^-]_c\mathbf{X}'\mathbf{X}\hat{\beta}_b \quad \text{und} \\ D(\hat{\beta}_c) = [(\mathbf{X}'\mathbf{X})_{rs}^-]_c\mathbf{X}'\mathbf{X}D(\hat{\beta}_b)\mathbf{X}'\mathbf{X}[(\mathbf{X}'\mathbf{X})_{rs}^-]_c, \quad (333.21)$$

worin für den Projektionsoperator aus (155.16) folgt

$$[(\mathbf{X}'\mathbf{X})_{rs}^-]_c\mathbf{X}'\mathbf{X} = \mathbf{I} - \mathbf{E}'(\mathbf{C}\mathbf{E}')^{-1}\mathbf{C}. \quad (333.22)$$

Die Gleichung (333.21) gibt die *Datumstransformation* an, mit der die Schätzwerte der unbekannt Parameter und ihre Kovarianzmatrix von einer Datumsdefinition in eine andere überführt werden können. Diese Transformation entspricht der *S-Transformation*, die von (BAARDA 1973) eingeführt wurde (MIERLO 1981; KOCH 1982C).

Für Hypothesentests in freien Netzen sollte man wissen, welche Größen invariant in bezug auf Datumstransformationen bleiben. Aufgrund der Überlegungen, die zu (331.6) führen, muß die Gestalt des freien Netzes, die durch die Schätzwerte $\hat{\mathbf{y}}$ der Erwartungswerte $E(\mathbf{y})$ der Beobachtungen \mathbf{y} bestimmt wird, und die Kovarianzmatrix $D(\hat{\mathbf{y}})$ invariant bei Datumstransformationen bleiben. Es gilt

$$\mathbf{X}\hat{\beta}_b = \mathbf{y} + \hat{\mathbf{e}} = \hat{\mathbf{y}} \quad \text{mit} \quad D(\hat{\mathbf{y}}) = \mathbf{X}D(\hat{\beta}_b)\mathbf{X}'.$$

Mit (153.5) und (333.21) erhält man

$$\mathbf{X}\hat{\beta}_c = \mathbf{X}[(\mathbf{X}'\mathbf{X})_{rs}^-]_c\mathbf{X}'\mathbf{X}\hat{\beta}_b = \mathbf{X}\hat{\beta}_b = \hat{\mathbf{y}}$$

sowie

$$\mathbf{X}D(\hat{\beta}_c)\mathbf{X}' = \mathbf{X}[(\mathbf{X}'\mathbf{X})_{rs}^-]_c\mathbf{X}'\mathbf{X}D(\hat{\beta}_b)\mathbf{X}'\mathbf{X}[(\mathbf{X}'\mathbf{X})_{rs}^-]_c\mathbf{X}' \\ = \mathbf{X}D(\hat{\beta}_b)\mathbf{X}' = D(\hat{\mathbf{y}}),$$

wodurch die Invarianz bestätigt wird.

Invariant gegenüber Datumstransformationen sind auch, sofern

$$\mathbf{X}_p\mathbf{E}' = \mathbf{0} \quad (333.23)$$

gilt, die Beobachtungen $\hat{\mathbf{y}}_p$, die aufgrund der Schätzwerte der unbekannt Parameter vorhergesagt werden mit

$$\mathbf{X}_p \hat{\boldsymbol{\beta}}_b = \hat{\mathbf{y}}_p \quad \text{und} \quad D(\hat{\mathbf{y}}_p) = \mathbf{X}_p D(\hat{\boldsymbol{\beta}}_b) \mathbf{X}_p', \quad (333.24)$$

worin \mathbf{X}_p die Matrix der Koeffizienten bedeutet, die zur Vorhersage benötigt wird. Mit (333.21) und (333.22) folgt nämlich

$$\mathbf{X}_p \hat{\boldsymbol{\beta}}_c = \mathbf{X}_p [(\mathbf{X}' \mathbf{X})_{rs}^-]_c \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}_b = \mathbf{X}_p \hat{\boldsymbol{\beta}}_b = \hat{\mathbf{y}}_p$$

und

$$\begin{aligned} \mathbf{X}_p D(\hat{\boldsymbol{\beta}}_c) \mathbf{X}_p' &= \mathbf{X}_p [(\mathbf{X}' \mathbf{X})_{rs}^-]_c \mathbf{X}' \mathbf{X} D(\hat{\boldsymbol{\beta}}_b) \mathbf{X}' \mathbf{X} [(\mathbf{X}' \mathbf{X})_{rs}^-]_c \mathbf{X}_p' \\ &= \mathbf{X}_p D(\hat{\boldsymbol{\beta}}_b) \mathbf{X}_p' = D(\hat{\mathbf{y}}_p). \end{aligned}$$

Die Bedingung (333.23) ist zum Beispiel für den Typ von Beobachtungen erfüllt, mit dem das Netz aufgebaut wurde.

334 Gauß–Markoff–Modell mit nicht vollem Rang und Restriktionen

Im Gauß–Markoff–Modell (331.2) mit nicht vollem Rang werden jetzt (327.1) entsprechend Restriktionen eingeführt. Da lediglich die Schätzwerte schätzbarer Funktionen gesucht werden, sollen auch die Restriktionen als schätzbare Funktionen vorausgesetzt werden, für die dann (332.2) und (332.3) gelten. Man erhält somit das Modell

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad \text{rg} \mathbf{X} = q < u, \quad \mathbf{H}\boldsymbol{\beta} = \mathbf{w} \quad \text{und} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I} \quad (334.1)$$

sowie

$$\mathbf{H} = \mathbf{C}\mathbf{X} \quad \text{und} \quad \mathbf{H}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X} = \mathbf{H}, \quad (334.2)$$

worin mit $r \leq q$ die $r \times u$ Matrix \mathbf{H} wieder vollen Zeilenrang besitze und \mathbf{C} eine $r \times n$ Matrix bedeute.

Für den Vektor \mathbf{a} einer schätzbaren Funktion $\alpha = \mathbf{a}'\boldsymbol{\beta}$ der unbekannt Parameter $\boldsymbol{\beta}$ im Modell (334.1) folgt (332.2) entsprechend aus (327.3) $\mathbf{c}'\mathbf{X} + \mathbf{d}'\mathbf{H} = \mathbf{a}'$ und mit (334.2) $\mathbf{c}'\mathbf{X} + \mathbf{d}'\mathbf{C}\mathbf{X} = \mathbf{a}'$ oder $\bar{\mathbf{c}}'\mathbf{X} = \mathbf{a}'$ mit $\bar{\mathbf{c}}' = \mathbf{c}' + \mathbf{d}'\mathbf{C}$, so daß auch für den Vektor \mathbf{a} einer schätzbaren Funktion im Modell (334.1) mit Restriktionen die Beziehung (332.2) und damit (332.3) erfüllt ist und folglich der Begriff der Schätzbarkeit sich nicht ändert. Entsprechend (332.7) erhält man dann aus (327.8) den

Satz: Die beste lineare erwartungstreue und eindeutige Schätzung $\tilde{\alpha}$ der schätzbaren Funktion $\alpha = \mathbf{a}'\boldsymbol{\beta}$ der unbekannt Parameter $\boldsymbol{\beta}$ im Modell (334.1) ist gegeben durch

$$\tilde{\alpha} = \mathbf{a}'(\mathbf{X}'\mathbf{X})^{-}[\mathbf{X}'\mathbf{y} + \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}(\mathbf{w} - \mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y})]$$

und die eindeutige Varianz $V(\tilde{\alpha})$ der Schätzung durch

$$V(\tilde{\alpha}) = \sigma^2 \mathbf{a}'[(\mathbf{X}'\mathbf{X})^{-} - (\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}]\mathbf{a} . \quad (334.3)$$

Beweis: Zuerst soll gezeigt werden, daß die Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'$ positiv definit ist, so daß ihre Inverse nach (143.3) existiert. Es sei $(\mathbf{X}'\mathbf{X})^{-}$ eine positiv definite generalisierte Inverse von $\mathbf{X}'\mathbf{X}$, deren Existenz aus (153.3) mit $\mathbf{Q} = \mathbf{P}'$, da $\mathbf{X}'\mathbf{X}$ symmetrisch ist, $\mathbf{S} = \mathbf{0}, \mathbf{R} = \mathbf{0}, \mathbf{T} = \mathbf{I}$ und aus (143.7) folgt. Dann ist auch die Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'$ mit $\text{rg}\mathbf{H} = r$ wegen (143.7) positiv definit und wegen (143.9) auch ihre Inverse. Da nach (334.2) $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}' = \mathbf{C}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{C}'$ gilt, ist nach (153.8) $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'$ invariant gegenüber der Wahl von $(\mathbf{X}'\mathbf{X})^{-}$, so daß $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'$ und ihre Inverse auch für beliebige generalisierte Inversen positiv definit sind. Da α eine schätzbare Funktion ist, folgt mit (332.3) und (334.2) wegen $\mathbf{H}\boldsymbol{\beta} - \mathbf{w} = \mathbf{0}$ die Erwartungstreue $E(\tilde{\alpha}) = \mathbf{a}'\boldsymbol{\beta} = \alpha$ der Schätzung und mit (153.8), (332.2) und (334.2) ihre Eindeutigkeit. Mit $\tilde{\alpha}$ und mit (153.5), (153.7), (153.8), (233.2), (332.2), (334.1) und (334.2) ergibt sich der angegebene eindeutige Ausdruck für die Varianz $V(\tilde{\alpha})$, die minimal ist, wie sich analog zum Beweis für (327.7) zeigen läßt, so daß die Aussagen folgen.

Als nicht erwartungstreuer Schätzwert $\tilde{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$ im Modell (334.1), sofern nicht die bereits bei (331.3) erwähnten Sonderfälle vorliegen, bietet sich mit $\mathbf{a}' = |1, 0, 0, \dots|$, $\mathbf{a}' = |0, 1, 0, \dots|$ und so fort in (334.3) der Ausdruck an

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}[\mathbf{X}'\mathbf{y} + \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}(\mathbf{w} - \mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y})] . \quad (334.4)$$

Mit (331.3) folgt dann

$$\tilde{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}(\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}) . \quad (334.5)$$

Definiert man (327.16) entsprechend die wegen (153.8) eindeutigen Residuen $\tilde{\mathbf{e}}$ mit

$$\tilde{\mathbf{e}} = \mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{y} , \quad (334.6)$$

ergibt sich mit (153.5), (153.8), (233.2), (334.2) und (334.4) die Kovarianzmatrix $D(\tilde{\epsilon})$ der Residuen zu

$$D(\tilde{\epsilon}) = \sigma^2 [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1} \mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}']$$

mit $\text{rg}D(\tilde{\epsilon}) = n - q + r$ und $\text{sp}D(\tilde{\epsilon}) = \sigma^2(n - q + r)$ (334.7)

wegen (137.3), (152.3), (153.4), (153.5) und (334.2), da die Matrix $D(\tilde{\epsilon})/\sigma^2$ idempotent ist. Die Kovarianzmatrix $D(\tilde{\epsilon})$ ist also nach (133.1) singulär, aber eindeutig.

Die Quadratsumme Ω_H der Residuen $\tilde{\epsilon}$ ergibt sich mit (153.5), (331.3) und (334.6) entsprechend (327.18) zu

$$\Omega_H = (\mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{y}) = (\mathbf{X}\bar{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\bar{\boldsymbol{\beta}} - \mathbf{y}) + (\tilde{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}})' \mathbf{X}' \mathbf{X} (\tilde{\boldsymbol{\beta}} - \bar{\boldsymbol{\beta}}). \quad (334.8)$$

Da $\tilde{\epsilon}$ eindeutig ist, ist auch Ω_H eindeutig und analog zu (327.21) und (327.22) läßt sich zeigen, daß Ω_H minimal ist. Somit bedeutet (334.4) die Schätzung nach der Methode (313.1) der kleinsten Quadrate, die identisch ist mit der Maximum–Likelihood–Schätzung (314.2) im Falle normalverteilter Beobachtungen wegen (324.2). Man erhält also den

Satz: Im Gauß–Markoff–Modell (334.1) mit nicht vollem Rang und Restriktionen liefern die Methode der kleinsten Quadrate und die Maximum–Likelihood–Methode im Falle normalverteilter Beobachtungen identische Schätzwerte $\tilde{\boldsymbol{\beta}}$, die in (334.4) angegeben sind, für die Parameter $\boldsymbol{\beta}$. (334.9)

Entsprechend (327.14) ergibt sich die Maximum–Likelihood–Schätzung $\bar{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit zu

$$\bar{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}) = \frac{1}{n}\Omega_H. \quad (334.10)$$

Um eine erwartungstreue Schätzung zu gewinnen, soll $E(\Omega_H)$ berechnet werden. Der zweite Summand auf der rechten Seite von (334.8) läßt sich mit (334.5) sowie (153.5), (153.7) und (334.2) umformen, so daß mit (331.5) folgt

$$\Omega_H = \Omega + R \quad \text{mit} \quad R = (\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}(\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}). \quad (334.11)$$

Wegen (334.2) ist $\mathbf{H}\bar{\boldsymbol{\beta}}$ eindeutig, so daß auch R eindeutig ist. Weiter gilt $R \geq 0$, da $(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{H}')^{-1}$, wie für (334.3) bewiesen, positiv definit ist. Mit

$$E(\mathbf{H}\bar{\boldsymbol{\beta}}) = \mathbf{H}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{H}\boldsymbol{\beta} \quad (334.12)$$

wegen (331.3) sowie (334.2) und

$$D(\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}) = \sigma^2 \mathbf{H}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X}((\mathbf{X}'\mathbf{X})^-)' \mathbf{H}' = \sigma^2 \mathbf{H}(\mathbf{X}'\mathbf{X})^- \mathbf{H}' \quad (334.13)$$

wegen (153.5), (153.7) und (233.2) folgt für den Erwartungswert von Ω_H mit (271.1) und (331.11) aus (334.11)

$$E(\Omega_H) = E(\Omega) + \sigma^2 \text{sp} \mathbf{I}_r = \sigma^2(n - q + r), \quad (334.14)$$

so daß sich

$$\tilde{\sigma}^2 = \Omega_H / (n - q + r) \quad (334.15)$$

als erwartungstreue Schätzung der Varianz σ^2 der Gewichtseinheit ergibt.

Im Gauß–Markoff–Modell (334.1) mit Restriktionen sollen jetzt die nach (333.3) oder (333.5) projizierten Parameter geschätzt werden, so daß in (334.3) \mathbf{a}' durch die Matrizen $(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}$ oder $(\mathbf{X}'\mathbf{X})^+ \mathbf{X}'\mathbf{X}$ ersetzt wird, die (332.3) erfüllen, und die Matrix \mathbf{H} der Restriktionen durch $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}$ oder $\mathbf{H}(\mathbf{X}'\mathbf{X})^+ \mathbf{X}'\mathbf{X}$, die den Gleichungen (334.2) genügen und von denen vorausgesetzt wird, daß sie vollen Zeilenrang besitzen. Mit (333.3) und (333.5) ergeben sich dann als Restriktionen $\mathbf{H}\boldsymbol{\beta}_b = \mathbf{w}$ oder $\mathbf{H}\boldsymbol{\beta}_e = \mathbf{w}$ und mit (153.5) und (153.11) folgt der

Satz: Die beste lineare erwartungstreue Schätzung der nach (333.3) oder (333.5) projizierten Parameter ist im Gauß–Markoff–Modell (334.1) mit den Restriktionen $\mathbf{H}\boldsymbol{\beta}_b = \mathbf{w}$ oder $\mathbf{H}\boldsymbol{\beta}_e = \mathbf{w}$ durch

$$\tilde{\boldsymbol{\beta}}_b = (\mathbf{X}'\mathbf{X})_{rs}^- [\mathbf{X}'\mathbf{y} + \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{H}')^{-1}(\mathbf{w} - \mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{y})]$$

mit der Kovarianzmatrix

$$D(\tilde{\boldsymbol{\beta}}_b) = \sigma^2 [(\mathbf{X}'\mathbf{X})_{rs}^- - (\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{H}')^{-1} \mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^-]$$

gegeben oder entsprechend durch $\tilde{\boldsymbol{\beta}}_e$ und $D(\tilde{\boldsymbol{\beta}}_e)$, falls $(\mathbf{X}'\mathbf{X})_{rs}^-$ durch $(\mathbf{X}'\mathbf{X})^+$ ersetzt wird. (334.16)

Für die Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}$ der in (334.16) eingeführten Restriktionen wird im Modell vorausgesetzt, daß sie vollen Zeilenrang besitzt. Die Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{H}' = \mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{H}'$ ist dann regulär, wie sich aus dem Beweis zu (334.3) ergibt. Ist andererseits diese Matrix regulär, besitzt $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}$ vollen Zeilenrang, da wegen (153.8) eine positiv definite generalisierte Inverse in $\mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}'$ gewählt werden kann und

mit (143.5) beispielsweise $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' = \mathbf{X}\mathbf{G}\mathbf{G}'\mathbf{X}'$ erhalten wird. Hiermit folgt für die reguläre Matrix

$$\begin{aligned} \mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{H}' = \\ \mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{X}'\mathbf{X}\mathbf{G}\mathbf{G}'\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{H}' \end{aligned}$$

und mit (132.4) und (135.6), daß $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{X}'\mathbf{X}$ vollen Zeilenrang aufweist. Ist $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{H}'$ singulär, besitzt $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{X}'\mathbf{X}$ nicht vollen Zeilenrang. Die Regularität der Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{H}'$ gewährleistet also, daß die für die projizierten Parameter eingeführten Restriktionen die Modellvoraussetzungen erfüllen.

Verletzt werden die Voraussetzungen zum Beispiel durch Restriktionen, die die wiederholen oder denen widersprechen, die aus \mathbf{B} oder \mathbf{E} in den Modellen (333.9) oder (333.10) resultieren. Dies ergibt sich aus dem

Satz: Wird eine generalisierte Inverse $(\mathbf{X}'\mathbf{X})_{rs}^{-}$ in (334.16) mit der Matrix \mathbf{B} in (155.15) berechnet, so wird die Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{H}'$ singulär, falls die Matrix $|\mathbf{B}', \mathbf{H}'|$ nicht vollen Spaltenrang besitzt. Entsprechendes gilt auch bei der Berechnung der Pseudoinversen $(\mathbf{X}'\mathbf{X})^{+}$ mit der Matrix \mathbf{E} in (155.21).
(334.17)

Beweis: Für die Determinante der um \mathbf{B} und \mathbf{H} erweiterten Matrix $\mathbf{X}'\mathbf{X}$ gilt mit (136.12), (155.7) und (155.15)

$$\begin{aligned} \det \begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{B}' & \mathbf{H}' \\ \mathbf{B} & \mathbf{0} & \mathbf{0} \\ \mathbf{H} & \mathbf{0} & \mathbf{0} \end{vmatrix} = \det \mathbf{D} \det(-|\mathbf{H}, \mathbf{0}| \mathbf{D}^{-1} \begin{vmatrix} \mathbf{H}' \\ \mathbf{0} \end{vmatrix}) \\ = \det \mathbf{D} \det(-\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{H}'). \end{aligned}$$

Besitzt $|\mathbf{B}', \mathbf{H}'|$ nicht vollen Spaltenrang, ist die erweiterte Matrix singulär, so daß $\det(-\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{H}') = 0$ wegen $\det \mathbf{D} \neq 0$ folgt. Wegen (136.17) ist dann die Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-}\mathbf{H}'$ singulär.

Beispiel: Für die Projektion von Koordinaten eines ebenen, freien Streckennetzes seien für die Teilmenge von k Punkten die aus (333.16) folgenden Restriktionen eingeführt

$$x_1 + \dots + x_i + \dots + x_k = 0 \quad (334.18)$$

$$y_1 + \dots + y_i + \dots + y_k = 0 \quad (334.19)$$

$$-y_{10}x_1 + x_{10}y_1 - \dots - y_{i0}x_i + x_{i0}y_i - \dots - y_{k0}x_k + x_{k0}y_k = 0. \quad (334.20)$$

Als Restriktionen, beispielsweise für eine Nullhypothese, sollen die Identitäten der Koordinatenkorrekturen dieser k Punkte mit Null eingeführt werden. Zur

Vermeidung einer singulären Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^{-1}\mathbf{H}'$ werden nur die folgenden $2k - 3$ Restriktionen formuliert

$$x_1 = 0, \dots, x_{i-2} = 0, x_{i+1} = 0, \dots, x_k = 0, \quad (334.21)$$

$$y_1 = 0, \dots, y_{i-1} = 0, y_{i+1} = 0, \dots, y_k = 0. \quad (334.22)$$

Aus (334.19) und (334.22) folgt

$$y_i = 0 \quad \text{für beliebiges } i. \quad (334.23)$$

Aus (334.18) und (334.21) erhält man

$$x_{i-1} + x_i = 0 \quad \text{und daher } x_{i-1} = -x_i$$

und weiter mit (334.21) bis (334.23) aus (334.20)

$$-y_{i-1,0}x_{i-1} - y_{i0}x_i = (y_{i-1,0} - y_{i0})x_i = 0$$

folglich, da $y_{i-1,0} \neq y_{i0}$ vorausgesetzt werden kann,

$$x_{i-1} = 0 \quad \text{und } x_i = 0 \quad \text{für beliebiges } i. \quad (334.24)$$

Durch die Restriktionen (334.18) bis (334.20) werden also die drei in (334.21) und (334.22) fehlenden Restriktionen erfüllt. Entsprechende Aussagen folgen für die Differenzen von Koordinatenkorrekturen, falls auch die Restriktionen (334.18) bis (334.20) sich für diese Differenzen aufstellen lassen, wie das zum Beispiel bei der Deformationsanalyse der Fall ist (KOCH 1985A).

34 Spezielle Gauß–Markoff–Modelle

341 Polynommodell

Das Polynommodell dient häufig dazu, aus gemessenen Daten die Gleichung einer ebenen Kurve oder einer Fläche im dreidimensionalen Raum E^3 abzuleiten. Ist beispielsweise eine Kurve in E^2 gegeben, in der ein rechtwinkliges (x, y) -Koordinatensystem definiert ist, und wählt man auf der Kurve Punkte P_i mit den Koordinaten (x_i, y_i) derart aus, daß für gegebene Abszissen x_i die Ordinaten y_i gemessen werden, so läßt sich zur Darstellung der Kurve das Polynom aufbauen

$$\beta_0 + x_i\beta_1 + x_i^2\beta_2 + \dots + x_i^{u-1}\beta_{u-1} = E(y_i) \quad \text{mit } i \in \{1, \dots, n\}, \quad (341.1)$$

worin $\beta_0, \beta_1, \dots, \beta_{u-1}$ die unbekanntenen Parameter der Polynomentwicklung angeben. Besitzen die Ordinaten y_i gleiche Varianzen und sind sie unkorreliert,

ergibt sich mit $\mathbf{y} = (y_i)$ die Kovarianzmatrix $D(\mathbf{y}) = \sigma^2 \mathbf{I}$, so daß das Modell (321.5) vorliegt. Im E^3 mit einem (x, y, z) -Koordinatensystem läßt sich zur Darstellung einer Fläche folgendes Polynommodell mit wieder u unbekanntem Parametern wählen

$$\beta_0 + x_i \beta_1 + y_i \beta_2 + x_i^2 \beta_3 + x_i y_i \beta_4 + y_i^2 \beta_5 + \dots = E(z_i) \quad (341.2)$$

und entsprechende Polynommodelle im E^4 , E^5 und so fort. Die Schätzwerte $\hat{\beta}_i$ der unbekanntem Parameter β_i ergeben sich aus (322.9).

Beispiel: In den folgenden Koordinatenpaaren (x_i, y_i) mit $i \in \{1, \dots, 5\}$, die eine ebene Kurve repräsentieren,

x_i	-1	0	1	2	3
y_i	1,1	1,2	1,9	2,5	3,2

seien die Werte x_i gegeben und die Werte y_i gemessen, wobei die y_i unkorreliert seien und gleiche Varianzen besitzen sollen. Für die Darstellung des Kurvenverlaufs genüge ein Polynom 2. Grades und der Punkt $x_6 = 4$; $y_6 = 4,6$ liege auf der Kurve. Aus (341.1) ergibt sich dann $\beta_0 + x_i \beta_1 + x_i^2 \beta_2 = E(y_i)$ mit $i \in \{1, \dots, 5\}$ und $\beta_0 + x_j \beta_1 + x_j^2 \beta_2 = y_j$ für $j = 6$ oder

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad \mathbf{H}\boldsymbol{\beta} = \mathbf{w} \quad \text{und} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I},$$

so daß das Modell (327.2) vorliegt mit

$$\mathbf{X} = \begin{vmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{vmatrix}, \quad \boldsymbol{\beta} = \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{vmatrix}, \quad \mathbf{y} = \begin{vmatrix} 1,1 \\ 1,2 \\ 1,9 \\ 2,5 \\ 3,2 \end{vmatrix}, \quad \mathbf{H}' = \begin{vmatrix} 1 \\ 4 \\ 16 \end{vmatrix}, \quad \mathbf{w} = 4,6.$$

Das nach (327.9) erweiterte Normalgleichungssystem, der Absolutgliedvektor und die Quadratsumme $\mathbf{y}'\mathbf{y}$ ergeben sich zu

$$\begin{vmatrix} \mathbf{X}'\mathbf{X} & \mathbf{H}' \\ \mathbf{H} & \mathbf{0} \end{vmatrix} = \begin{vmatrix} 5 & 5 & 15 & 1 \\ & 15 & 35 & 4 \\ & & 99 & 16 \\ & & & 0 \end{vmatrix}, \quad \begin{vmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{w} \end{vmatrix} = \begin{vmatrix} 9,9 \\ 15,4 \\ 41,8 \\ 4,6 \end{vmatrix}, \quad \mathbf{y}'\mathbf{y} = 22,75.$$

Ergänzt man (326.4) entsprechend die erweiterte Normalgleichungsmatrix um den Absolutgliedvektor, um $\mathbf{y}'\mathbf{y}$ und um die 4×4 Einheitsmatrix, ergibt sich wegen (327.24) nach vier Eliminationsschritten $\Omega_H = 0,0749$ und daher $\tilde{\sigma}^2 = \Omega_H / (5 - 3 + 1) = 0,0250$ sowie $\tilde{\sigma} = 0,16$ und weiter

$$\tilde{\boldsymbol{\beta}} = \begin{vmatrix} 1,2966 \\ 0,3289 \\ 0,1242 \end{vmatrix} \quad \text{mit} \quad D(\tilde{\boldsymbol{\beta}}) = \sigma^2 \begin{vmatrix} 0,2932 & -0,0484 & -0,0062 \\ & 0,2792 & -0,0668 \\ & & 0,0171 \end{vmatrix}.$$

Der Vektor \tilde{e} der Residuen folgt mit

$$\tilde{e} = \begin{pmatrix} -0,0081 \\ 0,0966 \\ -0,1502 \\ -0,0486 \\ 0,2015 \end{pmatrix} \quad \text{und als Probe} \quad \Omega_H = 0,0749 .$$

Für die Restriktion gilt

$$\tilde{\beta}_0 + 4\tilde{\beta}_1 + 16\tilde{\beta}_2 = 4,599 .$$

Bei Polynommodellen erhebt sich häufig die Frage, bis zu welchem Grad die Polynomentwicklung zu treiben ist. Wird der Grad des Polynoms zu niedrig oder zu hoch gewählt, enthält das Polynommodell zu wenige oder zu viele Parameter. Die Folgen dieser Modellabweichungen wurden im Kapitel 329 erörtert. Zweckmäßig trifft man die Entscheidung über den Grad der Polynomentwicklung mit Hilfe des noch zu behandelnden Hypothesentests (423.10) unter Beachtung von (423.12).

Wird der Grad der Polynomentwicklung erhöht oder erniedrigt, muß die Schätzung sämtlicher Parameter wiederholt werden. Um das zu vermeiden, kann man mit *Orthogonalpolynomen* arbeiten, deren Koeffizienten, wie im folgenden gezeigt wird, vom Grad der Polynomentwicklung unabhängig sind. Das allgemeine Polynommodell (341.1) oder (341.2) sei durch

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1} \quad (341.3)$$

gegeben, so daß die Spalten von \mathbf{X} aus den Grundpolynomen $x_i^0 = 1, x_i^1, x_i^2, \dots, x_i^{u-1}$ oder $x_i^0 = 1, x_i^1, y_i^1, \dots$ aufgebaut sind. Um Orthogonalpolynome zu erhalten, ist dieses Modell in $z_i^{(0)}\gamma_0 + z_i^{(1)}\gamma_1 + \dots = E(y_i)$ oder in

$$\mathbf{Z}\boldsymbol{\gamma} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1} \quad (341.4)$$

derart zu transformieren, daß aus (322.9) die Normalgleichungsmatrix $\mathbf{Z}'\mathbf{P}\mathbf{Z} = \mathbf{I}$ erhalten wird. Die Schätzer $\hat{\gamma}$ der Parameter $\boldsymbol{\gamma}$ ergeben sich dann unkorreliert und können als Koeffizienten orthogonaler Polynome interpretiert werden, da die Spalten von \mathbf{Z} , also die neu gewonnenen Polynome, wegen $\mathbf{Z}'\mathbf{P}\mathbf{Z} = \mathbf{I}$ zueinander orthogonal bezüglich des mit der Matrix \mathbf{P} verallgemeinerten Skalarproduktes sind.

Die Matrix \mathbf{X} besitze vollen Spaltenrang und \mathbf{P} sei positiv definit, so daß nach (143.5) und (143.7) die Cholesky-Zerlegung $\mathbf{X}'\mathbf{P}\mathbf{X} = \mathbf{G}\mathbf{G}'$ folgt, in der \mathbf{G} eine reguläre untere Dreiecksmatrix bedeutet. Mit

$$\boldsymbol{\gamma} = \mathbf{G}'\boldsymbol{\beta} \quad \text{und} \quad \mathbf{Z} = \mathbf{X}(\mathbf{G}')^{-1} \quad (341.5)$$

ergeben sich dann die gesuchten Orthogonalpolynome, denn es gilt $\mathbf{Z}\boldsymbol{\gamma} = \mathbf{X}\boldsymbol{\beta}$ und $\mathbf{Z}'\mathbf{P}\mathbf{Z} = \mathbf{G}^{-1}\mathbf{X}'\mathbf{P}\mathbf{X}(\mathbf{G}')^{-1} = \mathbf{I}$. Aus (322.9) folgen dann die Schätzwerte $\hat{\boldsymbol{\gamma}} = \mathbf{Z}'\mathbf{P}\mathbf{y}$ oder mit (341.5)

$$\hat{\boldsymbol{\gamma}} = \mathbf{G}^{-1}\mathbf{X}'\mathbf{P}\mathbf{y}. \quad (341.6)$$

Nimmt man also in den für (341.3) sich ergebenden Normalgleichungen $\mathbf{X}'\mathbf{P}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{P}\mathbf{y}$ eine Cholesky–Elimination nach (133.23) vor, ergibt sich nach der Elimination anstelle des Absolutgliedvektors der Vektor $\hat{\boldsymbol{\gamma}}$. Die Schätzwerte $\hat{\mathbf{y}}$ der Erwartungswerte der Beobachtungen folgen mit (341.5) zu $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{G}')^{-1}\hat{\boldsymbol{\gamma}}$. Zu ihrer Berechnung wie auch zur Berechnung der Orthogonalpolynome ist die aus der Cholesky–Elimination gewonnene Matrix \mathbf{G}' zu invertieren.

Die Koeffizienten $\boldsymbol{\gamma}$ der Orthogonalpolynome sind vom Grad der Polynomentwicklung unabhängig, denn erweitert man die Matrix \mathbf{X} in (341.3) um die Spalte \mathbf{x}_{u+1} der entsprechenden Grundpolynome, so führt die Cholesky–Zerlegung der sich ergebenden Normalgleichungsmatrix auf

$$\left| \begin{array}{cc} \mathbf{X}'\mathbf{P}\mathbf{X} & \mathbf{X}'\mathbf{P}\mathbf{x}_{u+1} \\ \mathbf{x}'_{u+1}\mathbf{P}\mathbf{X} & \mathbf{x}'_{u+1}\mathbf{P}\mathbf{x}_{u+1} \end{array} \right| = \left| \begin{array}{c|c} \mathbf{G}\mathbf{G}' & |\mathbf{G}, \mathbf{0}_u| \mathbf{g}_{u+1} \\ \mathbf{g}'_{u+1} | \mathbf{G}, \mathbf{0}_u|' & \mathbf{g}'_{u+1} \mathbf{g}_{u+1} \end{array} \right| = \bar{\mathbf{G}}\bar{\mathbf{G}}',$$

worin $\mathbf{0}_u$ der $u \times 1$ Nullvektor und $\bar{\mathbf{G}} = \left| \begin{array}{c} \mathbf{G}, \mathbf{0}_u \\ \mathbf{g}'_{u+1} \end{array} \right|$ die entsprechend um den $(u+1) \times 1$ Vektor \mathbf{g}_{u+1} erweiterte untere Dreiecksmatrix bedeutet, die die gesamte Matrix \mathbf{G} enthält.

342 Varianzanalyse

Bestehen mathematische oder physikalische Gesetzmäßigkeiten zwischen den unbekanntem Parametern und den Beobachtungen, dann bestimmen sie den Aufbau der Koeffizientenmatrix im Gauß–Markoff–Modell. Häufig sind aber solche Gesetze unbekannt, und es können lediglich *qualitative* Aussagen in der Weise gemacht werden, daß die Beobachtungsergebnisse aus den Einflüssen verschiedener *Faktoren* resultieren. Diese Einflüsse, die als *Effekte* bezeichnet werden, bilden die unbekanntem Parameter. Die zu beobachtende Größe wird auch als *Merkmal* bezeichnet, so daß die Aufgabe besteht, die Effekte von Faktoren auf ein Merkmal zu analysieren. Beispielsweise hängen die beobachteten Ertragsergebnisse von Weizen in einem Gebiet von den zur Aussaat gelangten Sorten ab, die somit die Faktoren bilden, deren Effekte zu bestimmen sind. Da man zwar aufgrund des Versuchsaufbaus weiß, welche Faktoren das beobachtete Merkmal beeinflussen, ein Maß für die Beeinflussung aber nicht kennt, setzt man die unbekanntem Effekte additiv zusammen, so daß die

Koeffizientenmatrix \mathbf{X} in (321.1) lediglich Nullen oder Einsen enthält. Da die Addition der Effekte aus dem *Versuchsplan* folgt, bezeichnet man die Koeffizientenmatrix \mathbf{X} als *Versuchsplanmatrix*.

Im folgenden werden Versuchspläne mit *festen* Effekten behandelt, die sogenannten Modelle I, in denen in Übereinstimmung mit (321.1) die unbekannt Parameter feste Größen darstellen. Die Modelle mit Zufallsparametern, sogenannte Modelle II, befinden sich im Kapitel 35. Im Modell II interessieren häufig nur die Varianzen der Effekte, so daß Varianzkomponenten zu schätzen sind, was im Kapitel 36 behandelt wird. Ob die Effekte einzelner Faktoren überhaupt ein Merkmal beeinflussen, wird mit Hilfe des noch zu behandelnden Hypothesentests (423.10) geprüft. In die Testgröße (423.11) gehen Varianzen ein, so daß man die Schätzung der Effekte zusammen mit den Hypothesentests als *Varianzanalyse* bezeichnet.

Im allgemeinen wird sich das beobachtete Merkmal aus unterschiedlichen Einflüssen zusammensetzen, die in Faktoren mit unterschiedlichen *Stufen* zusammengefaßt werden. Je nach Anzahl und Anordnung der Faktoren unterscheidet man verschiedene Versuchspläne, die im folgenden behandelt werden. Jeder Versuchsplan enthält einen unbekannt Parameter, zu dem die Effekte sämtlicher Faktoren addiert werden und der, wie sich aus den noch abzuleitenden Beziehungen (343.5) ergibt, als Gesamtmittel aller Beobachtungen interpretiert werden kann.

a) Einfache Klassifikation (Einwegklassifikation)

Bei der einfachen Klassifikation besteht das Problem darin, die Effekte eines Faktors A in p Stufen auf ein Merkmal zu analysieren. Bedeutet μ das Mittel aller Beobachtungen, α_i der Effekt der i -ten Stufe mit $i \in \{1, \dots, p\}$ des Faktors A , y_{ij} die Beobachtung j mit $j \in \{1, \dots, n_i\}$ und $n_i \in \mathbb{N}$ unter der Einwirkung der i -ten Stufe des Faktors A , ergibt sich mit (321.1)

$$\mu + \alpha_i = E(y_{ij}) \quad \text{mit } i \in \{1, \dots, p\}, j \in \{1, \dots, n_i\}. \quad (342.1)$$

Beispiel: Die Wirkungen dreier Typen von Glühlampen auf ihre Brenndauer wurden untersucht und die Ergebnisse in der folgenden Tabelle zusammengefaßt. Gesucht sind die Effekte der drei Typen. Die Versuchsplanmatrix \mathbf{X} , der Vektor $\boldsymbol{\beta}$ der unbekannt Parameter und der Beobachtungsvektor \mathbf{y} in

Stufen des Faktors A		Brenndauer in Stunden
1	Typ A_1	822, 830, 842
2	Typ A_2	883, 917
3	Typ A_3	913, 924

(321.1) ergeben sich dann mit (342.1) zu

$$\mathbf{X} = \begin{vmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{vmatrix}, \boldsymbol{\beta} = \begin{vmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{vmatrix}, \mathbf{y} = \begin{vmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{vmatrix} = \begin{vmatrix} 822 \\ 830 \\ 842 \\ 883 \\ 917 \\ 913 \\ 924 \end{vmatrix}.$$

b) Zweifache Kreuzklassifikation (Zweiwegkreuzklassifikation)

Bei der zweifachen Kreuzklassifikation sind die Effekte zweier Faktoren A und B auf ein Merkmal zu untersuchen, wobei der Faktor A in p Stufen und der Faktor B in q Stufen auftreten. Bezeichnet man mit n_{ij} die Anzahl der Beobachtungen des Merkmals unter dem Einfluß der i -ten Stufe des Faktors A und der j -ten Stufe des Faktors B , lassen sich die Versuchsergebnisse in der folgenden Tabelle zusammenfassen. Bedeuten α_i und β_j die Effekte der

		Stufen des Faktors B			
		1	2	...	q
Stufen des Faktors A	1	n_{11}	n_{12}	...	n_{1q}
	2	n_{21}	n_{22}	...	n_{2q}
			
	p	n_{p1}	n_{p2}	...	n_{pq}

Faktoren A und B , μ das Mittel aller Beobachtungen, ergibt sich mit (321.1)

$$\mu + \alpha_i + \beta_j = E(y_{ijk})$$

mit $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$, $k \in \{1, \dots, n_{ij}\}$. (342.2)

Gilt $n_{ij} = 0$ für mindestens ein Paar (i, j) , bezeichnet man den Versuchsplan als *unvollständig* und andernfalls als *vollständig*. Versuchspläne, in denen n_{ij} für alle Paare (i, j) gleich sind, heißen *balanciert*. Wie sich aus dem folgenden Kapitel ergibt, braucht für die Parameterschätzung zwischen diesen verschiedenen Versuchsplänen nicht unterschieden zu werden, sofern durch die fehlenden Beobachtungen kein Rangabfall in der Versuchsplanmatrix hervorgerufen wird.

Beispiel: Für eine Marktanalyse sollen die beobachteten Einschaltzeiten von Fernsehgeräten in Abhängigkeit von Großstädten, Mittel- und Kleinstädten, die den Stufen des Faktors A entsprechen, und von der Anzahl der Personen,

die die Stufen des Faktors B bilden, in den Haushalten mit den Fernsehgeräten untersucht werden. Die Analyse führt auf den Versuchsplan (342.2).

c) Zweifache hierarchische Klassifikation (hierarchische Zweiwegklassifikation)

Das Problem besteht nun darin, die Effekte eines Faktors A in p Stufen und eines Faktors B in q Stufen, dessen Stufen jeweils nur in Verbindung mit einer Stufe des Faktors A auftritt, auf ein Merkmal zu untersuchen. Der Versuchsplan läßt sich mit Hilfe der folgenden Tabelle veranschaulichen. Bezeichnet

Stufen des Faktors A	1	2	...	p
Stufen des Faktors B innerhalb der Stufen von A	$1, \dots, q_1$	$1, \dots, q_2$...	$1, \dots, q_p$

man wieder das Mittel mit μ , die Effekte des Faktors A mit α_i und die des Faktors B innerhalb der Stufen von A mit β_{ij} , ergibt sich mit (321.1)

$$\mu + \alpha_i + \beta_{ij} = E(y_{ijk})$$

mit $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q_i\}$, $k \in \{1, \dots, n_{ij}\}$. (342.3)

Beispiel: An einer Schule wird in den drei Fächern Englisch, Französisch und Geschichte eine neue Unterrichtsmethode bei insgesamt acht verschiedenen Klassen erprobt, wobei drei Klassen auf das Fach Englisch, zwei auf Französisch und drei auf Geschichte entfallen. Die Reaktionen der Schüler auf die neue Methode werden gemessen, so daß die drei Effekte des Faktors A , die den Wirkungen der neuen Methode in den drei Fächern entsprechen, und die Effekte des Faktors B innerhalb von A , die den Einfluß der Klassen in den einzelnen Fächern ausdrücken, im Versuchsplan (342.3) zu bestimmen sind.

d) Zweifache Klassifikation mit Wechselwirkung

Bis jetzt wurde davon ausgegangen, daß sich die Effekte der Faktoren A und B in den einzelnen Stufen addieren. Ist diese Voraussetzung nicht erfüllt, beeinflussen die Faktoren also das beobachtete Merkmal nicht unabhängig voneinander, spricht man von *Wechselwirkungen* zwischen den Faktoren, deren Effekte $(\alpha\beta)_{ij}$ wie folgt zu berücksichtigen sind,

$$\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} = E(y_{ijk})$$

mit $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$, $k \in \{1, \dots, n_{ij}\}$. (342.4)

Die Anlage der Experimente, die auf die Versuchspläne führt, wurde bislang nicht erwähnt. Wenn möglich, sollten die Versuche zufällig angeordnet sein, so daß man von einer *Randomisierung* der Experimente spricht (SCHEFFÉ 1959, S.105). Sie soll verhindern, daß außer den in Ansatz gebrachten Faktoren

weitere systematische Einflüsse auf das Versuchsergebnis wirken. Werden beispielsweise die Effekte von i Düngemitteln auf j Versuchsfeldern untersucht, so sollte die Anwendung der i Dünger auf jedem der j Felder in zufälliger Art und nicht auf systematische Weise geschehen.

e) Höhere Klassifikation mit Wechselwirkungen

Der Versuchsplan (342.4) der zweifachen Klassifikation mit Wechselwirkungen läßt sich auf beliebige höhere Klassifikationen mit Wechselwirkungen ausdehnen, so daß man allgemein erhält, wobei die Anzahl der Effekte und Wechselwirkungen endlich ist,

$$\begin{aligned} \mu + \alpha_i + \beta_j + \gamma_k + \dots + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \dots + (\alpha\beta\gamma)_{ijk} + \dots \\ = E(y_{ijk\dots l}) \quad (342.5) \end{aligned}$$

mit $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$, $k \in \{1, \dots, r\}$, $l \in \{1, \dots, n_{ijk\dots}\}$.

Die aus den Versuchsergebnissen resultierenden Beobachtungen werden im allgemeinen unkorreliert sein und gleiche Varianzen besitzen. Jedoch auch unterschiedliche Varianzen und von Null verschiedene Kovarianzen sind denkbar, so daß sich zur Vervollständigung der Versuchspläne (342.1) bis (342.5) entsprechend (321.5) oder (321.1) ergibt

$$D(\mathbf{y}) = \sigma^2 \mathbf{I} \quad \text{oder} \quad D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}. \quad (342.6)$$

343 Parameterschätzung für die Varianzanalyse mit einer symmetrischen reflexiven generalisierten Inversen

Die Besonderheit der Varianzanalyse besteht darin, daß die Versuchsplanmatrix nicht vollen Spaltenrang besitzt, da zum Mittel μ aller Beobachtungen die Effekte $\alpha_i, \beta_j, \gamma_k, \dots$ der Faktoren A, B, C, \dots in den einzelnen Stufen hinzuaddiert werden. Diese Vorgehensweise ist aber praktisch, da sich dann die Versuchsplanmatrizen bequem aufstellen lassen. Versuchsplanmatrizen mit vollem Rang würde man erhalten, wenn lediglich die Differenzen in bezug auf eine Stufe, beispielsweise $\alpha_i - \alpha_1, \beta_j - \beta_1, \gamma_k - \gamma_1, \dots$, bestimmt würden. Man erkennt dies für die Effekte α_i an der Versuchsplanmatrix \mathbf{X} des Beispiels zu (342.1), in der die erste Spalte sich aus der Summe der drei folgenden Spalten ergibt, so daß zum Beispiel die zweite Spalte zu streichen ist, was der Bestimmung von $\alpha_i - \alpha_1$ entspricht, um eine Matrix mit vollem Rang zu erhalten. Für die Effekte β_j, γ_k, \dots ist dies aus entsprechend erweiterten Versuchsplanmatrizen ersichtlich. Daß in der Varianzanalyse lediglich Differenzen von Effekten erwartungstreu im Sinne von (332.1) schätzbar sind, folgt auch aus (332.4) und (332.5), wonach bei drei Faktoren die Summen

$\mu + \alpha_i + \beta_j + \gamma_k$ für alle Tripel (i, j, k) und lineare Funktionen von ihnen, beispielsweise $\alpha_i - \alpha_1$, erwartungstreu schätzbar sind.

Der Rang einer Versuchsplanmatrix mit den Effekten α_i und β_j für $i \in \{1, \dots, p\}$ und $j \in \{1, \dots, q\}$ beträgt aufgrund der vorangegangenen Überlegungen $1 + (p - 1) + (q - 1)$. Kommen noch pq Wechselwirkungen $(\alpha\beta)_{ij}$ hinzu, so erhöhen nur die Wechselwirkungen den Rang der Versuchsplanmatrix jeweils um Eins, die sich auf die Differenzen $\alpha_i - \alpha_1$ und $\beta_i - \beta_1$ beziehen, insgesamt also $(p - 1)(q - 1)$ Wechselwirkungen, wie aus einer entsprechenden Versuchsplanmatrix durch Streichen der Spalten für α_1 und β_1 und der restlichen linear abhängigen Spalten ersichtlich ist. Auf diese Weise läßt sich der Rang der Matrizen der behandelten Versuchspläne angeben, wie es in Tabelle (343.1) geschehen ist, die zusätzlich die Anzahl der unbekannt Parameter in den Versuchsplänen enthält.

Versuchsplan	Anzahl der Parameter	Rang der Versuchsplanmatrix
(342.1)	$1 + p$	$1 + (p - 1)$
(342.2)	$1 + p + q$	$1 + (p - 1) + (q - 1)$
(342.3)	$1 + p + \Sigma q_i$	$1 + (p - 1) + \Sigma(q_i - 1)$
(342.4)	$1 + p + q + pq$	$1 + (p - 1) + (q - 1) + (p - 1)(q - 1) = pq$
(342.5)	$1 + p + q + r + \dots$	$1 + (p - 1) + (q - 1) + (r - 1) + \dots$
	$+pq + pr + qr + \dots$	$+(p - 1)(q - 1) + (p - 1)(r - 1)$
	$+pqr + \dots$	$+(q - 1)(r - 1) + \dots + (p - 1)$
		$(q - 1)(r - 1) + \dots = pqr \dots$

(343.1)

Wegen des Rangdefektes der Versuchsplanmatrizen stellen die unbekannt Parameter keine schätzbaren Funktionen dar, so daß an ihrer Stelle die nach (333.3) projizierten Parameter geschätzt werden sollen. Ihre Schätzwerte folgen nach (333.4) mit Hilfe einer symmetrischen reflexiven generalisierten Inversen der Normalgleichungsmatrix oder nach (333.7) mit Hilfe einer Matrix \mathbf{B} , die sich, wie im folgenden gezeigt wird, aufgrund der Überlegungen angeben läßt, die zur Matrix \mathbf{E} in (333.8) für die Schätzung mit der Pseudoinversen führen. In \mathbf{E} sind nämlich, wie schon bei (333.11) erwähnt, die Änderungen enthalten, die die Parameter β vornehmen können, ohne daß in den Versuchsplänen (342.1) bis (342.5) die Erwartungswerte der Beobachtungen sich ändern, denn die Spalten von \mathbf{E}' bilden nach (155.11) eine Basis für den Nullraum der Versuchsplanmatrix.

Wie oben gezeigt wurde, lassen sich nicht die Effekte $\alpha_i, \beta_j, \gamma_k, \dots$ der Stufen der Faktoren A, B, C, \dots , sondern nur die jeweiligen Differenzen, beispielsweise $\alpha_i - \alpha_1, \beta_j - \beta_1, \gamma_k - \gamma_1, \dots$, erwartungstreu schätzen. Zu den Effekten

jeden Faktors können daher die Konstanten

$$\Delta\alpha, \Delta\beta, \Delta\gamma, \dots \quad (343.2)$$

addiert werden, ohne daß die Erwartungswerte der Beobachtungen sich ändern. Zu den pq Wechselwirkungen $(\alpha\beta)_{ij}$, von denen nur $(p-1)(q-1)$ Wechselwirkungen erwartungstreu schätzbar sind, lassen sich $pq - (p-1)(q-1) = p+q-1$ Konstanten addieren und zwar zu den Wechselwirkungen $(\alpha\beta)_{i1}, (\alpha\beta)_{i2}, \dots, (\alpha\beta)_{iq}$ mit $i \in \{1, \dots, p\}$ die Konstanten

$$\Delta(\alpha\beta_{.1}), \Delta(\alpha\beta_{.2}), \dots, \Delta(\alpha\beta_{.q}) \quad (343.3)$$

und zu den Wechselwirkungen $(\alpha\beta)_{2j}, (\alpha\beta)_{3j}, \dots, (\alpha\beta)_{pj}$ mit $j \in \{1, \dots, q\}$ die Konstanten

$$\Delta(\alpha\beta_{2.}), \Delta(\alpha\beta_{3.}), \dots, \Delta(\alpha\beta_{p.}). \quad (343.4)$$

Daß die Konstante $\Delta(\alpha\beta_{.1})$ nicht addiert zu werden braucht, kann man dadurch interpretieren, daß $\Delta(\alpha\beta_{.1}) = \Delta(\alpha\beta_{1.})$ gilt. Mit Hilfe der Konstanten läßt sich, wie im folgenden Beispiel gezeigt wird, eine Matrix \mathbf{B} in (333.7) für eine Versuchsplanmatrix aufbauen, wobei die Anzahl der Zeilen von \mathbf{B} der aus (343.1) berechenbaren Differenz zwischen Anzahl der Parameter und Rang der Versuchsplanmatrix gleicht.

Beispiel: Gegeben sei das Modell (342.4) mit $p = 2$ und $q = 2$, so daß sich der Vektor $\boldsymbol{\beta}$ der unbekannt Parameter ergibt zu

$$\boldsymbol{\beta} = |\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, (\alpha\beta)_{11}, (\alpha\beta)_{12}, (\alpha\beta)_{21}, (\alpha\beta)_{22}|'.$$

Faßt man im Vektor \mathbf{t} die Konstanten zusammen, die zu den unbekannt Parametern addiert werden können, ohne daß sich die Erwartungswerte der Beobachtungen ändern, ergibt sich mit (343.2) bis (343.4)

$$\mathbf{t} = |\Delta\alpha, \Delta\beta, \Delta(\alpha\beta_{.1}), \Delta(\alpha\beta_{.2}), \Delta(\alpha\beta_{2.})|'.$$

Der durch den Vektor \mathbf{t} geänderte Parametervektor sei mit $\boldsymbol{\beta}_t$ bezeichnet, dann erhält man $\boldsymbol{\beta}_t = \boldsymbol{\beta} + \mathbf{B}'\mathbf{t}$ mit

$$\mathbf{B} = \begin{vmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{vmatrix}.$$

In der Matrix \mathbf{B} sind noch nicht alle Änderungen erfaßt, die die Parameter vornehmen können, da $\mathbf{XB}' \neq \mathbf{0}$ gilt. Ausgehend von \mathbf{B} läßt sich aber die Matrix \mathbf{E} konstruieren, für die $\mathbf{XE}' = \mathbf{0}$ gilt. Man erhält

$$\mathbf{E} = \begin{pmatrix} -1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Weiter besitzt die Matrix \mathbf{BE}' vollen Rang, so daß (155.12) gilt und mit \mathbf{B} eine symmetrische reflexive generalisierte Inverse für die Parameterschätzung zu berechnen ist.

Diesem Beispiel entsprechend erhält man für die verschiedenen Versuchspläne Matrizen \mathbf{B} . Nach (333.9) lassen sich mit \mathbf{B} auch die Restriktionen $\mathbf{B}\boldsymbol{\beta} = \mathbf{0}$ aufstellen. Für den Versuchsplan (342.5) ergeben sich mit \mathbf{B} die Restriktionen

$$\begin{aligned} \sum_{i=1}^p \alpha_i &= 0, \quad \sum_{j=1}^q \beta_j = 0, \quad \sum_{k=1}^r \gamma_k = 0, \dots, \quad \sum_{i=1}^p (\alpha\beta)_{ij} = 0, \\ \sum_{j=1}^q (\alpha\beta)_{ij} &= 0 \quad \text{für } i \neq 1, \quad \sum_{i=1}^p (\alpha\gamma)_{ik} = 0, \quad \sum_{k=1}^r (\alpha\gamma)_{ik} = 0 \quad \text{für } i \neq 1, \\ \sum_{j=1}^q (\beta\gamma)_{jk} &= 0, \quad \sum_{k=1}^r (\beta\gamma)_{jk} = 0 \quad \text{für } j \neq 1, \dots, \\ \sum_{i=1}^p (\alpha\beta\gamma)_{ijk} &= 0 \quad \text{für } j \neq 1, \quad \sum_{j=1}^q (\alpha\beta\gamma)_{ijk} = 0 \quad \text{für } k \neq 1, \\ \sum_{k=1}^r (\alpha\beta\gamma)_{ijk} &= 0 \quad \text{für } i \neq 1, \text{ ausgenommen } i = j = 1, \dots \end{aligned} \quad (343.5)$$

Umgekehrt läßt sich aus diesen Restriktionen mit $\mathbf{B}\boldsymbol{\beta} = \mathbf{0}$ eine Matrix \mathbf{B} in (333.7) oder (333.9) angeben.

Die Restriktionen (343.5) werden häufig zur Parameterschätzung für die Varianzanalyse benutzt (AHRENS 1968; JOHNSON und LEONE 1977; SCHEFFÉ 1959). Sie bewirken also eine Lösung mit einer symmetrischen reflexiven generalisierten Inversen der Normalgleichungsmatrix. Wünscht man eine Lösung mit der Pseudoinversen, ist aus der Matrix \mathbf{B} die Matrix \mathbf{E} zu konstruieren, wie im obigen Beispiel gezeigt. Die Pseudoinverse führt aber im Vergleich mit (343.5) zu unübersichtlichen Restriktionen, so daß die symmetrische re-

flexive generalisierte Inverse aufgrund von (343.5) vorzuziehen ist. Ein Rechenbeispiel für eine Parameterschätzung in der Varianzanalyse mit Hilfe der Restriktionen (343.5) befindet sich im Zusammenhang mit der multivariaten Parameterschätzung im Kapitel 374.

344 Kovarianzanalyse

Das Modell der *Kovarianzanalyse* entsteht durch die Vereinigung eines Gauß–Markoff–Modells, das aufgrund einer qualitativen Aussage entsteht, und eines Gauß–Markoff–Modells aufgrund quantitativer Aussagen. Es wird also das Modell einer Varianzanalyse mit einer Regressionsanalyse kombiniert, und man erhält

$$\mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}, \quad (344.1)$$

worin \mathbf{X} die Versuchsplanmatrix bedeutet, die als Elemente lediglich Nullen oder Einsen enthält, $\boldsymbol{\alpha}$ der Vektor der Effekte der Faktoren des Versuchsplans, \mathbf{Z} die Koeffizientenmatrix aufgrund quantitativer Aussagen, $\boldsymbol{\beta}$ der zugehörige Parametervektor und \mathbf{P} die positiv definite Gewichtsmatrix des Beobachtungsvektors \mathbf{y} eines Merkmals.

Der Name Kovarianzanalyse rührt daher, daß die Variablen, mit denen die Koeffizienten der Matrix \mathbf{Z} aufgebaut werden, beispielsweise die Variablen x_i mit $x_i^0, x_i^1, x_i^2, \dots, x_i^m$ bei einer Polynomentwicklung, als *Kovariablen* bezeichnet werden. Die Variablen, aus denen die Versuchsplanmatrix \mathbf{X} folgt, nennt man daher auch *Scheinvariable*, da sie nur die Werte Null und Eins annehmen können.

Es besitze \mathbf{Z} vollen Spaltenrang, während \mathbf{X} als Versuchsplanmatrix einen Rangdefekt aufweise, so daß wie im vorangegangenen Kapitel die Normalgleichungen mit Hilfe der Restriktionen (343.5) gelöst werden sollen, die wieder in der Matrix \mathbf{B} zusammengefaßt seien. Mit (333.7) ergeben sich dann die Schätzwerte der Parameter im Modell (344.1) zu

$$\begin{pmatrix} \hat{\boldsymbol{\alpha}}_b \\ \hat{\boldsymbol{\beta}} \\ \mathbf{k} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{P}\mathbf{X} & \mathbf{X}'\mathbf{P}\mathbf{Z} & \mathbf{B}' \\ \mathbf{Z}'\mathbf{P}\mathbf{X} & \mathbf{Z}'\mathbf{P}\mathbf{Z} & \mathbf{0} \\ \mathbf{B} & \mathbf{0} & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'\mathbf{P}\mathbf{y} \\ \mathbf{Z}'\mathbf{P}\mathbf{y} \\ \mathbf{0} \end{pmatrix}. \quad (344.2)$$

Bei der Berechnung der inversen Normalgleichungsmatrix gilt das bei (133.11) Erwähnte, da $\mathbf{X}'\mathbf{P}\mathbf{X}$ singular ist. Die Kovarianzanalyse eignet sich beispielsweise zur Untersuchung von Immobilienwerten (KOCH 1995).

345 Gauß–Markoff–Modell für Ausreißer in den Beobachtungen

Bei Meßreihen können einzelne Beobachtungen aus irgendwelchen Gründen grob verfälscht sein; man sagt dann, daß diese Beobachtungen *Ausreißer* enthalten. Die Ausreißer müssen entdeckt werden, damit sie die Parameterschätzung nicht beeinflussen. Hierzu dienen Ausreißertests für den Fall, daß nur wenige Ausreißer in sich gegenseitig kontrollierenden Beobachtungen erwartet werden. Die Ausreißertests werden im Kapitel 44 abgeleitet. Sie benötigen ein spezielles Gauß–Markoff–Modell, das im folgenden behandelt wird. Kann man nicht erwarten, daß die Ausreißer durch Ausreißertests gefunden werden, sind die robusten Parameterschätzungen des Kapitels 38 anzuwenden.

Den Ausgangspunkt bildet das Gauß–Markoff–Modell mit nicht vollem Rang

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) = \mathbf{y} + \mathbf{e} \quad \text{mit} \quad \text{rg}\mathbf{X} = q < u, \quad E(\mathbf{e}) = \mathbf{0} \\ \text{und} \quad D(\mathbf{y}) = D(\mathbf{e}) = \sigma^2 \mathbf{P}^{-1}, \quad (345.1)$$

in dem wieder \mathbf{X} die $n \times u$ Matrix gegebener Koeffizienten, $\boldsymbol{\beta}$ der $u \times 1$ Vektor unbekannter Parameter, \mathbf{y} der $n \times 1$ Zufallsvektor der Beobachtungen, \mathbf{e} der $n \times 1$ Fehlervektor, σ^2 die Varianz der Gewichtseinheit und \mathbf{P} die $n \times n$ positiv definite Gewichtsmatrix der Beobachtungen bedeuten. Als schätzbare Funktionen werden (333.3) entsprechend die projizierten Parameter $\boldsymbol{\beta}_b$ eingeführt

$$\boldsymbol{\beta}_b = (\mathbf{X}'\mathbf{P}\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{P}\mathbf{X}\boldsymbol{\beta}, \quad (345.2)$$

deren beste lineare erwartungstreue Schätzung $\hat{\boldsymbol{\beta}}_b$ sich nach (333.4) zu

$$\hat{\boldsymbol{\beta}}_b = (\mathbf{X}'\mathbf{P}\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{P}\mathbf{y} \quad (345.3)$$

ergibt. Dann folgt der Vektor $\hat{\mathbf{e}}$ der Residuen mit

$$\hat{\mathbf{e}} = -(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{P})\mathbf{y} \quad (345.4)$$

und seine Kovarianzmatrix $D(\hat{\mathbf{e}})$ aufgrund von (153.6) und (233.2) mit

$$D(\hat{\mathbf{e}}) = \sigma^2 \mathbf{Q}_e = \sigma^2 (\mathbf{P}^{-1} - \mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})_{rs}^- \mathbf{X}') \quad (345.5)$$

und schließlich die quadratische Form Ω der Residuen aus (325.1) mit

$$\Omega = \mathbf{y}'(\mathbf{P} - \mathbf{P}\mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{P})\mathbf{y}. \quad (345.6)$$

Das Gauß–Markoff–Modell (345.1) wird jetzt um den $r \times 1$ Vektor $\boldsymbol{\epsilon}$ zusätzlicher unbekannter Parameter und um die zugehörige $n \times r$ Matrix \mathbf{Z} bekannter

Koeffizienten erweitert

$$|\mathbf{X}, \mathbf{Z}| \begin{vmatrix} \boldsymbol{\beta} \\ \boldsymbol{\epsilon} \end{vmatrix} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = \sigma_\epsilon^2 \mathbf{P}^{-1}, \quad (345.7)$$

worin σ_ϵ^2 die Varianz der Gewichtseinheit des erweiterten Modells bezeichnet. Der Vektor $\boldsymbol{\epsilon}$ enthält die Ausreißer, die jeweils von den Beobachtungen subtrahiert werden. Soll zum Beispiel nur der Ausreißer Δy_k in der Beobachtung y_k mit $\mathbf{y} = (y_i)$ vorliegen, dann definiert man mit $r = 1$

$$\boldsymbol{\epsilon} = \Delta y_k \quad \text{und} \quad \mathbf{Z} = \mathbf{e}_k \quad \text{mit} \quad \mathbf{e}_k = |0, \dots, 0, 1, 0, \dots, 0|', \quad (345.8)$$

wobei die k -te Komponente von \mathbf{e}_k den Wert Eins erhält. Mit $\mathbf{X} = |\mathbf{x}_1, \dots, \mathbf{x}_n|'$ lautet die Beobachtungsgleichung für die k -te Beobachtung

$$\mathbf{x}'_k \boldsymbol{\beta} = E(y_k) - \Delta y_k$$

und für alle übrigen Beobachtungen $\mathbf{x}'_i \boldsymbol{\beta} = E(y_i)$. Will man beispielsweise die Ausreißer $\Delta y_k, \Delta y_{k+1}, \dots, \Delta y_l$ in den Beobachtungen y_k, y_{k+1}, \dots, y_l vorsehen, setzt man

$$\boldsymbol{\epsilon} = |\Delta y_k, \Delta y_{k+1}, \dots, \Delta y_l|' \quad \text{und} \quad \mathbf{Z} = |\mathbf{e}_k, \mathbf{e}_{k+1}, \dots, \mathbf{e}_l|. \quad (345.9)$$

Entsprechend (345.3) erhält man die Schätzwerte der projizierten Parameter im erweiterten Modell (345.7) zu

$$\begin{vmatrix} \hat{\boldsymbol{\beta}}_b \\ \hat{\boldsymbol{\epsilon}} \end{vmatrix} = \begin{vmatrix} \mathbf{X}'\mathbf{P}\mathbf{X} & \mathbf{X}'\mathbf{P}\mathbf{Z} \\ \mathbf{Z}'\mathbf{P}\mathbf{X} & \mathbf{Z}'\mathbf{P}\mathbf{Z} \end{vmatrix}^{-1}_{rs} \begin{vmatrix} \mathbf{X}'\mathbf{P}\mathbf{y} \\ \mathbf{Z}'\mathbf{P}\mathbf{y} \end{vmatrix}. \quad (345.10)$$

Es wird vorausgesetzt, daß die Matrizen \mathbf{X} und $|\mathbf{X}, \mathbf{Z}|$ wegen des speziellen Aufbaus der Matrix \mathbf{Z} identische Spaltenrangdefekte besitzen, dann gilt (ROHDE 1965)

$$\begin{vmatrix} \mathbf{X}'\mathbf{P}\mathbf{X} & \mathbf{X}'\mathbf{P}\mathbf{Z} \\ \mathbf{Z}'\mathbf{P}\mathbf{X} & \mathbf{Z}'\mathbf{P}\mathbf{Z} \end{vmatrix}^{-1}_{rs} = \begin{vmatrix} (\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}_{rs} (\mathbf{I} + \mathbf{X}'\mathbf{P}\mathbf{Z}\mathbf{S}\mathbf{Z}'\mathbf{P}\mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}_{rs}) & -(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}_{rs} \mathbf{X}'\mathbf{P}\mathbf{Z}\mathbf{S} \\ -\mathbf{S}\mathbf{Z}'\mathbf{P}\mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}_{rs} & \mathbf{S} \end{vmatrix} \quad (345.11)$$

mit

$$\mathbf{S} = [\mathbf{Z}'(\mathbf{P} - \mathbf{P}\mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}_{rs}\mathbf{X}'\mathbf{P})\mathbf{Z}]^{-1} = (\mathbf{Z}'\mathbf{P}\mathbf{Q}_e\mathbf{P}\mathbf{Z})^{-1}$$

worin \mathbf{Q}_e in (345.5) definiert ist. Mit (345.10) und (345.11) folgt dann

$$\hat{\boldsymbol{\epsilon}} = \mathbf{S}\mathbf{Z}'\mathbf{P}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{P}\mathbf{X})^{-1}_{rs}\mathbf{X}'\mathbf{P})\mathbf{y}$$

und schließlich mit (345.4)

$$\hat{\epsilon} = -(\mathbf{Z}'\mathbf{P}\mathbf{Q}_e\mathbf{P}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}\hat{\epsilon}. \quad (345.12)$$

Die Schätzwerte $\hat{\epsilon}$ für die Ausreißer ϵ im Modell (345.7) lassen sich also durch den Vektor $\hat{\epsilon}$ der Residuen des Modells (345.1) und durch die Matrix \mathbf{Q}_e in $D(\hat{\epsilon})$ ausdrücken.

35 Verallgemeinerte lineare Modelle

351 Regressionsmodell

Bislang wurden die unbekannt Parameter als feste Größen behandelt, im folgenden sollen sie als Zufallsvariable eingeführt werden.

Definition: Es sei γ ein $u \times 1$ Vektor unbekannter Parameter, die als Zufallsvariablen definiert sind und deren $u \times 1$ Erwartungswertvektor $E(\gamma) = \mu_\gamma$ sowie deren positiv definite $u \times u$ Kovarianzmatrix $D(\gamma) = \Sigma_{\gamma\gamma}$ gegeben seien. Ferner sei \mathbf{y} der $n \times 1$ Zufallsvektor der Beobachtungen. Bekannt seien der $n \times 1$ Erwartungswertvektor $E(\mathbf{y}) = \mu_y$ von \mathbf{y} , die positiv definite $n \times n$ Kovarianzmatrix $D(\mathbf{y}) = \Sigma_{yy}$ von \mathbf{y} und die Kovarianzmatrix $C(\gamma, \mathbf{y}) = \Sigma_{\gamma y}$ von γ und \mathbf{y} . Dann bezeichnet man die Schätzung eines Zufallsparameters γ_i mit $\gamma = (\gamma_i)$ oder einer linearen Funktion der Parameter γ , beispielsweise $\mathbf{b}'\gamma$, durch eine Konstante und durch eine lineare Funktion der Beobachtungen \mathbf{y} , beispielsweise durch $d + \mathbf{c}'\mathbf{y}$, als *lineare Regression* und die Gesamtheit der Voraussetzungen als *Regressionsmodell*. (351.1)

Im Gegensatz zum Gauß–Markoff–Modell oder zur Regressionsanalyse (321.1), in der lineare Funktionen fester Parameter als lineare Funktionen der Beobachtungen geschätzt werden, sind im Regressionsmodell (351.1) lineare Funktionen von Zufallsparametern zu schätzen, was einer Vorhersage der Zufallsparameter durch die Beobachtungen entspricht. Hierzu soll die beste lineare erwartungstreue Schätzung benutzt werden, mit der die lineare Funktion $\mathbf{b}'\gamma$ der unbekannt Parameter γ durch die lineare Funktion $\mathbf{c}'\mathbf{y}$ der Beobachtungen \mathbf{y} geschätzt wird, zu der wegen (351.1) die Konstante d zu addieren ist, wobei der $u \times 1$ Vektor \mathbf{b} gegeben und der $n \times 1$ Vektor \mathbf{c} und die Konstante d zu bestimmen sind. Für eine beste lineare erwartungstreue Schätzung ist nach (312.3) zu fordern

$$1) \quad E(\mathbf{c}'\mathbf{y} + d) = E(\mathbf{b}'\gamma), \quad (351.2)$$

woraus mit (351.1)

$$\mathbf{c}'\mu_y + d - \mathbf{b}'\mu_\gamma = 0 \quad (351.3)$$

folgt. Ferner muß

$$2) \quad V(\mathbf{c}'\mathbf{y} + d - \mathbf{b}'\boldsymbol{\gamma}) \quad \text{minimal}$$

werden. Mit (233.13), (233.15) und $\boldsymbol{\Sigma}_{y\gamma} = \boldsymbol{\Sigma}'_{\gamma y}$ erhält man

$$V(\mathbf{c}'\mathbf{y} + d - \mathbf{b}'\boldsymbol{\gamma}) = \mathbf{c}'\boldsymbol{\Sigma}_{yy}\mathbf{c} - 2\mathbf{c}'\boldsymbol{\Sigma}_{y\gamma}\mathbf{b} + \mathbf{b}'\boldsymbol{\Sigma}_{\gamma\gamma}\mathbf{b}. \quad (351.4)$$

Diese Varianz ist unter der Bedingung (351.3) zu minimieren. Bildet man hierzu nach (171.6) die Lagrange'sche Funktion $w(\mathbf{c}, d)$ mit dem Lagrange'schen Multiplikator $-2k$, ergibt sich $w(\mathbf{c}, d) = \mathbf{c}'\boldsymbol{\Sigma}_{yy}\mathbf{c} - 2\mathbf{c}'\boldsymbol{\Sigma}_{y\gamma}\mathbf{b} + \mathbf{b}'\boldsymbol{\Sigma}_{\gamma\gamma}\mathbf{b} - 2k(\mathbf{c}'\boldsymbol{\mu}_y + d - \mathbf{b}'\boldsymbol{\mu}_\gamma)$. Für $\partial w(\mathbf{c}, d)/\partial \mathbf{c} = \mathbf{0}$ und $\partial w(\mathbf{c}, d)/\partial d = 0$ folgen mit (172.1) und (172.2) die Beziehungen $\boldsymbol{\Sigma}_{yy}\mathbf{c} - \boldsymbol{\Sigma}_{y\gamma}\mathbf{b} - k\boldsymbol{\mu}_y = \mathbf{0}$ und $k = 0$ sowie mit (351.3)

$$\mathbf{c} = \boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{y\gamma}\mathbf{b} \quad \text{und} \quad d = \mathbf{b}'\boldsymbol{\mu}_\gamma - \mathbf{c}'\boldsymbol{\mu}_y. \quad (351.5)$$

Der Vektor \mathbf{c} und die Konstante d sind eindeutig bestimmt, da nach Voraussetzung $\boldsymbol{\Sigma}_{yy}$ positiv definit ist, so daß $\boldsymbol{\Sigma}_{yy}^{-1}$ wegen (143.3) und (143.9) existiert.

Mit (351.5) ergibt sich anstelle von (351.4) die Varianz

$$V(\mathbf{c}'\mathbf{y} + d - \mathbf{b}'\boldsymbol{\gamma}) = \mathbf{b}'\boldsymbol{\Sigma}_{\gamma\gamma}\mathbf{b} - \mathbf{b}'\boldsymbol{\Sigma}_{\gamma y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{y\gamma}\mathbf{b}, \quad (351.6)$$

von der jetzt gezeigt werden soll, daß sie minimal ist, also die zweite Bedingung in (312.3) $V(\mathbf{c}'\mathbf{y} + d - \mathbf{b}'\boldsymbol{\gamma}) \leq V(\mathbf{c}'^*\mathbf{y} + d^* - \mathbf{b}'\boldsymbol{\gamma})$ erfüllt, worin $\mathbf{c}'^*\mathbf{y} + d^*$ eine beliebige erwartungstreue Schätzung von $\mathbf{b}'\boldsymbol{\gamma}$ bedeutet. Man erhält mit (233.13)

$$\begin{aligned} V(\mathbf{c}'^*\mathbf{y} + d^* - \mathbf{b}'\boldsymbol{\gamma}) &= V(\mathbf{c}'^*\mathbf{y} + d^* - \mathbf{c}'\mathbf{y} - d + \mathbf{c}'\mathbf{y} + d - \mathbf{b}'\boldsymbol{\gamma}) \\ &= V((\mathbf{c}^* - \mathbf{c})'\mathbf{y} + d^* - d) + V(\mathbf{c}'\mathbf{y} + d - \mathbf{b}'\boldsymbol{\gamma}) + 2f, \end{aligned}$$

woraus die obige Bedingung resultiert, da für die Varianz $V((\mathbf{c}^* - \mathbf{c})'\mathbf{y} + d^* - d) \geq 0$ gilt und mit (233.15) und (233.17) folgt

$$f = (\mathbf{c}^* - \mathbf{c})'D(\mathbf{y})\mathbf{c} - (\mathbf{c}^* - \mathbf{c})'C(\mathbf{y}, \boldsymbol{\gamma})\mathbf{b} = 0.$$

Bezeichnet man mit $\widehat{\mathbf{b}'\boldsymbol{\gamma}}$ die beste lineare erwartungstreue Schätzung von $\mathbf{b}'\boldsymbol{\gamma}$, ergibt sich mit $\widehat{\mathbf{b}'\boldsymbol{\gamma}} = \mathbf{c}'\mathbf{y} + d$ aus (351.5)

$$\widehat{\mathbf{b}'\boldsymbol{\gamma}} = \mathbf{b}'(\boldsymbol{\mu}_\gamma + \boldsymbol{\Sigma}_{\gamma y}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)) \quad (351.7)$$

und aus (351.6) die minimale Varianz

$$V(\widehat{\mathbf{b}'\boldsymbol{\gamma}} - \mathbf{b}'\boldsymbol{\gamma}) = \mathbf{b}'\boldsymbol{\Sigma}_{\gamma\gamma}\mathbf{b} - \mathbf{b}'\boldsymbol{\Sigma}_{\gamma y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{y\gamma}\mathbf{b}. \quad (351.8)$$

Die Varianz von $\widehat{\mathbf{b}'\boldsymbol{\gamma}}$ berechnet sich mit (233.2) aus (351.7) zu

$$V(\widehat{\mathbf{b}'\boldsymbol{\gamma}}) = \mathbf{b}'\boldsymbol{\Sigma}_{\boldsymbol{\gamma}y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{y\boldsymbol{\gamma}}\mathbf{b} \quad (351.9)$$

und die Kovarianz von $\mathbf{b}'\boldsymbol{\gamma}$ und $\widehat{\mathbf{b}'\boldsymbol{\gamma}}$ mit (233.15) zu

$$C(\mathbf{b}'\boldsymbol{\gamma}, \widehat{\mathbf{b}'\boldsymbol{\gamma}}) = \mathbf{b}'\boldsymbol{\Sigma}_{\boldsymbol{\gamma}y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{y\boldsymbol{\gamma}}\mathbf{b}. \quad (351.10)$$

Setzt man $\mathbf{b}' = |1, 0, 0, \dots|$, $\mathbf{b}' = |0, 1, 0, \dots|$ und so fort, ergeben sich aus (351.7) die Schätzwerte $\hat{\boldsymbol{\gamma}}$ von $\boldsymbol{\gamma}$ und damit $D(\hat{\boldsymbol{\gamma}})$ aus (233.2), $C(\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}})$ aus (233.15) und schließlich $D(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ aus (233.13). Somit folgt der

Satz: Die beste lineare erwartungstreue Schätzung $\hat{\boldsymbol{\gamma}}$ der unbekanntem Zufallsparameter $\boldsymbol{\gamma}$ im Regressionsmodell (351.1) ist gegeben durch

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{\mu}_{\boldsymbol{\gamma}} + \boldsymbol{\Sigma}_{\boldsymbol{\gamma}y}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)$$

und die Kovarianzmatrizen $D(\hat{\boldsymbol{\gamma}})$ und $C(\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}})$ durch

$$D(\hat{\boldsymbol{\gamma}}) = C(\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}) = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{y\boldsymbol{\gamma}}$$

sowie die Kovarianzmatrix $D(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$, die minimale Spur besitzt, durch

$$D(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} - \boldsymbol{\Sigma}_{\boldsymbol{\gamma}y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{y\boldsymbol{\gamma}}. \quad (351.11)$$

Die Schätzung der Zufallsparameter $\boldsymbol{\gamma}$ nach (351.11) ist auf Ausnahmefälle beschränkt, da zur Berechnung von $\hat{\boldsymbol{\gamma}}$ die Erwartungswertvektoren $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ und $\boldsymbol{\mu}_y$ bekannt sein müssen. Dieser Nachteil ist in dem Modell der Prädiktion und Filterung des Kapitels 356 behoben. Mit (351.11) übereinstimmende Ergebnisse erhält man auch durch die im Kapitel 312 definierte beste lineare Schätzung, die darin besteht, daß der Erwartungswert $E((\mathbf{c}'\mathbf{y} + d - \mathbf{b}'\boldsymbol{\gamma})^2)$ minimiert wird. Man kann $\hat{\boldsymbol{\gamma}}$ außerdem mit Hilfe der bedingten Normalverteilung ableiten. Stammen nämlich die Zufallsvektoren $\boldsymbol{\gamma}$ und \mathbf{y} aus der Zerlegung eines normalverteilten Zufallsvektors, so gleicht $\hat{\boldsymbol{\gamma}}$ nach (253.2) dem Erwartungswertvektor der Normalverteilung des Zufallsvektors $\boldsymbol{\gamma}$ unter der Bedingung, daß der Zufallsvektor \mathbf{y} die gemessenen Werte annimmt.

Die Elemente der nach (233.8) aus $D(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ folgenden Korrelationsmatrix bezeichnet man als *partielle Korrelationskoeffizienten* und die Elemente der Korrelationsmatrix aus $C(\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}})$ als *multiple Korrelationskoeffizienten*. Auf Verfahren zur Schätzung und zum Test dieser Korrelationskoeffizienten (ANDERSON 1958, S.60; GRAYBILL 1976, S.390) wird hier nicht eingegangen, jedoch wird die Schätzung von Kovarianzmatrizen im Kapitel 375 behandelt.

352 Gemischtes Modell

Ersetzt man den Fehlervektor \mathbf{e} im Gauß–Markoff–Modell (321.2) durch eine Linearkombination $-\mathbf{Z}\boldsymbol{\gamma}$ unbekannter Parameter $\boldsymbol{\gamma}$, die als Zufallsvariable definiert sind, ergibt sich das folgende Modell mit festen Parametern und Zufallsparametern.

Definition: Es seien \mathbf{X} eine $n \times u$ und \mathbf{Z} eine $n \times r$ Matrix bekannter Koeffizienten mit $\text{rg}\mathbf{Z} = n$, $\boldsymbol{\beta}$ ein $u \times 1$ Vektor unbekannter, fester Parameter, $\boldsymbol{\gamma}$ ein $r \times 1$ Vektor unbekannter Zufallsparameter mit $E(\boldsymbol{\gamma}) = \mathbf{0}$ und $D(\boldsymbol{\gamma}) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$, wobei der positive Faktor σ^2 unbekannt und die $r \times r$ Matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$ gegeben sowie positiv definit sei, und \mathbf{y} ein $n \times 1$ Beobachtungsvektor, dann bezeichnet man

$$\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} = \mathbf{y} \quad \text{mit} \quad E(\boldsymbol{\gamma}) = \mathbf{0} \quad \text{und} \quad D(\boldsymbol{\gamma}) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}$$

als *gemischtes* Modell. (352.1)

Ein Anwendungsbeispiel für dieses Modell, verbunden mit einem Rechenbeispiel für die Parameterschätzung, befindet sich im Kapitel 354.

Denkt man sich den Beobachtungsvektor \mathbf{y} durch die lineare Transformation $\mathbf{y} = \mathbf{Z}\bar{\mathbf{y}} + \mathbf{z}$ entstanden, in der $\bar{\mathbf{y}}$ ein $r \times 1$ Vektor von Beobachtungen und \mathbf{z} ein $n \times 1$ Vektor von Konstanten bedeuten, so läßt sich der Vektor $\boldsymbol{\gamma}$ als Fehlervektor von $\bar{\mathbf{y}}$ interpretieren. Das durch die Transformation gewonnene Modell $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} = \mathbf{y}$ wurde als Allgemeinfeld der Ausgleichsrechnung von HELMERT (1872, S.215) eingeführt und daher auch als *Gauß–Helmert–Modell* bezeichnet (WOLF 1978A). Eine Verallgemeinerung des gemischten Modells befindet sich in (SCHAFFRIN 1985, S.26).

Die Kovarianzmatrix $D(\mathbf{y})$ der Beobachtungen \mathbf{y} ergibt sich aus (233.2) zu

$$D(\mathbf{y}) = D(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}) = \sigma^2 \mathbf{Z}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathbf{Z}' = \sigma^2 \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}}, \quad (352.2)$$

worin $\boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}} = \mathbf{Z}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathbf{Z}'$ wegen (143.7) positiv definit ist, da \mathbf{Z} vollen Zeilenrang besitzt. Die Kovarianzmatrix $C(\boldsymbol{\gamma}, \mathbf{y})$ von $\boldsymbol{\gamma}$ und \mathbf{y} erhält man mit (233.15) zu

$$C(\boldsymbol{\gamma}, \mathbf{y}) = C(\boldsymbol{\gamma}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathbf{Z}' = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\mathbf{y}}. \quad (352.3)$$

Zur Parameterschätzung im gemischten Modell (352.1) wird zunächst $-\mathbf{Z}\boldsymbol{\gamma} = \mathbf{e}$ gesetzt, so daß sich mit (233.2) und (352.2) $D(\mathbf{e}) = \sigma^2 \mathbf{Z}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\mathbf{Z}' = \sigma^2 \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}}$ ergibt. Damit erhält man das Modell

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{y} + \mathbf{e} \quad \text{mit} \quad E(\mathbf{e}) = \mathbf{0} \quad \text{und} \quad D(\mathbf{e}) = D(\mathbf{y}) = \sigma^2 \boldsymbol{\Sigma}_{\mathbf{y}\mathbf{y}}, \quad (352.4)$$

das mit dem Gauß–Markoff–Modell (321.2) identisch ist, so daß als beste lineare erwartungstreue Schätzung $\hat{\beta}$ von β aus (322.9) folgt

$$\hat{\beta} = (\mathbf{X}'\Sigma_{yy}^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_{yy}^{-1}\mathbf{y}. \quad (352.5)$$

Die Schätzung der festen Parameter β im gemischten Modell (352.1) unterscheidet sich also nicht von der Parameterschätzung im Gauß–Markoff–Modell. Besitzt die Koeffizientenmatrix \mathbf{X} keinen vollen Spaltenrang, gewinnt man nach (333.11) die erwartungstreuen Schätzwerte projizierter Parameter, indem die inverse Normalgleichungsmatrix in (352.5) durch $(\mathbf{X}'\Sigma_{yy}^{-1}\mathbf{X})_{rs}^-$ oder $(\mathbf{X}'\Sigma_{yy}^{-1}\mathbf{X})^+$ ersetzt wird. Die erwartungstreue Schätzung $\hat{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit folgt aus (325.6) oder (331.12).

Falls in (352.1) der Vektor β bekannt ist, ergibt sich

$$\mathbf{Z}\gamma = \mathbf{y} - \mathbf{X}\beta = \bar{\mathbf{y}} \quad \text{mit} \quad E(\gamma) = \mathbf{0}, \quad E(\bar{\mathbf{y}}) = \mathbf{0}. \quad (352.6)$$

In diesem Modell soll die beste lineare erwartungstreue Schätzung von $\mathbf{b}'\gamma$ durch $\mathbf{c}'\bar{\mathbf{y}} + d$ nach (312.3) bestimmt werden, wobei der $r \times 1$ Vektor \mathbf{b} gegeben und der $n \times 1$ Vektor \mathbf{c} sowie die Konstante d zu bestimmen sind. Aus der Forderung nach Erwartungstreue folgt $d = 0$, und die Minimierung der Varianz $V(\mathbf{c}'\bar{\mathbf{y}} - \mathbf{b}'\gamma)$ wurde bereits mit (351.5) gelöst. Setzt man $\mathbf{b}' = |1, 0, 0, \dots|$, $\mathbf{b}' = |0, 1, 0, \dots|$ und so weiter, ergibt sich mit $D(\bar{\mathbf{y}}) = D(\mathbf{y})$ wegen (233.2) und $C(\gamma, \bar{\mathbf{y}}) = C(\gamma, \mathbf{y})$ wegen (233.15) aus (351.11)

$$\gamma^* = \Sigma_{\gamma y} \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{X}\beta), \quad (352.7)$$

falls γ^* die beste lineare erwartungstreue Schätzung von γ bedeutet. Da aber β in (352.1) nicht als bekannt vorausgesetzt werden darf, soll β in (352.7) durch seinen Schätzwert $\hat{\beta}$ aus (352.5) ersetzt werden, so daß sich die Schätzung $\hat{\gamma}$ von γ ergibt zu

$$\hat{\gamma} = \Sigma_{\gamma y} \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}). \quad (352.8)$$

Diese Schätzung ist zulässig, da sie zusammen mit (352.5) das Modell (352.1) erfüllt, denn mit $\mathbf{Z}\Sigma_{\gamma y} = \Sigma_{yy}$ aus (352.2) und (352.3) folgt

$$\mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\gamma} = \mathbf{y}. \quad (352.9)$$

Die Schätzung für γ ist aber auch wie die für β die beste erwartungstreue Schätzung, wie im folgenden Kapitel gezeigt wird. Um $\hat{\gamma}$ zu berechnen, braucht, wenn $\Sigma_{\gamma y}$ gegeben ist, \mathbf{Z} nicht bekannt zu sein.

353 Beste lineare erwartungstreue Schätzung im gemischten Modell

Zur Vereinfachung der Ableitungen in den folgenden Kapiteln wird voller Spaltenrang für die Koeffizientenmatrix \mathbf{X} im gemischten Modell (352.1) vorausgesetzt; gemischte Modelle mit beliebigem Rang für \mathbf{X} und singulärer Kovarianzmatrix für die Beobachtungen werden bei (HARVILLE 1976) behandelt. Die Eigenschaften der Schätzwerte folgen dann aus dem

Satz: Im gemischten Modell (352.1) ist die beste lineare erwartungstreue Schätzung von $\mathbf{a}'\boldsymbol{\beta} + \mathbf{b}'\boldsymbol{\gamma}$, worin \mathbf{a} und \mathbf{b} bekannte $u \times 1$ und $r \times 1$ Vektoren bedeuten, durch $\mathbf{a}'\hat{\boldsymbol{\beta}} + \mathbf{b}'\hat{\boldsymbol{\gamma}}$ mit $\hat{\boldsymbol{\beta}}$ und $\hat{\boldsymbol{\gamma}}$ aus (352.5) und (352.8) gegeben. (353.1)

Beweis: Die Linearkombination $\mathbf{a}'\boldsymbol{\beta} + \mathbf{b}'\boldsymbol{\gamma}$ werde mit (352.5) und (352.8) durch $\mathbf{a}'\hat{\boldsymbol{\beta}} + \mathbf{b}'\hat{\boldsymbol{\gamma}} = \mathbf{c}'\mathbf{y} + d$ geschätzt, worin \mathbf{c} ein $n \times 1$ Vektor und d eine Konstante bedeuten, denn es gilt $\mathbf{a}'\hat{\boldsymbol{\beta}} + \mathbf{b}'\hat{\boldsymbol{\gamma}} = \mathbf{a}'(\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{y} + \mathbf{b}'\boldsymbol{\Sigma}_{\gamma y}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1})\mathbf{y}$ und daher

$$\mathbf{c}' = \mathbf{a}'(\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1} + \mathbf{b}'\boldsymbol{\Sigma}_{\gamma y}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1})$$

sowie $d = 0$. (353.2)

Zu zeigen ist, daß $\mathbf{c}'\mathbf{y} + d$ beste lineare erwartungstreue Schätzung von $\mathbf{a}'\boldsymbol{\beta} + \mathbf{b}'\boldsymbol{\gamma}$ ist. Die Schätzung ist erwartungstreu, denn mit $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ aus (352.1) folgt aus (353.2) $E(\mathbf{c}'\mathbf{y} + d) = \mathbf{a}'\boldsymbol{\beta} = E(\mathbf{a}'\boldsymbol{\beta} + \mathbf{b}'\boldsymbol{\gamma})$, so daß die erste Bedingung in (312.3) erfüllt ist. Die zweite Bedingung fordert $V(\mathbf{c}'\mathbf{y} + d - \mathbf{a}'\boldsymbol{\beta} - \mathbf{b}'\boldsymbol{\gamma}) \leq V(\mathbf{c}^*\mathbf{y} + d^* - \mathbf{a}'\boldsymbol{\beta} - \mathbf{b}'\boldsymbol{\gamma})$, worin $\mathbf{c}^*\mathbf{y} + d^*$ eine beliebige erwartungstreue Schätzung von $\mathbf{a}'\boldsymbol{\beta} + \mathbf{b}'\boldsymbol{\gamma}$ bedeutet, also $E(\mathbf{c}^*\mathbf{y} + d^*) = E(\mathbf{c}'\mathbf{y} + d) = E(\mathbf{a}'\boldsymbol{\beta} + \mathbf{b}'\boldsymbol{\gamma})$ oder $\mathbf{c}^*\mathbf{X}\boldsymbol{\beta} + d^* = \mathbf{c}'\mathbf{X}\boldsymbol{\beta} + d = \mathbf{a}'\boldsymbol{\beta}$. Da diese Gleichung für alle $\boldsymbol{\beta}$ gilt, folgt

$$\mathbf{c}^*\mathbf{X} = \mathbf{a}' \quad \text{und} \quad d^* = d = 0. \quad (353.3)$$

Weiter erhält man mit (233.13) $V(\mathbf{c}^*\mathbf{y} + d^* - \mathbf{a}'\boldsymbol{\beta} - \mathbf{b}'\boldsymbol{\gamma}) = V(\mathbf{c}^*\mathbf{y} + d^* - \mathbf{c}'\mathbf{y} - d + \mathbf{c}'\mathbf{y} + d - \mathbf{a}'\boldsymbol{\beta} - \mathbf{b}'\boldsymbol{\gamma}) = V((\mathbf{c}^* - \mathbf{c})'\mathbf{y}) + V(\mathbf{c}'\mathbf{y} + d - \mathbf{a}'\boldsymbol{\beta} - \mathbf{b}'\boldsymbol{\gamma}) + 2f$, woraus die obige Bedingung resultiert, da für die Varianz $V((\mathbf{c}^* - \mathbf{c})'\mathbf{y}) \geq 0$ gilt und da mit (233.15) und (233.17) folgt $f = (\mathbf{c}^* - \mathbf{c})'D(\mathbf{y})\mathbf{c} - (\mathbf{c}^* - \mathbf{c})'C(\mathbf{y}, \boldsymbol{\gamma})\mathbf{b}$ und wegen $\boldsymbol{\Sigma}_{y\gamma} = \boldsymbol{\Sigma}'_{\gamma y}$, (352.2), (352.3) und (353.2) $f = \sigma^2(\mathbf{c}^* - \mathbf{c})'[\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{X})^{-1}\mathbf{a} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{y\gamma}\mathbf{b}]$ und schließlich mit (353.3) $f = 0$, so daß die Aussage folgt.

Setzt man in (352.1) $r = n$ und $\mathbf{Z} = \mathbf{I}_n$, ergibt sich mit $\boldsymbol{\gamma} = -\mathbf{e}$ das Gauß-Markoff-Modell (321.2). In diesem Fall findet man als Schätzung für $\hat{\boldsymbol{\gamma}}$ aus

(352.2), (352.3) und (352.8) $\hat{\gamma} = \Sigma_{yy} \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}) = -\hat{\mathbf{e}}$ in Übereinstimmung mit (323.4). Das Gauß–Markoff–Modell läßt sich daher auch als gemischtes Modell ansehen, in dem feste Parameter und Zufallsparameter zu schätzen sind, so daß das gemischte Modell eine Verallgemeinerung des Gauß–Markoff–Modells darstellt.

Mit den Schätzwerten $\hat{\beta}$ und $\hat{\gamma}$ aus (352.5) und (352.8) ergeben sich die folgenden Kovarianzmatrizen. Mit (233.2) und (352.2) folgt

$$D(\hat{\beta} - \beta) = D(\hat{\beta}) = \sigma^2 (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \quad (353.4)$$

in Übereinstimmung mit (322.9). Aus (233.2) ergibt sich mit $\hat{\gamma} = \Sigma_{\gamma y} \Sigma_{yy}^{-1} (\mathbf{I} - \mathbf{X} (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{yy}^{-1}) \mathbf{y}$ die Kovarianzmatrix $D(\hat{\gamma})$ und entsprechend aus (233.15) die Kovarianzmatrix $C(\hat{\gamma}, \gamma)$ mit $\Sigma_{y\gamma} = \Sigma'_{\gamma y}$ zu

$$\begin{aligned} D(\hat{\gamma}) &= \sigma^2 (\Sigma_{\gamma y} \Sigma_{yy}^{-1} \Sigma_{y\gamma} - \Sigma_{\gamma y} \Sigma_{yy}^{-1} \mathbf{X} (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{yy}^{-1} \Sigma_{y\gamma}) \\ &= C(\hat{\gamma}, \gamma) . \end{aligned} \quad (353.5)$$

Da diese Matrix sich mit $\Sigma_{yy}^{-1} = \mathbf{G} \mathbf{G}'$ wegen (143.5) als Produkt mit der Matrix $\mathbf{I} - \mathbf{G}' \mathbf{X} (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{G}$ darstellen läßt, deren Rang wegen (152.3) mit (152.4) $n - u$ beträgt, sind $D(\hat{\gamma})$ und $C(\hat{\gamma}, \gamma)$ singulär. Weiter folgt mit (233.13) und (353.5)

$$\begin{aligned} D(\hat{\gamma} - \gamma) &= \sigma^2 (\Sigma_{\gamma\gamma} - \Sigma_{\gamma y} \Sigma_{yy}^{-1} \Sigma_{y\gamma} \\ &\quad + \Sigma_{\gamma y} \Sigma_{yy}^{-1} \mathbf{X} (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{yy}^{-1} \Sigma_{y\gamma}) . \end{aligned} \quad (353.6)$$

Ferner ergibt sich aus (233.15)

$$C(\hat{\beta}, \hat{\gamma}) = \mathbf{0} \quad (353.7)$$

und aus (233.15) und (233.17)

$$C(\hat{\beta} - \beta, \hat{\gamma} - \gamma) = C(\hat{\beta}, \hat{\gamma} - \gamma) = -\sigma^2 (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{yy}^{-1} \Sigma_{y\gamma} . \quad (353.8)$$

Mit (353.4), (353.6) und (353.8) erhält man aus (233.13) für gegebene Matrizen \mathbf{A} und \mathbf{B} entsprechender Dimensionen die Kovarianzmatrix $D(\mathbf{A} \hat{\beta} + \mathbf{B}(\hat{\gamma} - \gamma))$, die wegen (353.1) minimale Spur besitzt.

Für die numerische Berechnung der Schätzwerte $\hat{\beta}$ und $\hat{\gamma}$ sowie ihrer Kovarianzmatrizen eignen sich die im folgenden Kapitel angegebenen Rechenformeln.

354 Methode der kleinsten Quadrate und Maximum–Likelihood–Methode für das gemischte Modell

Wie schon bei (352.1) erläutert, läßt sich der Vektor γ als Fehlervektor interpretieren, so daß zur Parameterschätzung im gemischten Modell (352.1) nach der Methode (313.1) der kleinsten Quadrate die quadratische Form $\gamma' \Sigma_{\gamma\gamma}^{-1} \gamma / \sigma^2$ der unbekannt Parameter γ unter der Bedingung $\mathbf{X}\beta + \mathbf{Z}\gamma = \mathbf{y}$ zu minimieren ist. Hierfür wird mit der Lagrangeschen Funktion $w(\beta, \gamma) = \gamma' \Sigma_{\gamma\gamma}^{-1} \gamma / \sigma^2 - 2\mathbf{k}'(\mathbf{X}\beta + \mathbf{Z}\gamma - \mathbf{y}) / \sigma^2$ aus (171.6), in der $-2\mathbf{k}' / \sigma^2$ der $n \times 1$ Vektor der Lagrangeschen Multiplikatoren bedeutet, mit (172.1) und (172.2) $\partial w(\beta, \gamma) / \partial \gamma = 2\Sigma_{\gamma\gamma}^{-1} \gamma / \sigma^2 - 2\mathbf{Z}'\mathbf{k} / \sigma^2 = \mathbf{0}$ und $\partial w(\beta, \gamma) / \partial \beta = -2\mathbf{X}'\mathbf{k} / \sigma^2 = \mathbf{0}$ gebildet. Damit folgt

$$\hat{\gamma} = \Sigma_{\gamma\gamma} \mathbf{Z}' \mathbf{k} \quad (354.1)$$

und mit (352.1) die Normalgleichungen

$$\left| \begin{array}{cc|c} \mathbf{Z}\Sigma_{\gamma\gamma}\mathbf{Z}' & \mathbf{X} & \mathbf{k} \\ \mathbf{X}' & \mathbf{0} & \hat{\beta} \end{array} \right| = \left| \begin{array}{c} \mathbf{y} \\ \mathbf{0} \end{array} \right|. \quad (354.2)$$

Würde dagegen die quadratische Form $(\mathbf{y} - \mathbf{X}\beta)' \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{X}\beta) / \sigma^2 = \gamma' \mathbf{Z}' \Sigma_{yy}^{-1} \mathbf{Z}\gamma / \sigma^2$ unter der Bedingung $\mathbf{X}\beta + \mathbf{Z}\gamma = \mathbf{y}$ minimiert, müßte zur Berechnung der Schätzwerte für γ die Matrix $\mathbf{Z}' \Sigma_{yy}^{-1} \mathbf{Z}$ invertiert werden, die aber singular ist.

Da für die Koeffizientenmatrix \mathbf{X} voller Rang vorausgesetzt wurde, ist die Normalgleichungsmatrix auf der linken Seite von (354.2) regulär, was wie für (327.9) gezeigt werden kann, so daß $\hat{\beta}$ sowie \mathbf{k} und damit $\hat{\gamma}$ aus (354.1) eindeutig bestimmt werden. Man erhält mit (134.8) und (352.2) $\hat{\beta} = (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{yy}^{-1} \mathbf{y}$ in Übereinstimmung mit (352.5) und weiter $\mathbf{k} = \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta})$ sowie mit (352.3) aus (354.1) $\hat{\gamma} = \Sigma_{\gamma\gamma} \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta})$ in Übereinstimmung mit (352.8).

Nimmt man an, daß der Zufallsvektor γ normalverteilt ist, so erhält man aus (251.1) mit (352.1) $\gamma \sim N(\mathbf{0}, \sigma^2 \Sigma_{\gamma\gamma})$. Bei der Maximum–Likelihood–Methode zur Parameterschätzung nach (314.2) ist dann, abgesehen von dem Faktor 1/2, die gleiche quadratische Form unter den gleichen Bedingungen wie bei der Methode der kleinsten Quadrate zu minimieren, so daß sich identische Schätzwerte ergeben.

Die Inverse der Normalgleichungsmatrix in (354.2) berechnet sich mit (134.3) und (352.2) zu

$$\left| \begin{array}{cc|c} \mathbf{Z}\Sigma_{\gamma\gamma}\mathbf{Z}' & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{array} \right|^{-1} = \left| \begin{array}{cc|c} \Sigma_{yy} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{array} \right|^{-1} = \left| \begin{array}{cc} \Sigma_{kk} & \Sigma_{k\beta} \\ \Sigma_{\beta k} & \Sigma_{\beta\beta} \end{array} \right| \quad (354.3)$$

mit

$$\begin{aligned}\Sigma'_{k\beta} &= \Sigma_{\beta k}, \quad \Sigma_{kk} = \Sigma_{yy}^{-1} - \Sigma_{yy}^{-1} \mathbf{X} (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{yy}^{-1} \\ \Sigma_{\beta k} &= (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{yy}^{-1} \quad \text{und} \quad \Sigma_{\beta\beta} = -(\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1}.\end{aligned}$$

Mit der inversen Normalgleichungsmatrix ergeben sich daher die Kovarianzmatrizen (353.4), (353.5) und (353.8) wegen (352.3) zu

$$\begin{aligned}D(\hat{\beta}) &= -\sigma^2 \Sigma_{\beta\beta}, \quad D(\hat{\gamma}) = -\sigma^2 \Sigma_{\gamma\gamma} \mathbf{Z}' \Sigma_{kk} \mathbf{Z} \Sigma_{\gamma\gamma} \\ \text{und} \quad C(\hat{\beta}, \hat{\gamma} - \gamma) &= -\sigma^2 \Sigma_{\beta k} \mathbf{Z} \Sigma_{\gamma\gamma}.\end{aligned}\quad (354.4)$$

Wie bereits bei (352.5) erwähnt, ergibt sich die erwartungstreue Schätzung $\hat{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit zu

$$\hat{\sigma}^2 = \Omega / (n - u) \quad (354.5)$$

mit $\Omega = (\mathbf{X} \hat{\beta} - \mathbf{y})' \Sigma_{yy}^{-1} (\mathbf{X} \hat{\beta} - \mathbf{y}) = \hat{\gamma}' \mathbf{Z}' \Sigma_{yy}^{-1} \mathbf{Z} \hat{\gamma}$ wegen (325.1) und (352.9). Weiter folgt mit (352.2) und (354.1) $\Omega = \mathbf{k}' \mathbf{Z} \Sigma_{\gamma\gamma} \mathbf{Z}' \Sigma_{yy}^{-1} \mathbf{Z} \hat{\gamma} = \mathbf{k}' \mathbf{Z} \hat{\gamma} = \hat{\gamma}' \Sigma_{\gamma\gamma}^{-1} \hat{\gamma}$ und somit

$$\Omega = \hat{\gamma}' \mathbf{Z}' \Sigma_{yy}^{-1} \mathbf{Z} \hat{\gamma} = \hat{\gamma}' \Sigma_{\gamma\gamma}^{-1} \hat{\gamma}. \quad (354.6)$$

Ferner erhält man mit (352.9) $\Omega = \mathbf{k}' \mathbf{Z} \hat{\gamma} = \mathbf{k}' (\mathbf{y} - \mathbf{X} \hat{\beta})$ und schließlich mit (354.2)

$$\Omega = \mathbf{y}' \mathbf{k}. \quad (354.7)$$

Durch Erweitern der Normalgleichungsmatrix in (354.2) um die Absolutgliedvektoren \mathbf{y} und $\mathbf{0}$ und um ein zusätzliches Nullelement läßt sich Ω entsprechend (327.24) wegen $0 - |\mathbf{y}', \mathbf{0}' | \mathbf{k}', \hat{\beta}'|' = -\Omega$ durch die Gaußsche Elimination berechnen.

Die Beziehungen (354.1) bis (354.7) stellen geeignete Rechenformeln für die Parameterschätzung im gemischten Modell dar.

Beispiel: Folgendes inverses Problem wird gelöst. Aus Messungen der Anomalien der Vertikalanziehung an der Erdoberfläche soll auf Dichteanomalien in einem begrenzten Teil des Erdinneren geschlossen werden. Der Einfluß der außerhalb des begrenzten Teiles liegenden Masse auf die Anomalien der Vertikalanziehung sei durch eine Polynomentwicklung zu berücksichtigen. Dann bilden die Koeffizienten der Polynome die unbekanntesten, festen Parameter β im Modell (352.1), und die Dichteanomalien werden als Zufallsparameter γ angesetzt, für die $E(\gamma) = \mathbf{0}$ gilt, da sie als Anomalien Abweichungen von einem Durchschnittswert repräsentieren.

An den vier Punkten P_1, P_2, P_3, P_4 auf einem ebenen Teil der Erdoberfläche seien die Anomalien der Vertikalanziehung gemessen worden. Die Lage der Punkte in bezug auf den Ausschnitt der Erdkruste, in dem die Dichteanomalien zu bestimmen sind, ergibt sich aus der Abbildung 354-1. Die Meßwerte betragen 24, 25, 25, 26 in der Dimension $10^{-3} \text{cm sec}^{-2}$. In der Abbildung 354-2 ist eine Seitenansicht des Teils der Erdkruste gezeigt, in der die Dich-

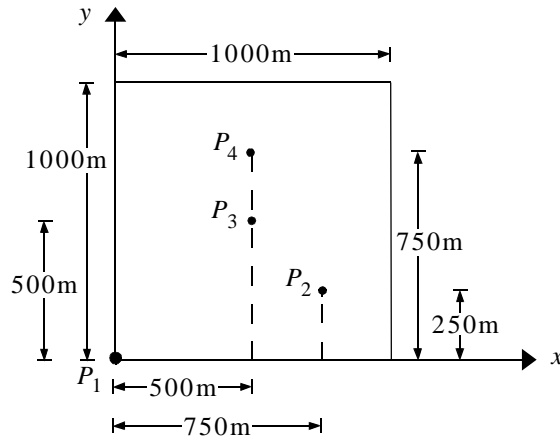


Abb. 354-1: Ausschnitt

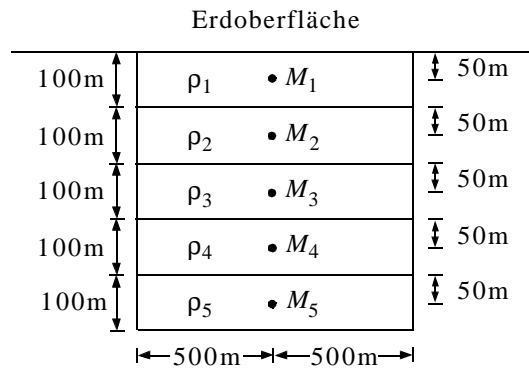


Abb. 354-2: Seitenansicht

teanomalien ρ_1, \dots, ρ_5 von fünf Quadrern zu schätzen sind, die die Dimensionen $1 \text{ km} \cdot 1 \text{ km} \cdot 0,1 \text{ km}$ besitzen und in denen die Dichteanomali-

en ρ_i mit $i \in \{1, \dots, 5\}$ als konstant angenommen werden. Die Anomalien Δg_j mit $j \in \{1, \dots, 4\}$ der Vertikalanziehung werden aus der Näherungsformel $\Delta g_j = \sum_{i=1}^5 k \rho_i h_i \Delta \tau_i / r_{ij}^3$ erhalten, in der h_i der Abstand des Mittelpunktes M_i des Quaders i von der Erdoberfläche, $\Delta \tau_i$ das Volumen des Quaders i , r_{ij} der Abstand des Punktes M_i vom Meßpunkt P_j und $k = 6,67 \cdot 10^{-8} \text{cm}^3 \text{g}^{-1} \text{sec}^{-2}$ die Gravitationskonstante bedeuten.

Die Dichteanomalien seien unkorreliert und ihre Varianzen betragen σ^2 . Als Polynomansatz gelte $a_0 + x_j a_1 + y_j a_2$, wobei (x_j, y_j) die Koordinaten des Punktes P_j in dem in Abbildung 354-1 angegebenen Koordinatensystem bedeuten. Dann ergibt sich folgendes gemischte Modell

$$\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} = \mathbf{y} \quad \text{mit} \quad E(\boldsymbol{\gamma}) = \mathbf{0} \quad \text{und} \quad D(\boldsymbol{\gamma}) = \sigma^2 \mathbf{I} \quad (354.8)$$

sowie

$$\mathbf{X} = \begin{vmatrix} 1 & 0 & 0 \\ 1 & 7,5 & 2,5 \\ 1 & 5,0 & 5,0 \\ 1 & 5,0 & 7,5 \end{vmatrix}, \quad \boldsymbol{\beta} = \begin{vmatrix} 10^1 a_0 \\ 10^5 a_1 \\ 10^5 a_2 \end{vmatrix}, \quad \mathbf{y} = \begin{vmatrix} 0,24 \\ 0,25 \\ 0,25 \\ 0,26 \end{vmatrix}, \quad \boldsymbol{\gamma} = \begin{vmatrix} 10^{-2} \rho_1 \\ 10^{-2} \rho_2 \\ 10^{-2} \rho_3 \\ 10^{-2} \rho_4 \\ 10^{-2} \rho_5 \end{vmatrix}$$

$$\mathbf{Z} = \begin{vmatrix} 0,09362 & 0,2649 & 0,3953 & 0,4753 & 0,5098 \\ 0,7325 & 1,766 & 2,054 & 1,896 & 1,601 \\ 266,8 & 29,64 & 10,672 & 5,445 & 3,294 \\ 2,012 & 4,037 & 3,773 & 2,934 & 2,200 \end{vmatrix},$$

wobei ρ_i die Dimension g cm^{-3} besitzt. Die Aufstellung und Lösung des Normalgleichungssystems (354.2) ergibt mit (354.1) die Schätzwerte $\hat{\boldsymbol{\beta}}$ und $\hat{\boldsymbol{\gamma}}$ sowie $\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\gamma}}$ als Rechenprobe wegen (352.9)

$$\hat{\boldsymbol{\beta}} = \begin{vmatrix} 2,400E-01 \\ 5,716E-04 \\ 2,291E-03 \end{vmatrix}, \quad \hat{\boldsymbol{\gamma}} = \begin{vmatrix} -1,597E-05 \\ -1,612E-06 \\ -4,736E-07 \\ -1,901E-07 \\ -9,065E-08 \end{vmatrix}, \quad \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\gamma}} = \begin{vmatrix} 0,2400 \\ 0,2500 \\ 0,2500 \\ 0,2600 \end{vmatrix}.$$

Die Schätzung $\hat{\sigma}^2$ der Varianz der Gewichtseinheit folgt aus (354.5) mit

$$\hat{\sigma}^2 = 2,56E - 10$$

und mit Hilfe der inversen Normalgleichungsmatrix (354.3) die mit $\hat{\sigma}^2$ berechnete Kovarianzmatrix $\hat{D}(\hat{\boldsymbol{\beta}}) = -\hat{\sigma}^2 \boldsymbol{\Sigma}_{\beta\beta}$ aus (354.4)

$$\hat{D}(\hat{\boldsymbol{\beta}}) = 2,56E - 10 \begin{vmatrix} 0,701 & 0,141 & 0,467 \\ & 0,029 & 0,094 \\ & & 0,436 \end{vmatrix}$$

und entsprechend $\hat{D}(\hat{\beta})$

$$\hat{D}(\hat{\beta}) = 2,56E-10 \begin{vmatrix} 9,89E-01 & 9,99E-02 & 2,93E-02 & 1,18E-02 & 5,62E-03 \\ & 1,01E-02 & 2,96E-03 & 1,19E-03 & 5,67E-04 \\ & & 8,70E-04 & 3,49E-04 & 1,66E-04 \\ & & & 1,40E-04 & 6,69E-05 \\ & & & & 3,19E-05 \end{vmatrix}.$$

355 Modell der Ausgleichung nach bedingten Beobachtungen

Setzt man im gemischten Modell (352.1) $\mathbf{X} = \mathbf{0}$, ergibt sich das Modell

$$\mathbf{Z}\boldsymbol{\gamma} = \mathbf{y} \quad \text{mit} \quad E(\boldsymbol{\gamma}) = \mathbf{0} \quad \text{und} \quad D(\boldsymbol{\gamma}) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}. \quad (355.1)$$

Denkt man sich wieder \mathbf{y} durch die bei (352.1) erwähnte Transformation $\mathbf{y} = \mathbf{Z}\bar{\mathbf{y}} + \mathbf{z}$ entstanden, so wird die Parameterschätzung im Modell (355.1) in der Ausgleichungsrechnung als *Ausgleichung nach bedingten Beobachtungen* bezeichnet. Mit (352.2), (352.3) und (353.1) ergibt sich die beste lineare erwartungstreue Schätzung $\hat{\boldsymbol{\gamma}}$ von $\boldsymbol{\gamma}$ zu

$$\hat{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \mathbf{Z}' (\mathbf{Z} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \mathbf{Z}')^{-1} \mathbf{y} \quad (355.2)$$

mit der Kovarianzmatrix $D(\hat{\boldsymbol{\gamma}})$, die singularär ist, aus (353.5) und $D(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$, die wegen (353.1) minimale Spur besitzt, aus (353.6).

Die Bedeutung des Modells (355.1) liegt darin, daß mit der Bedingung

$$\mathbf{Z}\mathbf{X} = \mathbf{0} \quad (355.3)$$

die Parameterschätzung im Gauß-Markoff-Modell $\mathbf{X}\boldsymbol{\beta} = \mathbf{y}^* + \mathbf{e}$ in die Parameterschätzung im Modell der bedingten Beobachtungen überführt werden kann, wobei \mathbf{y}^* wegen (321.8) dargestellt wird durch $\mathbf{y}^* = \bar{\mathbf{y}} - \mathbf{y}_0$ mit \mathbf{y}_0 einem Vektor von Konstanten. Aus $\mathbf{Z}\mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\bar{\mathbf{y}} - \mathbf{Z}\mathbf{y}_0 + \mathbf{Z}\mathbf{e} = \mathbf{0}$ folgt nämlich mit $-\mathbf{Z}\mathbf{y}_0 = \mathbf{z}$ und $-\mathbf{e} = \boldsymbol{\gamma}$ das Modell $\mathbf{Z}\boldsymbol{\gamma} = \mathbf{Z}\bar{\mathbf{y}} + \mathbf{z}$ und hieraus mit der oben angegebenen Transformation das Modell (355.1). Wegen (321.1) gilt $D(\mathbf{e}) = D(\bar{\mathbf{y}})$ und somit $D(\boldsymbol{\gamma}) = D(\bar{\mathbf{y}})$. Der Transformation des Gauß-Markoff-Modells in das Modell (355.1) mittels (355.3) entspricht die Elimination der u unbekannt Parameter aus den n Beobachtungsgleichungen, so daß $n-u$ Bedingungsgleichungen verbleiben. Für $n-u < u$ ist die Anzahl der im Modell der bedingten Beobachtungen zu schätzenden Parameter kleiner als die im Gauß-Markoff-Modell, so daß in der Zeit vor den elektronischen Datenverarbeitungsanlagen zum Beispiel geodätische Netze häufig nach bedingten Beobachtungen ausgeglichen wurden (WOLF 1968, S.345).

Beispiel: In einem ebenen Dreieck wurden die drei Winkel α, β und γ gemessen, wobei die Beobachtungen voneinander unabhängig sind und gleiche Varianzen besitzen. Ihre negativen Fehler e_α, e_β und e_γ seien die unbekanntes Zufallsparameter.

Mit $\bar{\mathbf{y}} = |\alpha, \beta, \gamma|'$, $\boldsymbol{\gamma} = -|e_\alpha, e_\beta, e_\gamma|'$ folgt $D(\bar{\mathbf{y}}) = \sigma^2 \mathbf{I}$ und, wie oben angegeben, $D(\boldsymbol{\gamma}) = D(\bar{\mathbf{y}})$. Weiter gilt in einem ebenen Dreieck die Bedingungsgleichung

$$\alpha + e_\alpha + \beta + e_\beta + \gamma + e_\gamma = 180^\circ$$

oder entsprechend (355.1)

$$\mathbf{Z}\boldsymbol{\gamma} = \mathbf{y}$$

mit

$$\mathbf{Z} = |1, 1, 1| \quad \text{und} \quad \mathbf{y} = |\alpha + \beta + \gamma - 180^\circ|.$$

Die Schätzwerte für $\boldsymbol{\gamma}$ erhält man dann aus (355.2) zu

$$\hat{\boldsymbol{\gamma}} = \frac{1}{3} \begin{vmatrix} 1 \\ 1 \\ 1 \end{vmatrix} \mathbf{y}.$$

356 Prädiktion und Filterung

Mit (351.11) wurden bereits Schätzungen für die Vorhersage von Zufallsparametern durch Beobachtungen angegeben. Diese Vorhersage bezeichnet man auch als *Prädiktion*. Werden, wie im folgenden erläutert wird, die Beobachtungen selbst prädiziert, spricht man von *Filterung*. Für die Schätzung nach (351.11) benötigt man die Erwartungswerte der Beobachtungen, die im allgemeinen unbekannt sind, so daß sie durch die Einführung sogenannter *systematischer Anteile* oder *Trendanteile* zu schätzen sind. Dies geschieht durch eine Spezialisierung des gemischten Modells (352.1).

Setzt man anstelle von (352.1)

$$\mathbf{X}\boldsymbol{\beta} + \bar{\mathbf{Z}}\bar{\boldsymbol{\gamma}} = \mathbf{y} \quad \text{mit} \quad E(\bar{\boldsymbol{\gamma}}) = \mathbf{0} \quad (356.1)$$

und

$$\bar{\mathbf{Z}} = |\mathbf{Z}, -\mathbf{I}_n|, \quad \bar{\boldsymbol{\gamma}} = |\boldsymbol{\gamma}', \mathbf{e}'|', \quad D(\bar{\boldsymbol{\gamma}}) = \sigma^2 \begin{vmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}\boldsymbol{\gamma}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{e}\mathbf{e}} \end{vmatrix},$$

worin \mathbf{Z} eine $n \times (r - n)$ Matrix bekannter Koeffizienten, $\boldsymbol{\gamma}$ ein $(r - n) \times 1$ Vektor und \mathbf{e} ein $n \times 1$ Vektor bedeuten, ergibt sich das Modell

$$\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} = \mathbf{y} + \mathbf{e} \quad \text{mit} \quad E(\boldsymbol{\gamma}) = \mathbf{0} \quad \text{und} \quad E(\mathbf{e}) = \mathbf{0}. \quad (356.2)$$

In diesem Modell wird also die Summe des Beobachtungsvektors \mathbf{y} und des Vektors \mathbf{e} dargestellt durch einen systematischen Anteil $\mathbf{X}\boldsymbol{\beta}$ und einen zufälligen Anteil $\mathbf{Z}\boldsymbol{\gamma}$, der auch als *Signal* bezeichnet wird. Den Vektor \mathbf{e} interpretiert man als Fehler des Vektors \mathbf{y} , wobei jedoch im Unterschied zum Gauß-Markoff-Modell (321.2) $\mathbf{e} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} - \mathbf{y}$ gilt. Die Schätzwerte für die Summe $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma}$ von Trend und Signal ergeben die korrigierten oder gefilterten Beobachtungen. In Abbildung 356-1 ist mit $\mathbf{X}\boldsymbol{\beta} = ((\mathbf{X}\boldsymbol{\beta})_i)$, $\mathbf{Z}\boldsymbol{\gamma} = ((\mathbf{Z}\boldsymbol{\gamma})_i)$, $\mathbf{y} = (y_i)$ und $\mathbf{e} = (e_i)$ die Zusammensetzung der Beobachtungen y_i , die in der

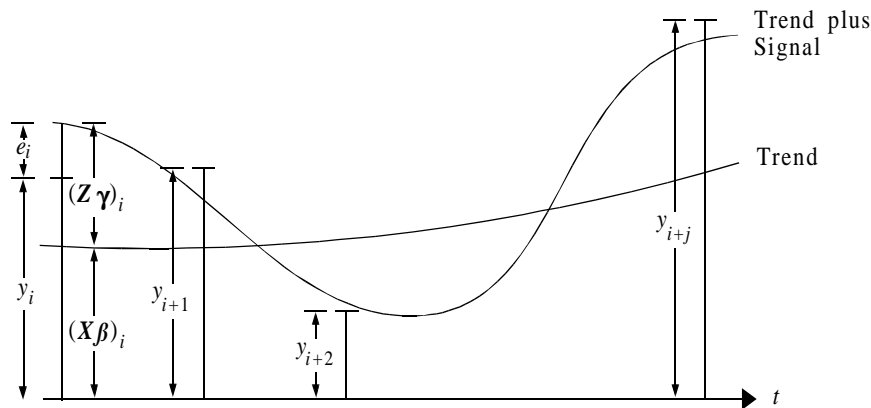


Abb. 356-1: Trend und Signal

Abbildung Funktionen der Zeit t sind, aus Trend $(\mathbf{X}\boldsymbol{\beta})_i$, Signal $(\mathbf{Z}\boldsymbol{\gamma})_i$ und Fehler e_i dargestellt.

Beispiel: In der physikalischen Geodäsie wird gezeigt, daß der Absolutbetrag der Schwerkraft an der Erdoberfläche und seine Richtung Funktionen des Störpotentials der Erdschwere und des Referenzpotentials sind, auf das das Störpotential sich bezieht. Falls das Störpotential in Kugelfunktionen entwickelt wird, lassen sich ihre unbekannt harmonischen Koeffizienten mit den Zufallsparametern $\boldsymbol{\gamma}$ identifizieren. Die unbekannt Parameter des Referenzpotentials, zum Beispiel ebenfalls harmonische Koeffizienten, kann man den Parametern $\boldsymbol{\beta}$ zuordnen. Dann enthält, wenn erforderlich nach einer Linearisierung, $\mathbf{X}\boldsymbol{\beta}$ den Trend und $\mathbf{Z}\boldsymbol{\gamma}$ das Signal für die an der Erdoberfläche gemessene Schwerkraft und ihre Richtung, die in den Beobachtungsvektor \mathbf{y}

eingehen.

Mit (356.1) ergibt sich anstelle von (352.2)

$$D(\mathbf{y}) = \sigma^2 \bar{\mathbf{Z}} \begin{vmatrix} \Sigma_{\gamma\gamma} & \mathbf{0} \\ \mathbf{0} & \Sigma_{ee} \end{vmatrix} \bar{\mathbf{Z}}' = \sigma^2 (\mathbf{Z} \Sigma_{\gamma\gamma} \mathbf{Z}' + \Sigma_{ee}) = \sigma^2 \Sigma_{yy}. \quad (356.3)$$

Die Matrix Σ_{yy} ist nach (143.7) positiv definit, da die Matrix $\bar{\mathbf{Z}} = |\mathbf{Z}, -\mathbf{I}_n|$, wie aus (132.8) ersichtlich, immer vollen Zeilenrang besitzt und $D(\bar{\gamma})$ nach Voraussetzung positiv definit ist. Anstelle von (352.3) erhält man

$$C(\bar{\gamma}, \mathbf{y}) = \sigma^2 \begin{vmatrix} \Sigma_{\gamma\gamma} & \mathbf{0} \\ \mathbf{0} & \Sigma_{ee} \end{vmatrix} \bar{\mathbf{Z}}' = \sigma^2 \begin{vmatrix} \Sigma_{\gamma\gamma} \mathbf{Z}' \\ -\Sigma_{ee} \end{vmatrix} = \sigma^2 \begin{vmatrix} \Sigma_{\gamma y} \\ -\Sigma_{ee} \end{vmatrix}. \quad (356.4)$$

Aus (352.5) und (352.8) folgen dann die Schätzwerte, die lediglich in der unterschiedlich definierten Kovarianzmatrix Σ_{yy} voneinander abweichen,

$$\hat{\gamma} = \Sigma_{\gamma y} \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}) \quad \text{und} \quad \hat{\beta} = (\mathbf{X}' \Sigma_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma_{yy}^{-1} \mathbf{y}, \quad (356.5)$$

so daß bei entsprechender Definition von Σ_{yy} auch die Kovarianzmatrizen (353.4) bis (353.8) gelten. Die Schätzwerte $\hat{\gamma}$ lassen sich wieder ohne Kenntnis der Matrix \mathbf{Z} berechnen, sofern die Kovarianzmatrix $\Sigma_{\gamma y}$ gegeben ist.

Weiter erhält man aus (352.8) die Schätzung $\hat{e} = -\Sigma_{ee} \Sigma_{yy}^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta})$, so daß mit (356.3), (356.4) und $\hat{\mathbf{y}} = \mathbf{y} + \hat{e}$ folgt

$$\mathbf{X} \hat{\beta} + \mathbf{Z} \hat{\gamma} = \hat{\mathbf{y}}. \quad (356.6)$$

Die Schätzwerte $\hat{\mathbf{y}}$ ergeben die gefilterten Beobachtungen. Sie sind als Linearkombinationen von $\hat{\beta}$ und $\hat{\gamma}$ wegen (353.1) beste lineare erwartungstreue Schätzungen.

Sollen die Beobachtungen nicht nur gefiltert, sondern auch vorhergesagt, also prädiiziert werden, gilt (356.6) entsprechend

$$\mathbf{X}^* \hat{\beta} + \mathbf{Z}^* \hat{\gamma} = \hat{\mathbf{y}}^*, \quad (356.7)$$

worin \mathbf{X}^* und \mathbf{Z}^* die Berechnung der Trend- und Signalanteile für die zu prädiizierenden Beobachtungen bewirken und $\hat{\mathbf{y}}^*$ die prädiizierten Beobachtungen bezeichnet. Sind die Beobachtungen beispielsweise Funktionen der Zeit oder eines Ortsvektors und läßt sich der systematische Anteil $\mathbf{X} \beta$ durch eine Polynomentwicklung darstellen, wie im Beispiel des Kapitels 354 gezeigt wurde, erhält man \mathbf{X}^* mit den Zeiten oder den Ortsvektoren der zu prädiizierenden Beobachtungen. Die Kovarianzmatrix $D(\mathbf{X}^* \hat{\beta} + \mathbf{Z}^* \hat{\gamma})$ der prädiizierten Beobachtungen $\hat{\mathbf{y}}^*$ und die Kovarianzmatrix $D(\mathbf{X}^* \hat{\beta} + \mathbf{Z}^* (\hat{\gamma} - \gamma))$,

deren Varianzen wegen (353.1) minimal sind und nach (312.1) ein Maß für die Prädiktionsfehler angeben, erhält man mit (233.13), (233.15), (353.4) bis (353.8) und mit (356.3) zu

$$D(\mathbf{X}^* \hat{\boldsymbol{\beta}} + \mathbf{Z}^* \hat{\boldsymbol{\gamma}}) = \sigma^2 (\mathbf{X}^* (\mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}^{*'} + \mathbf{Z}^* \boldsymbol{\Sigma}_{\gamma y} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{y\gamma} \mathbf{Z}^{*'} - \mathbf{Z}^* \boldsymbol{\Sigma}_{\gamma y} \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{X} (\mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{y\gamma} \mathbf{Z}^{*'}) \quad (356.8)$$

sowie

$$D(\mathbf{X}^* \hat{\boldsymbol{\beta}} + \mathbf{Z}^* (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})) = \sigma^2 (\mathbf{Z}^* \boldsymbol{\Sigma}_{\gamma\gamma} \mathbf{Z}^{*'} - \mathbf{Z}^* \boldsymbol{\Sigma}_{\gamma y} \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{y\gamma} \mathbf{Z}^{*'} + (\mathbf{Z}^* \boldsymbol{\Sigma}_{\gamma y} \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{X} - \mathbf{X}^*) (\mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{X})^{-1} (\mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{y\gamma} \mathbf{Z}^{*'} - \mathbf{X}^{*'})) . \quad (356.9)$$

Soll lediglich das Signal $\mathbf{Z}\boldsymbol{\gamma} = \mathbf{s}$ geschätzt werden, erhält man mit

$$\mathbf{Z}\boldsymbol{\gamma} = \mathbf{s} , \quad \mathbf{Z}\hat{\boldsymbol{\gamma}} = \hat{\mathbf{s}} , \quad \mathbf{Z}\boldsymbol{\Sigma}_{\gamma\gamma} \mathbf{Z}' = \boldsymbol{\Sigma}_{ss} \quad \text{und} \quad \boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{ss} + \boldsymbol{\Sigma}_{ee} \quad (356.10)$$

und mit (356.4) aus (356.5) die Schätzung $\hat{\mathbf{s}}$ des Signals \mathbf{s}

$$\hat{\mathbf{s}} = \boldsymbol{\Sigma}_{ss} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad \text{und} \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{y} . \quad (356.11)$$

Aus (356.6) folgen dann die gefilterten Beobachtungen $\hat{\mathbf{y}}$

$$\mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{s}} = \hat{\mathbf{y}} \quad (356.12)$$

und mit

$$\mathbf{Z}^* \boldsymbol{\gamma} = \mathbf{s}^* , \quad \mathbf{Z}^* \hat{\boldsymbol{\gamma}} = \hat{\mathbf{s}}^* \quad \text{und} \quad \mathbf{Z}^* \boldsymbol{\Sigma}_{\gamma\gamma} \mathbf{Z}' = C(\mathbf{Z}^* \boldsymbol{\gamma}, \mathbf{Z}\boldsymbol{\gamma}) / \sigma^2 = \boldsymbol{\Sigma}_{s^*s} \quad (356.13)$$

anstelle von (356.11) das prädizierte Signal $\hat{\mathbf{s}}^*$

$$\hat{\mathbf{s}}^* = \boldsymbol{\Sigma}_{s^*s} \boldsymbol{\Sigma}_{yy}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad \text{und} \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}_{yy}^{-1} \mathbf{y} \quad (356.14)$$

sowie anstelle von (356.7) die prädizierten Beobachtungen $\hat{\mathbf{y}}^*$

$$\mathbf{X}^* \hat{\boldsymbol{\beta}} + \hat{\mathbf{s}}^* = \hat{\mathbf{y}}^* \quad (356.15)$$

mit den entsprechenden Kovarianzmatrizen aus (356.8) und (356.9).

Die Prädiktion nach (356.15) eignet sich vorzüglich zur Interpolation von Meßdaten (KOCH 1973; KRAUS 1971; PELZER 1978), aber auch die in den Beobachtungen enthaltenen zufälligen Anteile $\boldsymbol{\gamma}$ lassen sich nach (356.5) bequem schätzen (HEITZ 1968; MORITZ 1973, 1980), wobei aber zu beachten ist, daß $E(\mathbf{s}) = \mathbf{0}$ oder $E(\boldsymbol{\gamma}) = \mathbf{0}$ im Modell vorausgesetzt wird (KOCH 1977). Allerdings ist das gleichzeitige Auftreten unbekannter fester Parameter und

Zufallsparameter in einem Modell nicht unbedingt einleuchtend. Eine plausible Interpretation des Modells der Prädiktion und Filterung liefert jedoch die Bayes-Statistik, mit der auch Hypothesen für die Parameter getestet und Bereichsschätzungen vorgenommen werden können (KOCH 1990, S.109; KOCH 1994; KOCH 2000, S.129).

Die Schätzung der für die Prädiktion benötigten Kovarianzmatrizen und ihre Annäherung durch Kovarianzfunktionen werden im Kapitel 375 behandelt. Sind die in (356.5) und (356.14) auftretenden Vektoren und Matrizen durch den Abbruch einer für das Prädiktionsproblem eigentlich notwendigen unendlich-dimensionalen Basisentwicklung entsprechender Größe entstanden, ergibt sich das als *Kollokation* bekannte Prädiktionsverfahren (KRARUP 1969; MEISSL 1976; MORITZ 1973, 1980; RUMMEL 1976; TSCHERNING 1978).

Der Prädiktion nach (356.15) mit $\Sigma_{ee} = \mathbf{0}$ entspricht die optimale Vorhersage (TOUTENBURG 1975, S.70), die aus einer linearen Vorhersage der Beobachtungen im Gauß-Markoff-Modell folgt. Sind die Beobachtungen trendfrei, erhält man anstelle von (356.14) $\hat{\mathbf{s}}^* = \mathbf{H}\mathbf{y}$, wobei für \mathbf{H} die Gleichung $\Sigma_{s^*s} = \mathbf{H}\Sigma_{yy}$ gilt, die der *Wiener-Hopf-Integralgleichung* in der Theorie der stochastischen Prozesse entspricht (KOCH und SCHMIDT 1994, S.233; NEUBURGER 1972, S.59; PAPOULIS 1965, S.404).

Abschließend soll noch gezeigt werden, daß mit (356.5) übereinstimmende Schätzwerte auch in dem Gauß-Markoff-Modell

$$\bar{\mathbf{X}}\bar{\boldsymbol{\beta}} = E(\bar{\mathbf{y}}) \quad (356.16)$$

erhalten werden mit

$$\bar{\mathbf{X}} = \begin{vmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{0} & \mathbf{I} \end{vmatrix}, \quad \bar{\boldsymbol{\beta}} = \begin{vmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{vmatrix}, \quad \bar{\mathbf{y}} = \begin{vmatrix} \mathbf{y} \\ \mathbf{0} \end{vmatrix}$$

und $D(\bar{\mathbf{y}}) = \sigma^2 \begin{vmatrix} \Sigma_{ee} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\gamma\gamma} \end{vmatrix},$

worin $\boldsymbol{\gamma}$ jetzt aber feste Parameter bedeuten. Aus (322.9) folgt

$$\begin{vmatrix} \mathbf{X}'\Sigma_{ee}^{-1}\mathbf{X} & \mathbf{X}'\Sigma_{ee}^{-1}\mathbf{Z} \\ \mathbf{Z}'\Sigma_{ee}^{-1}\mathbf{X} & \mathbf{Z}'\Sigma_{ee}^{-1}\mathbf{Z} + \Sigma_{\gamma\gamma}^{-1} \end{vmatrix} \begin{vmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{vmatrix} = \begin{vmatrix} \mathbf{X}'\Sigma_{ee}^{-1}\mathbf{y} \\ \mathbf{Z}'\Sigma_{ee}^{-1}\mathbf{y} \end{vmatrix}. \quad (356.17)$$

Hieraus ergibt sich

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Z}'\Sigma_{ee}^{-1}\mathbf{Z} + \Sigma_{\gamma\gamma}^{-1})^{-1}\mathbf{Z}'\Sigma_{ee}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (356.18)$$

und mit der Identität (134.7) die Schätzung

$$\hat{\boldsymbol{\gamma}} = \Sigma_{\gamma\gamma}\mathbf{Z}'(\mathbf{Z}\Sigma_{\gamma\gamma}\mathbf{Z}' + \Sigma_{ee})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad (356.19)$$

die nach Substitution von (356.3) und (356.4) identisch ist mit der ersten Beziehung in (356.5). Setzt man (356.18) in (356.17) ein, ergibt sich

$$\begin{aligned} \mathbf{X}'(\boldsymbol{\Sigma}_{ee}^{-1} - \boldsymbol{\Sigma}_{ee}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{Z} + \boldsymbol{\Sigma}_{\gamma\gamma}^{-1})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}_{ee}^{-1})\mathbf{X}\hat{\boldsymbol{\beta}} = \\ \mathbf{X}'(\boldsymbol{\Sigma}_{ee}^{-1} - \boldsymbol{\Sigma}_{ee}^{-1}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}_{ee}^{-1}\mathbf{Z} + \boldsymbol{\Sigma}_{\gamma\gamma}^{-1})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}_{ee}^{-1})\mathbf{y} \end{aligned} \quad (356.20)$$

und mit der Identität (134.6) die Schätzung

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'(\mathbf{Z}\boldsymbol{\Sigma}_{\gamma\gamma}\mathbf{Z}' + \boldsymbol{\Sigma}_{ee})^{-1}\mathbf{X})^{-1}\mathbf{X}'(\mathbf{Z}\boldsymbol{\Sigma}_{\gamma\gamma}\mathbf{Z}' + \boldsymbol{\Sigma}_{ee})^{-1}\mathbf{y}, \quad (356.21)$$

die wegen (356.3) mit der zweiten Beziehung in (356.5) identisch ist.

36 Schätzung von Varianz- und Kovarianzkomponenten

361 Beste invariante quadratische erwartungstreue Schätzung

Der Zufallsvektor \mathbf{e} der Fehler im Gauß–Markoff–Modell (321.2) lasse sich wie im gemischten Modell (352.1) darstellen durch

$$\mathbf{e} = |\mathbf{U}_1, \dots, \mathbf{U}_l| \begin{vmatrix} \gamma_1 \\ \dots \\ \gamma_l \end{vmatrix}, \quad (361.1)$$

worin die Matrizen \mathbf{U}_i die Dimensionen $n \times r_i$ besitzen und γ_i unbekannte und Messungen nicht zugängliche Zufallsvektoren der Dimensionen $r_i \times 1$ bedeuten, für die $E(\gamma_i) = \mathbf{0}$ sowie $C(\gamma_i, \gamma_j) = \sigma_{ij}\mathbf{R}_{ij}$ mit $\sigma_{ij} = \sigma_{ji}$ und $\mathbf{R}'_{ij} = \mathbf{R}_{ji}$ für $i, j \in \{1, \dots, l\}$ gelte. Die Kovarianzmatrix $D(\mathbf{y}) = \boldsymbol{\Sigma}$ der Beobachtungen \mathbf{y} ergibt sich dann mit (233.2) zu $\boldsymbol{\Sigma} = \sigma_1^2\mathbf{U}_1\mathbf{R}_{11}\mathbf{U}'_1 + \sigma_{12}(\mathbf{U}_1\mathbf{R}_{12}\mathbf{U}'_2 + \mathbf{U}_2\mathbf{R}_{21}\mathbf{U}'_1) + \sigma_{13}(\mathbf{U}_1\mathbf{R}_{13}\mathbf{U}'_3 + \mathbf{U}_3\mathbf{R}_{31}\mathbf{U}'_1) + \dots + \sigma_l^2\mathbf{U}_l\mathbf{R}_{ll}\mathbf{U}'_l$. Mit $\mathbf{U}_1\mathbf{R}_{11}\mathbf{U}'_1 = \mathbf{V}_1 = \alpha_1^2\mathbf{T}_1$, $\mathbf{U}_1\mathbf{R}_{12}\mathbf{U}'_2 + \mathbf{U}_2\mathbf{R}_{21}\mathbf{U}'_1 = \mathbf{V}_2 = \alpha_{12}\mathbf{T}_2, \dots, \mathbf{U}_l\mathbf{R}_{ll}\mathbf{U}'_l = \mathbf{V}_k = \alpha_l^2\mathbf{T}_k$, wobei $k = l(l+1)/2$ gilt und die Matrizen \mathbf{V}_m und \mathbf{T}_m mit $m \in \{1, \dots, k\}$ symmetrisch sind, erhält man

$$\begin{aligned} \boldsymbol{\Sigma} &= \sigma_1^2\alpha_1^2\mathbf{T}_1 + \sigma_{12}\alpha_{12}\mathbf{T}_2 + \dots + \sigma_l^2\alpha_l^2\mathbf{T}_k \\ &= \sigma_1^2\mathbf{V}_1 + \sigma_{12}\mathbf{V}_2 + \dots + \sigma_l^2\mathbf{V}_k. \end{aligned} \quad (361.2)$$

In (361.2) seien die Matrizen \mathbf{T}_m und die Faktoren α_i^2 sowie α_{ij} und damit auch die Matrizen \mathbf{V}_m bekannt, während die Varianzen σ_i^2 und die Kovarianzen σ_{ij} unbekannt sind. Damit führt (361.2) auf ein Gauß–Markoff–Modell, in dem neben den unbekanntem Parametern $\boldsymbol{\beta}$ anstelle der Varianz σ^2 der

Gewichtseinheit die sogenannten Varianzkomponenten σ_i^2 und die Kovarianzkomponenten σ_{ij} für die zugehörigen Matrizen $\mathbf{V}_m \neq \mathbf{0}$ zu schätzen sind.

Definition: Es sei \mathbf{X} eine $n \times u$ Matrix gegebener Koeffizienten, $\boldsymbol{\beta}$ ein $u \times 1$ Vektor unbekannter Parameter, \mathbf{y} ein $n \times 1$ Zufallsvektor von Beobachtungen, dessen Kovarianzmatrix $D(\mathbf{y}) = \boldsymbol{\Sigma}$ positiv definit sei, dann bezeichnet man

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit}$$

$$\begin{aligned} D(\mathbf{y}) = \boldsymbol{\Sigma} &= \sigma_1^2 \alpha_1^2 \mathbf{T}_1 + \sigma_{12} \alpha_{12} \mathbf{T}_2 + \dots + \sigma_l^2 \alpha_l^2 \mathbf{T}_k \\ &= \sigma_1^2 \mathbf{V}_1 + \sigma_{12} \mathbf{V}_2 + \dots + \sigma_l^2 \mathbf{V}_k \end{aligned}$$

als Gauß–Markoff–Modell mit k unbekanntem *Varianz-* und *Kovarianzkomponenten* σ_i^2 und σ_{ij} mit $i \in \{1, \dots, l\}$, $i < j \leq l$ und $l \leq k \leq l(l+1)/2$. Die Faktoren α_i^2 und α_{ij} seien bekannt und Näherungswerte für die Produkte $\sigma_i^2 \alpha_i^2$ und $\sigma_{ij} \alpha_{ij}$. Weiter seien die $n \times n$ Matrizen \mathbf{T}_m und damit \mathbf{V}_m für $m \in \{1, \dots, k\}$ bekannt und symmetrisch, und $\sum_{m=1}^k \mathbf{V}_m$ sei positiv definit. (361.3)

Varianzkomponenten sind beispielsweise zu schätzen, wenn Gruppen verschiedenartiger Beobachtungen wie Strecken- und Winkelmessungen vorliegen, die voneinander unabhängig sind und von denen lediglich die Gewichtsmatrizen bekannt sind. Der Beobachtungsvektor \mathbf{y}_1 mit der Gewichtsmatrix \mathbf{P}_1 enthalte die Streckenmessungen und der Vektor \mathbf{y}_2 mit der Gewichtsmatrix \mathbf{P}_2 die Winkelmessungen. Die Kovarianzmatrix von \mathbf{y}_1 und \mathbf{y}_2 mit der unbekanntem Varianzkomponente σ_1^2 für die Strecken und der unbekanntem Varianzkomponente σ_2^2 für die Winkel ergibt sich dann zu

$$D\left(\begin{array}{c} \mathbf{y}_1 \\ \mathbf{y}_2 \end{array}\right) = \sigma_1^2 \begin{vmatrix} \mathbf{P}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} + \sigma_2^2 \begin{vmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2^{-1} \end{vmatrix}.$$

Sind die Beobachtungsgruppen voneinander abhängig und die Kovarianzen bis auf gemeinsame Faktoren bekannt, sind diese Faktoren als Kovarianzkomponenten zu schätzen. Für das Beispiel der Strecken- und Winkelmessungen erhält man mit der Kovarianzmatrix \mathbf{R}_{12} der Strecken und Winkel und der unbekanntem Kovarianzkomponente σ_{12} die Kovarianzmatrix

$$D\left(\begin{array}{c} \mathbf{y}_1 \\ \mathbf{y}_2 \end{array}\right) = \sigma_1^2 \begin{vmatrix} \mathbf{P}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix} + \sigma_{12} \begin{vmatrix} \mathbf{0} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{0} \end{vmatrix} + \sigma_2^2 \begin{vmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2^{-1} \end{vmatrix},$$

in der $\mathbf{R}_{21} = \mathbf{R}'_{12}$ gilt.

Um unterschiedliche Einflüsse auf die Varianzen von Beobachtungen angeben zu können, müssen ebenfalls Varianzkomponenten geschätzt werden, wie

zum Beispiel bei den Streckenmessungen, deren Varianzen man sich häufig aus einem von der Entfernung abhängigen Anteil und einem restlichen Anteil zusammengesetzt vorstellt (FRÖHLICH und DUDDEK 1983; KOCH 1981). Ein weiteres Beispiel ist die Schätzung von Regularisierungsparametern aus Varianzkomponenten (KOCH und KUSCHE 2002).

Die Schätzung der Varianz- und Kovarianzkomponenten soll eine Verallgemeinerung der Schätzung der Varianz der Gewichtseinheit darstellen, die, wie aus (331.5), (331.9) und (331.12) ersichtlich, mit Hilfe einer quadratischen Form der Beobachtungen beziehungsweise Residuen gefunden wird, wobei die Erwartungstreue der Schätzung dadurch erzielt wird, daß man in (331.11) den Erwartungswert der quadratischen Form durch den berechneten Wert und die Varianz der Gewichtseinheit durch ihren Schätzwert ersetzt, so daß (331.12) erhalten wird. Entsprechend läßt sich auch bei der Schätzung der Varianzkomponenten vorgehen, was bereits auf HELMERT (1924, S.358) zurückgeht. Sein Verfahren besitzt auch die Eigenschaft einer besten invarianten Schätzung (GRAFAREND et al. 1980; KELM 1978; WELSCH 1978), so daß es mit der im folgenden abgeleiteten Schätzung übereinstimmt. Varianzkomponenten lassen sich auch nach der Maximum-Likelihood-Methode schätzen, wobei im allgemeinen iterativ vorzugehen ist (EBNER 1972; HARVILLE 1977; KOCH 1990, S.124; KUBIK 1970; SEARLE 1971, S.462).

Im folgenden sollen wie für die unbekannt Parameter β im Gauß-Markoff-Modell auch für die unbekannt Varianz- und Kovarianzkomponenten beste erwartungstreue Schätzungen im Sinne von (312.2) abgeleitet werden. Wie bereits erwähnt, soll die Schätzung eine Verallgemeinerung der Schätzung der Varianz der Gewichtseinheit darstellen, so daß die lineare Funktion $\mathbf{p}'\sigma$ der Varianz- und Kovarianzkomponenten, in der \mathbf{p} ein gegebener $k \times 1$ Vektor und $\sigma = |\sigma_1^2, \sigma_{12}, \dots, \sigma_l^2|'$ bedeuten, mit Hilfe der skalaren Größe $\mathbf{y}'\mathbf{D}\mathbf{y}$ geschätzt wird, in der \mathbf{D} eine unbekannte $n \times n$ Matrix bezeichnet. Da $\mathbf{y}'\mathbf{D}\mathbf{y} = \mathbf{y}'\mathbf{D}'\mathbf{y} = \mathbf{y}'(\mathbf{D} + \mathbf{D}')\mathbf{y}/2$ gilt, wird \mathbf{D} im folgenden als symmetrische Matrix gesucht, so daß $\mathbf{y}'\mathbf{D}\mathbf{y}$ eine quadratische Form der Beobachtungen \mathbf{y} darstellt. Somit wird eine *beste quadratische erwartungstreue* Schätzung hergeleitet, für die nach (312.2) zu fordern ist

$$1) \quad E(\mathbf{y}'\mathbf{D}\mathbf{y}) = \mathbf{p}'\sigma$$

oder mit (271.1) und (361.3)

$$\text{sp}(\mathbf{D}\Sigma) + \beta' \mathbf{X}' \mathbf{D} \mathbf{X} \beta = \mathbf{p}'\sigma . \quad (361.4)$$

Ferner muß

$$2) \quad V(\mathbf{y}'\mathbf{D}\mathbf{y}) \quad \text{minimal}$$

werden. Diese Varianz soll mit Hilfe der Erwartungswerte und der Kovarianzmatrix der Beobachtungen angegeben werden, so daß \mathbf{y} als normalverteilt angenommen wird und mit (361.3) aus (251.1) $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$ folgt. Mit (271.2) und (361.3) ergibt sich dann

$$V(\mathbf{y}'\mathbf{D}\mathbf{y}) = 2\text{sp}(\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}\boldsymbol{\Sigma}) + 4\boldsymbol{\beta}'\mathbf{X}'\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}\mathbf{X}\boldsymbol{\beta}. \quad (361.5)$$

Um die Varianz- und Kovarianzkomponenten $\boldsymbol{\sigma}$ unabhängig von den unbekanntem Parametern $\boldsymbol{\beta}$ zu schätzen, müssen (361.4) und (361.5) unabhängig von den Werten für $\boldsymbol{\beta}$ sein, wozu es notwendig und hinreichend ist, daß

$$\mathbf{D}\mathbf{X} = \mathbf{0} \quad (361.6)$$

gilt. Hieraus folgt auch die *Translations-Invarianz* der quadratischen Form, denn es gilt $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*)'\mathbf{D}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^*) = \mathbf{y}'\mathbf{D}\mathbf{y}$, worin $\boldsymbol{\beta}^*$ ein beliebiger $u \times 1$ Vektor bedeutet. Die Matrizen der quadratischen Formen (325.2) oder (331.9), mit denen die Varianz σ^2 der Gewichtseinheit geschätzt wird, erfüllen (361.6).

Fordert man lediglich die Erwartungstreue der quadratischen Form $\mathbf{y}'\mathbf{D}\mathbf{y}$ für beliebige Werte für $\boldsymbol{\beta}$, ist es wegen (361.4) notwendig und hinreichend, daß $\mathbf{X}'\mathbf{D}\mathbf{X} = \mathbf{0}$ gilt. Dies ist im allgemeinen eine schwächere Forderung als (361.6); äquivalent sind beide beispielsweise dann, wenn \mathbf{D} positiv semidefinit ist. Nach (143.10) gilt dann nämlich $\mathbf{D} = \mathbf{H}\mathbf{H}'$, so daß $\mathbf{X}'\mathbf{H}\mathbf{H}'\mathbf{X} = \mathbf{0}$ und mit (131.11) $\mathbf{H}'\mathbf{X} = \mathbf{0}$ und schließlich $\mathbf{D}\mathbf{X} = \mathbf{0}$ folgt. Da aber die Matrix \mathbf{D} lediglich als symmetrisch gesucht wird, soll die Bedingung (361.6) erfüllt sein, also eine *beste invariante quadratische erwartungstreue Schätzung* abgeleitet werden. Schätzungen, für die lediglich $\mathbf{X}'\mathbf{D}\mathbf{X} = \mathbf{0}$ gefordert wird, befinden sich bei (DRYGAS 1977; GRAFAREND und D'HONE 1978; KLEFFE und PINCUS 1974; LAMOTTE 1973B).

Mit (361.6) ergibt sich anstelle von (361.5)

$$V(\mathbf{y}'\mathbf{D}\mathbf{y}) = 2\text{sp}(\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}\boldsymbol{\Sigma}) \quad (361.7)$$

und anstelle von (361.4) mit $\mathbf{p} = (p_i)$ und (361.3) $\sigma_1^2\text{sp}(\mathbf{D}\mathbf{V}_1) + \sigma_{12}\text{sp}(\mathbf{D}\mathbf{V}_2) + \dots + \sigma_k^2\text{sp}(\mathbf{D}\mathbf{V}_k) = p_1\sigma_1^2 + p_2\sigma_{12} + \dots + p_k\sigma_k^2$, so daß folgt

$$\text{sp}(\mathbf{D}\mathbf{V}_i) = p_i \quad \text{für } i \in \{1, \dots, k\}. \quad (361.8)$$

Die quadratische Form $\mathbf{y}'\mathbf{D}\mathbf{y}$ ist also beste invariante quadratische erwartungstreue Schätzung von $\mathbf{p}'\boldsymbol{\sigma}$ genau dann, wenn eine symmetrische Matrix \mathbf{D} existiert, mit der (361.7) minimal wird und die die Bedingungen (361.6) und (361.8) erfüllt.

Zur Minimierung von (361.7) werden die in $\boldsymbol{\Sigma}$ enthaltenen unbekanntem Varianz- und Kovarianzkomponenten σ_i^2 und σ_{ij} gleich Eins gesetzt, da nach

(361.3) mit α_i^2 und α_{ij} bereits Näherungswerte für $\sigma_i^2\alpha_i^2$ und $\sigma_{ij}\alpha_{ij}$ vorliegen, so daß die Werte für σ_i^2 und σ_{ij} nahe bei Eins liegen. Die genäherte Kovarianzmatrix Σ_0 ergibt sich zu

$$\Sigma_0 = \sum_{m=1}^k \mathbf{V}_m . \quad (361.9)$$

Die Schätzwerte der Varianz- und Kovarianzkomponenten hängen dann aber von den Näherungswerten α_i^2 und α_{ij} ab, so daß die Bedingungen in (312.2) nur in Ausnahmefällen für alle Komponenten σ_i^2 und σ_{ij} erfüllt sein werden. Damit wird keine gleichmäßig beste, sondern in Abhängigkeit von den Näherungswerten eine *lokal beste* invariante quadratische erwartungstreue Schätzung der Varianz- und Kovarianzkomponenten hergeleitet.

Beim *MINQUE-Verfahren* zur Schätzung von Varianzkomponenten (RAO 1973, S.302) minimiert man unter den Bedingungen (361.6) und (361.8) ebenfalls die Größe $\text{sp}(\mathbf{D}\Sigma_0\mathbf{D}\Sigma_0)$, die jedoch ohne Annahme der Normalverteilung abgeleitet wird, so daß sie nicht identisch mit der Varianz des Schätzwertes zu sein braucht.

362 Lokal beste Schätzung

Zur Minimierung von (361.7) unter den Bedingungen (361.6) und (361.8) wird nach (171.6) die Lagrangesche Funktion $w(\mathbf{D})$ gebildet

$$w(\mathbf{D}) = 2\text{sp}(\mathbf{D}\Sigma_0\mathbf{D}\Sigma_0) - 4\text{sp}(\mathbf{D}\mathbf{X}\Lambda') - 4 \sum_{i=1}^k \lambda_i (\text{sp}(\mathbf{D}\mathbf{V}_i) - p_i) ,$$

in der -4Λ die $n \times u$ Matrix der Lagrangeschen Multiplikatoren für die Bedingungen $\mathbf{D}\mathbf{X} = \mathbf{0}$ und $-4\lambda_i$ die k Lagrangeschen Multiplikatoren für die Bedingungen (361.8) bedeuten. Um nicht nach den Elementen der symmetrischen Matrix \mathbf{D} ableiten zu müssen, wird die bereits erwähnte Darstellung von \mathbf{D} durch beispielsweise $\mathbf{D} = (\mathbf{F} + \mathbf{F}')/2$ benutzt, so daß die Lagrangesche Funktion erhalten wird

$$\begin{aligned} w((\mathbf{F} + \mathbf{F}')/2) &= \frac{1}{2}\text{sp}[(\mathbf{F} + \mathbf{F}')\Sigma_0(\mathbf{F} + \mathbf{F}')\Sigma_0] - 2\text{sp}[(\mathbf{F} + \mathbf{F}')\mathbf{X}\Lambda'] \\ &\quad - 2 \sum_{i=1}^k \lambda_i [\text{sp}((\mathbf{F} + \mathbf{F}')\mathbf{V}_i) - 2p_i] \\ &= \frac{1}{2}\text{sp}(\mathbf{F}\Sigma_0\mathbf{F}\Sigma_0) + \text{sp}(\mathbf{F}\Sigma_0\mathbf{F}'\Sigma_0) + \frac{1}{2}\text{sp}(\mathbf{F}'\Sigma_0\mathbf{F}'\Sigma_0) \\ &\quad - 2\text{sp}(\mathbf{F}\mathbf{X}\Lambda') - 2\text{sp}(\mathbf{F}'\mathbf{X}\Lambda') - 2 \sum_{i=1}^k \lambda_i [\text{sp}((\mathbf{F} + \mathbf{F}')\mathbf{V}_i) - 2p_i] \end{aligned}$$

$$\begin{aligned}
&= \text{sp}(\mathbf{F}\Sigma_0\mathbf{F}\Sigma_0) + \text{sp}(\mathbf{F}\Sigma_0\mathbf{F}'\Sigma_0) - 2\text{sp}(\mathbf{F}\mathbf{X}\Lambda') - 2\text{sp}(\mathbf{F}'\mathbf{X}\Lambda') \\
&\quad - 2\sum_{i=1}^k \lambda_i [\text{sp}((\mathbf{F} + \mathbf{F}')\mathbf{V}_i) - 2p_i].
\end{aligned}$$

Aus $\partial w((\mathbf{F} + \mathbf{F}')/2)/\partial \mathbf{F} = \mathbf{0}$ ergibt sich dann mit (172.4) bis (172.6)

$$2\Sigma_0\mathbf{F}'\Sigma_0 + 2\Sigma_0\mathbf{F}\Sigma_0 - 2\Lambda\mathbf{X}' - 2\mathbf{X}\Lambda' - 4\sum_{i=1}^k \lambda_i \mathbf{V}_i = \mathbf{0}$$

oder

$$\Sigma_0\mathbf{D}\Sigma_0 - \frac{1}{2}(\Lambda\mathbf{X}' + \mathbf{X}\Lambda') - \sum_{i=1}^k \lambda_i \mathbf{V}_i = \mathbf{0}. \quad (362.1)$$

Die Matrix $\mathbf{I} - \mathbf{R}$ mit $\mathbf{R} = \mathbf{X}(\mathbf{X}'\Sigma_0^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_0^{-1}$, in der \mathbf{R} den orthogonalen Projektionsoperator bezüglich des durch Σ_0^{-1} verallgemeinerten Skalarproduktes für die Projektion auf den Spaltenraum $R(\mathbf{X})$ der Matrix \mathbf{X} mit beliebigem Rang bezeichnet, erfüllt nach (162.5) die Beziehung $(\mathbf{I} - \mathbf{R})\mathbf{X} = \mathbf{0}$. Da \mathbf{D} in (362.1) symmetrisch gesucht wird, führt man die symmetrische Matrix $\mathbf{W} = \Sigma_0^{-1}(\mathbf{I} - \mathbf{R})$ ein, für die ebenfalls $\mathbf{W}\mathbf{X} = \mathbf{0}$ gilt und außerdem mit (361.6)

$$\mathbf{W}\Sigma_0\mathbf{D}\Sigma_0\mathbf{W} = \mathbf{D}, \quad (362.2)$$

so daß durch rechts- und linksseitige Multiplikation von (362.1) mit \mathbf{W} die Lösung $\hat{\mathbf{D}}$ für \mathbf{D} sich ergibt

$$\hat{\mathbf{D}} = \sum_{i=1}^k \lambda_i \mathbf{W}\mathbf{V}_i\mathbf{W} \quad \text{mit} \quad \mathbf{W} = \Sigma_0^{-1} - \Sigma_0^{-1}\mathbf{X}(\mathbf{X}'\Sigma_0^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma_0^{-1}. \quad (362.3)$$

Hiermit folgt aus (361.8)

$$\sum_{i=1}^k \lambda_i \text{sp}(\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{V}_j) = p_j \quad \text{für} \quad j \in \{1, \dots, k\}. \quad (362.4)$$

Bezeichnet man mit $\widehat{\mathbf{p}'\boldsymbol{\sigma}}$ die Schätzung von $\mathbf{p}'\boldsymbol{\sigma}$, so gilt

$$\widehat{\mathbf{p}'\boldsymbol{\sigma}} = \mathbf{y}'\hat{\mathbf{D}}\mathbf{y} = \boldsymbol{\lambda}'\mathbf{q} \quad (362.5)$$

mit

$$\boldsymbol{\lambda} = (\lambda_i) \quad \text{und} \quad \mathbf{q} = (q_i) = (\mathbf{y}'\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{y}) \quad \text{für} \quad i \in \{1, \dots, k\}.$$

Wegen (325.2) und (331.9) ist \mathbf{W} die Matrix der quadratischen Form für Ω in dem Modell mit nicht vollem Rang $\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y})$ und $D(\mathbf{y}) = \sigma^2\boldsymbol{\Sigma}_0$. Für den Residuenvektor $\hat{\boldsymbol{e}}$ in diesem Modell gilt (323.9) entsprechend $\hat{\boldsymbol{e}} = -\boldsymbol{\Sigma}_0\mathbf{W}\mathbf{y}$ und damit

$$\mathbf{q} = (q_i) = (\hat{\boldsymbol{e}}'\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_i\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{e}}) \quad \text{für } i \in \{1, \dots, k\}.$$

Definiert man

$$\mathbf{S} = (\text{sp}(\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{V}_j)) \quad \text{mit } i, j \in \{1, \dots, k\},$$

dann führt (362.4) auf $\mathbf{S}'\boldsymbol{\lambda} = \mathbf{p}$. Ist diese Gleichung für einen bestimmten Vektor \mathbf{p} konsistent, gilt also nach (154.2) $\mathbf{p} \in R(\mathbf{S})$, da $\mathbf{S} = \mathbf{S}'$ ist, so folgt mit (154.5) $\boldsymbol{\lambda} = (\mathbf{S}^-)'\mathbf{p}$, worin $(\mathbf{S}^-)'$ eine generalisierte Inverse von \mathbf{S}' ist. Mit (362.5) ergibt sich dann die lokal beste invariante quadratische erwartungstreue Schätzung $\widehat{\mathbf{p}'\boldsymbol{\sigma}}$ von $\mathbf{p}'\boldsymbol{\sigma}$ zu

$$\widehat{\mathbf{p}'\boldsymbol{\sigma}} = \mathbf{p}'\mathbf{S}^-\mathbf{q}. \quad (362.6)$$

Diese Schätzung ist eindeutig, da sie unabhängig von der Wahl der generalisierten Inversen \mathbf{S}^- ist. Um das zu zeigen, wird zunächst $\mathbf{q} \in R(\mathbf{S})$ bewiesen. Hierfür zerlegt man $\boldsymbol{\Sigma}_0$ nach (143.5) in $\boldsymbol{\Sigma}_0 = \mathbf{G}\mathbf{G}'$ und definiert die $n^2 \times k$ Matrix \mathbf{F} nach (137.5) durch die Spalten

$$\mathbf{f}_i = \text{vec}(\mathbf{G}'\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{G}) \quad \text{mit } i \in \{1, \dots, k\}.$$

Wegen (153.6) gilt

$$\mathbf{W}\boldsymbol{\Sigma}_0\mathbf{W} = \mathbf{W} \quad (362.7)$$

und mit (137.3) und (137.7) $\mathbf{f}'_i\mathbf{f}_j = \text{sp}(\mathbf{G}'\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{G}\mathbf{G}'\mathbf{W}\mathbf{V}_j\mathbf{W}\mathbf{G}) = \text{sp}(\mathbf{G}'\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{V}_j\mathbf{W}\mathbf{G}) = \text{sp}(\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{V}_j)$, so daß $\mathbf{F}'\mathbf{F} = \mathbf{S}$ folgt. Weiter sei $\mathbf{h} = \text{vec}(\mathbf{G}^{-1}\mathbf{y}\mathbf{y}'(\mathbf{G}')^{-1})$, so daß $\mathbf{f}'_i\mathbf{h} = \text{sp}(\mathbf{G}'\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{G}\mathbf{G}^{-1}\mathbf{y}\mathbf{y}'(\mathbf{G}')^{-1}) = \text{sp}(\mathbf{y}'\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{y}) = q_i$ sich ergibt. Daher gilt $\mathbf{F}'\mathbf{h} = \mathbf{q}$ und folglich $\mathbf{q} \in R(\mathbf{F}') = R(\mathbf{F}'\mathbf{F}) = R(\mathbf{S})$ nach (135.6). Da $\mathbf{p} \in R(\mathbf{S})$ vorausgesetzt wurde, existieren zwei Vektoren \mathbf{v} und \mathbf{w} derart, daß $\mathbf{p} = \mathbf{S}\mathbf{v}$ und $\mathbf{q} = \mathbf{S}\mathbf{w}$ gilt. Dann ist in (362.6) $\mathbf{p}'\mathbf{S}^-\mathbf{q} = \mathbf{v}'\mathbf{S}\mathbf{S}^-\mathbf{S}\mathbf{w} = \mathbf{v}'\mathbf{S}\mathbf{w}$ wegen (153.1) unabhängig von der Wahl der generalisierten Inversen \mathbf{S}^- , so daß die Schätzung $\widehat{\mathbf{p}'\boldsymbol{\sigma}}$ eindeutig ist.

Setzt man in (362.6) $\mathbf{p}' = |1, 0, 0, \dots|$, $\mathbf{p}' = |0, 1, 0, \dots|$ und so weiter, ergibt sich

$$\hat{\boldsymbol{\sigma}} = \mathbf{S}^-\mathbf{q}. \quad (362.8)$$

Hieraus folgt mit $\hat{\sigma} = |\hat{\sigma}_1^2, \hat{\sigma}_{12}, \dots, \hat{\sigma}_k^2|'$ die lokal beste invariante quadratische erwartungstreue Schätzung $\hat{\sigma}_i^2$ und $\hat{\sigma}_{ij}$ der Varianz- und Kovarianzkomponenten σ_i^2 und σ_{ij} genau dann, wenn die Gleichung $\mathbf{S}'\boldsymbol{\lambda} = \mathbf{p}$ mit $\mathbf{S}' = \mathbf{S}$ für alle k Vektoren \mathbf{p} , die die Spalten einer Einheitsmatrix bilden, konsistent ist, wenn also $\dim R(\mathbf{S}) = k$ gilt, was nach (135.5) bedeutet, daß \mathbf{S} regulär ist. Dann ist $\hat{\sigma}$ wegen (153.22) eindeutig bestimmt durch

$$\hat{\sigma} = \mathbf{S}^{-1}\mathbf{q}. \quad (362.9)$$

Ist \mathbf{S} singulär, existiert die i -te Komponente von $\hat{\sigma}$ für ein $i \in \{1, \dots, k\}$ genau dann, wenn für $\mathbf{p}' = |0, \dots, 0, 1, 0, \dots, 0|$ mit der Eins als i -te Komponente $\mathbf{p} \in R(\mathbf{S})$ gilt.

Mit der Schätzung $\hat{\mathbf{D}}$ aus (362.3) ergibt sich mit (362.5) aus (361.7) die Varianz der Schätzung $\widehat{\mathbf{p}'\boldsymbol{\sigma}}$. Um die Kovarianz zweier Schätzungen zu erhalten, sei die Schätzung $\widehat{\mathbf{t}'\boldsymbol{\sigma}}$ der linearen Funktion $\mathbf{t}'\boldsymbol{\sigma}$ der Varianz- und Kovarianzkomponenten $\boldsymbol{\sigma}$ eingeführt, die nach (362.3) und (362.5) mit

$$\widehat{\mathbf{t}'\boldsymbol{\sigma}} = \mathbf{y}'\hat{\mathbf{D}}\mathbf{y} \quad \text{und} \quad \hat{\mathbf{D}} = \sum_{j=1}^k \bar{\lambda}_j \mathbf{W}\mathbf{V}_j\mathbf{W}$$

erhalten werde. Aus (271.2) folgt dann die Kovarianz der beiden Schätzungen, wobei $\boldsymbol{\Sigma}$ durch $\boldsymbol{\Sigma}_0$ aus (361.9) zu ersetzen ist, so daß die aufgrund der Näherungswerte α_i^2 und α_{ij} berechenbare Kovarianz $C(\widehat{\mathbf{p}'\boldsymbol{\sigma}}, \widehat{\mathbf{t}'\boldsymbol{\sigma}})$ sich ergibt zu

$$C(\widehat{\mathbf{p}'\boldsymbol{\sigma}}, \widehat{\mathbf{t}'\boldsymbol{\sigma}}) = 2\text{sp} \left(\left(\sum_{i=1}^k \lambda_i \mathbf{W}\mathbf{V}_i\mathbf{W} \right) \boldsymbol{\Sigma}_0 \left(\sum_{j=1}^k \bar{\lambda}_j \mathbf{W}\mathbf{V}_j\mathbf{W} \right) \boldsymbol{\Sigma}_0 \right).$$

Mit (137.3), (137.7) und mit $\boldsymbol{\Sigma}_0 = \mathbf{G}\mathbf{G}'$ nach (143.5) folgt weiter

$$\begin{aligned} C(\widehat{\mathbf{p}'\boldsymbol{\sigma}}, \widehat{\mathbf{t}'\boldsymbol{\sigma}}) &= 2\text{sp} \left(\mathbf{G}' \left(\sum_{i=1}^k \lambda_i \mathbf{W}\mathbf{V}_i\mathbf{W} \right) \mathbf{G}\mathbf{G}' \left(\sum_{j=1}^k \bar{\lambda}_j \mathbf{W}\mathbf{V}_j\mathbf{W} \right) \mathbf{G} \right) \\ &= 2(\text{vec} \sum_{i=1}^k \lambda_i \mathbf{G}'\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{G})' \text{vec} \left(\sum_{j=1}^k \bar{\lambda}_j \mathbf{G}'\mathbf{W}\mathbf{V}_j\mathbf{W}\mathbf{G} \right) \\ &= 2 \left(\sum_{i=1}^k \lambda_i \text{vec} \mathbf{G}'\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{G} \right)' \left(\sum_{j=1}^k \bar{\lambda}_j \text{vec} \mathbf{G}'\mathbf{W}\mathbf{V}_j\mathbf{W}\mathbf{G} \right) \\ &= 2(\mathbf{F}\boldsymbol{\lambda})' \mathbf{F}\bar{\boldsymbol{\lambda}} = 2\boldsymbol{\lambda}'\mathbf{S}\bar{\boldsymbol{\lambda}} = 2(\boldsymbol{\lambda}'\mathbf{S})\mathbf{S}^{-1}(\mathbf{S}\bar{\boldsymbol{\lambda}}). \end{aligned}$$

Mit $\mathbf{S}'\boldsymbol{\lambda} = \mathbf{p}$, $\mathbf{S}'\bar{\boldsymbol{\lambda}} = \mathbf{t}$ und $\mathbf{S}' = \mathbf{S}$ erhält man schließlich

$$C(\widehat{\mathbf{p}'\boldsymbol{\sigma}}, \widehat{\mathbf{t}'\boldsymbol{\sigma}}) = 2\mathbf{p}'\mathbf{S}^{-1}\mathbf{t}. \quad (362.10)$$

Für $\mathbf{p}, \mathbf{t} \in R(\mathbf{S})$ ist $C(\widehat{\mathbf{p}'\boldsymbol{\sigma}}, \widehat{\mathbf{t}'\boldsymbol{\sigma}})$ eindeutig bestimmt; dies läßt sich wie für (362.6) zeigen. Gilt $\mathbf{p}' = |0, \dots, 0, 1, 0, \dots, 0| \in R(\mathbf{S})$, wobei die i -te Komponente gleich Eins und $\hat{\sigma}_{mn}$ die i -te Komponente von $\hat{\boldsymbol{\sigma}}$ ist, und gilt entsprechend $\mathbf{t}' = |0, \dots, 0, 1, 0, \dots, 0| \in R(\mathbf{S})$, wobei die j -te Komponente gleich Eins und $\hat{\sigma}_{op}$ die j -te Komponente von $\hat{\boldsymbol{\sigma}}$ ist, dann folgt

$$C(\hat{\sigma}_{mn}, \hat{\sigma}_{op}) = 2s_{ij} \quad \text{mit} \quad \mathbf{S}^{-1} = (s_{ij}) \quad \text{oder} \quad \mathbf{S}^{-1} = (s_{ij}). \quad (362.11)$$

Es bleibt noch zu zeigen, daß die Lösung $\hat{\mathbf{D}}$ aus (362.3) auf eine beste Schätzung führt, daß also die zweite Bedingung in (312.2) $V(\mathbf{y}'\hat{\mathbf{D}}\mathbf{y}) \leq V(\mathbf{y}'\mathbf{D}^*\mathbf{y})$ gilt, wobei \mathbf{D}^* eine beliebige symmetrische Matrix bedeutet, die (361.6) und (361.8) erfüllt. Mit (233.13) erhält man $V(\mathbf{y}'\mathbf{D}^*\mathbf{y}) = V(\mathbf{y}'(\mathbf{D}^* - \hat{\mathbf{D}} + \hat{\mathbf{D}})\mathbf{y}) = V(\mathbf{y}'(\mathbf{D}^* - \hat{\mathbf{D}})\mathbf{y}) + V(\mathbf{y}'\hat{\mathbf{D}}\mathbf{y}) + 2f$, woraus die obige Bedingung folgt, da die Varianz $V(\mathbf{y}'(\mathbf{D}^* - \hat{\mathbf{D}})\mathbf{y}) \geq 0$ und $f = 0$ ist. Es gilt nämlich mit (271.2) $f = 2\text{sp}((\mathbf{D}^* - \hat{\mathbf{D}})\boldsymbol{\Sigma}_0\hat{\mathbf{D}}\boldsymbol{\Sigma}_0)$ und mit (137.3) und (362.3) $f = 2\sum_{i=1}^k \lambda_i \text{sp}((\mathbf{D}^* - \hat{\mathbf{D}})\boldsymbol{\Sigma}_0\mathbf{W}\mathbf{V}_i\mathbf{W}\boldsymbol{\Sigma}_0) = 2\sum_{i=1}^k \lambda_i \text{sp}(\mathbf{W}\boldsymbol{\Sigma}_0(\mathbf{D}^* - \hat{\mathbf{D}})\boldsymbol{\Sigma}_0\mathbf{W}\mathbf{V}_i)$. Mit (361.8) und (362.2) folgt schließlich $f = 2\sum_{i=1}^k \lambda_i \text{sp}((\mathbf{D}^* - \hat{\mathbf{D}})\mathbf{V}_i) = 2\sum_{i=1}^k \lambda_i(p_i - p_i) = 0$.

Zusammenfassend gilt der

Satz: Unter der Voraussetzung normalverteilter Beobachtungen existieren die lokal besten invarianten quadratischen erwartungstreuen Schätzungen $\hat{\boldsymbol{\sigma}}$ der Varianz- und Kovarianzkomponenten $\boldsymbol{\sigma}$ genau dann, wenn die Matrix \mathbf{S} regulär ist. In diesem Fall werden die Schätzungen $\hat{\boldsymbol{\sigma}}$ durch (362.9) und ihre aufgrund der Näherungswerte α_i^2 und α_{ij} berechenbaren Kovarianzen durch (362.11) eindeutig bestimmt. Ist \mathbf{S} singulär, so existiert die Schätzung der i -ten Komponente von $\boldsymbol{\sigma}$ für ein $i \in \{1, \dots, k\}$ genau dann, wenn für $\mathbf{p}' = |0, \dots, 0, 1, 0, \dots, 0|$ mit dem Wert Eins als i -te Komponente $\mathbf{p} \in R(\mathbf{S})$ gilt. In diesem Fall wird die i -te Komponente von $\boldsymbol{\sigma}$ durch (362.8) und ihre Varianz durch (362.11) mit Hilfe einer beliebigen generalisierten Inversen \mathbf{S}^{-} eindeutig bestimmt. (362.12)

Die Matrix \mathbf{D} wurde lediglich als symmetrisch vorausgesetzt, so daß $\hat{\mathbf{D}}$ aus (362.3) weder positiv definit noch positiv semidefinit zu sein braucht und daher negative Varianzkomponenten geschätzt werden können. Diese Schätzwerte lassen sich ebenso wie Schätzwerte, die keine positiv definite Kovarianzmatrix $\boldsymbol{\Sigma}$ ergeben, nicht weiter verwenden. Sind diese Schätzwerte nicht durch ungünstig gewählte Näherungen oder zu wenige Beobachtungen \mathbf{y} ver-

ursacht, dann ist das Modell (361.3) zu ändern, da es vermutlich die Genauigkeit der Beobachtungen nicht repräsentiert. Es läßt sich auch die Bedingung aufstellen, daß die Varianzkomponenten nicht negativ geschätzt werden, doch diese Schätzungen existieren im Vergleich zu den abgeleiteten Schätzungen nur für eine kleinere Anzahl von Aufgabenstellungen in der Varianzkomponentenschätzung (LAMOTTE 1973A; PUKELSHEIM 1979).

Leitet man die Schätzungen für die Parameter β und die Varianz- und Kovarianzkomponenten σ im Modell (361.3) nicht, wie hier geschehen, voneinander getrennt, sondern gemeinsam ab, ergeben sich identische Ergebnisse (KLEFFE 1978). Schätzwerte für die Varianz- und Kovarianzkomponenten im Gauß-Helmert-Modell mit singulärer Kovarianzmatrix für die Beobachtungen sind bei (SJÖBERG 1985) angegeben.

363 Iterierte Schätzungen

Um die Schätzungen $\hat{\sigma}_i^2$ oder $\hat{\sigma}_{ij}$, falls sie existieren, und ihre Varianzen und Kovarianzen unabhängig von den gewählten Näherungswerten α_i^2 und α_{ij} zu erhalten, sind mit der ersten Schätzung $(\hat{\sigma}_i^2)_1 = \hat{\sigma}_i^2$ und mit $(\alpha_i^2)_1 = \alpha_i^2$ die Produkte $(\hat{\sigma}_i^2)_1(\alpha_i^2)_1$ als neue Näherungswerte $(\alpha_i^2)_2 = (\hat{\sigma}_i^2)_1(\alpha_i^2)_1$ in eine zweite Schätzung der Varianzkomponenten mit dem Ergebnis $(\hat{\sigma}_i^2)_2$ einzuführen. Für die Kovarianzkomponenten gilt entsprechend $(\alpha_{ij})_2 = (\hat{\sigma}_{ij})_1(\alpha_{ij})_1$. Es ist dann, Konvergenz vorausgesetzt, so lange zu iterieren, bis nach m Iterationen $(\hat{\sigma}_i^2)_m = 1$ und $(\hat{\sigma}_{ij})_m = 1$ für alle k Komponenten erhalten wird. Im Punkt der Konvergenz reproduzieren sich die Schätzungen, da die Schätzwerte gleich den Näherungswerten werden. Die Produkte $(\hat{\sigma}_i^2)_m(\alpha_i^2)_m$ und $(\hat{\sigma}_{ij})_m(\alpha_{ij})_m$ liefern im Konvergenzpunkt die Varianz- und Kovarianzkomponenten der iterierten Schätzung mit

$$(\alpha_i^2)_m = \alpha_i^2 \prod_{n=1}^{m-1} (\hat{\sigma}_i^2)_n \quad \text{und} \quad (\alpha_{ij})_m = \alpha_{ij} \prod_{n=1}^{m-1} (\hat{\sigma}_{ij})_n \quad (363.1)$$

für

$$(\hat{\sigma}_i^2)_m = 1 \quad \text{und} \quad (\hat{\sigma}_{ij})_m = 1.$$

Diese Schätzungen sind unabhängig von den gewählten Näherungswerten α_i^2 und α_{ij} , da sie, Konvergenz vorausgesetzt, eindeutig für alle α_i^2 und α_{ij} folgen, die nicht zu stark voneinander abweichen, da nach Voraussetzung α_i^2 und α_{ij} Näherungswerte von $\sigma_i^2 \alpha_i^2$ und $\sigma_{ij} \alpha_{ij}$ sein sollen.

Im Konvergenzpunkt ergeben sich auch die Varianzen $V(\hat{\sigma}_i^2)$ und $V(\hat{\sigma}_{ij})$ unabhängig von den Näherungswerten. Sie berechnen sich dann mit den Schätzwerten der Varianz- und Kovarianzkomponenten, so daß die Varianzen

$\hat{V}((\hat{\sigma}_i^2)_m)$ und $\hat{V}((\hat{\sigma}_{ij})_m)$ erhalten werden, also

$$\hat{V}((\hat{\sigma}_i^2)_m) = V((\hat{\sigma}_i^2)_m) \quad \text{und} \quad \hat{V}((\hat{\sigma}_{ij})_m) = V((\hat{\sigma}_{ij})_m) \\ \text{für } (\hat{\sigma}_i^2)_m = (\hat{\sigma}_{ij})_m = 1. \quad (363.2)$$

Die Varianzen der Produkte $(\hat{\sigma}_i^2)_m(\alpha_i^2)_m$ und $(\hat{\sigma}_{ij})_m(\alpha_{ij})_m$ im Konvergenzpunkt, aus denen nach (363.1) die Varianz- und Kovarianzkomponenten der iterierten Schätzung folgen, lassen sich unabhängig von den Schätzungen der vorhergegangenen Iterationen angeben, denn die Näherungswerte $(\alpha_i^2)_m$ und $(\alpha_{ij})_m$ könnte man beispielsweise auch durch systematische Variation von α_i^2 und α_{ij} erhalten haben. Mit (233.2) und (363.2) folgt daher für die Varianzen der im Konvergenzpunkt geschätzten Varianz- und Kovarianzkomponenten

$$\hat{V}((\hat{\sigma}_i^2)_m(\alpha_i^2)_m) = (\alpha_i^2)_m^2 \hat{V}((\hat{\sigma}_i^2)_m) \quad \text{für } (\hat{\sigma}_i^2)_m = 1, \\ \hat{V}((\hat{\sigma}_{ij})_m(\alpha_{ij})_m) = (\alpha_{ij})_m^2 \hat{V}((\hat{\sigma}_{ij})_m) \quad \text{für } (\hat{\sigma}_{ij})_m = 1 \quad (363.3)$$

und mit (233.15) und (362.11) ganz entsprechend die Kovarianzen.

Im Konvergenzpunkt erhält man aus (362.9) $\mathbf{S}\hat{\boldsymbol{\sigma}} = \mathbf{q}$ mit $\hat{\boldsymbol{\sigma}} = |1, \dots, 1|'$ oder $\sum_{j=1}^k \text{sp}(\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{V}_j) = q_i$ und aus (361.9) und (362.7)

$$\text{sp}(\mathbf{W}\mathbf{V}_i) = q_i \quad \text{für } i \in \{1, \dots, k\}.$$

Die Schätzung

$$\hat{\boldsymbol{\sigma}} = \mathbf{H}\mathbf{q} \quad \text{mit} \quad \mathbf{H} = (\text{diag}[1/\text{sp}(\mathbf{W}\mathbf{V}_i)]) \\ \text{und} \quad \mathbf{q} = (\hat{\boldsymbol{e}}'\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_i\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{e}}), \quad (363.4)$$

die bereits bei (FÖRSTNER 1979B) auf unterschiedliche Weise abgeleitet wurde, führt daher, falls sie iterativ angewendet wird und falls die Iterationen konvergieren, im Konvergenzpunkt auf die gleichen Ergebnisse wie die iterierten Schätzungen nach (362.9). Der Rechenaufwand für (363.4) ist geringer als der für (362.9), beispielsweise benötigt man für diagonale Matrizen \mathbf{V}_i und $\boldsymbol{\Sigma}_0$ lediglich die Diagonalelemente von \mathbf{W} . Der geringere Rechenaufwand muß allerdings mit einer im allgemeinen langsameren Konvergenz der Iterationen erkauft werden. Lassen sich die Matrizen \mathbf{V}_i zerlegen in zum Beispiel $\mathbf{V}_i = \mathbf{B}_i\mathbf{B}_i'$, was nach (143.10) und (361.2) für die Matrizen \mathbf{V}_i der Fall ist, die den Varianzkomponenten σ_i^2 zugeordnet werden, ergeben sich positive Schätzwerte aus (363.4). Der Zähler läßt sich nämlich als Quadratsumme darstellen

$$\hat{\boldsymbol{e}}'\boldsymbol{\Sigma}_0^{-1}\mathbf{V}_i\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{e}} = (\mathbf{B}_i'\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{e}})'(\mathbf{B}_i'\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{e}})$$

und ebenso mit (361.9) und (362.7) der Nenner

$$\begin{aligned} \text{sp}(\mathbf{W}\mathbf{V}_i) &= \text{sp}(\mathbf{W}\mathbf{V}_i\mathbf{W}\boldsymbol{\Sigma}_0) = \sum_{j=1}^k \text{sp}(\mathbf{W}\mathbf{V}_i\mathbf{W}\mathbf{V}_j) \\ &= \sum_{j=1}^k \text{sp}(\mathbf{B}'_j\mathbf{W}\mathbf{B}_i\mathbf{B}'_i\mathbf{W}\mathbf{B}_j) = \sum_{j=1}^k \text{sp}((\mathbf{B}'_i\mathbf{W}\mathbf{B}_j)'(\mathbf{B}'_i\mathbf{W}\mathbf{B}_j)). \end{aligned}$$

Auch mit der Maximum-Likelihood-Methode kann man (362.9) und (363.4) ableiten (KOCH 1986; SCHAFFRIN 1983, S.86, 99).

Beispiel: In dem Modell

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad \boldsymbol{\Sigma} = \sigma_1^2\alpha_1^2\mathbf{T}_1 + \sigma_2^2\alpha_2^2\mathbf{T}_2 = \sigma_1^2\mathbf{V}_1 + \sigma_2^2\mathbf{V}_2$$

sowie $\boldsymbol{\beta} = \beta_1$, $\alpha_1^2 = 1 \cdot 10^{-4}$, $\alpha_2^2 = 4 \cdot 10^{-4}$ und

$$\mathbf{X} = \begin{vmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{vmatrix}, \quad \mathbf{y} = \begin{vmatrix} 1,25 \\ 1,26 \\ 1,24 \\ 1,22 \\ 1,27 \end{vmatrix}, \quad \mathbf{V}_1 = 1 \cdot 10^{-4} \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{vmatrix},$$

$$\mathbf{V}_2 = 4 \cdot 10^{-4} \begin{vmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

sollen die Varianzkomponenten σ_1^2 und σ_2^2 geschätzt werden. Das Problem besteht also darin, für die beiden unkorrelierten Beobachtungsgruppen 1,25; 1,26; 1,24 und 1,22; 1,27, deren Gewichtsmatrizen bekannt sind, die gemeinsame Kovarianzmatrix anzugeben. Man erhält \mathbf{W} aus (362.3) mit $\boldsymbol{\Sigma}_0 = \mathbf{V}_1 + \mathbf{V}_2$ aus (361.9) zu

$$\mathbf{W} = 10^4 \begin{vmatrix} 0,7143 & -0,2857 & -0,2857 & -0,07143 & -0,07143 \\ & 0,7143 & -0,2857 & -0,07143 & -0,07143 \\ & & 0,7143 & -0,07143 & -0,07143 \\ & & & 0,23214 & -0,01786 \\ & & & & 0,23214 \end{vmatrix}$$

und weiter

$$\begin{aligned} \text{sp}(\mathbf{W}\mathbf{V}_1\mathbf{W}\mathbf{V}_1) &= 2,0204, \quad \text{sp}(\mathbf{W}\mathbf{V}_1\mathbf{W}\mathbf{V}_2) = 0,1224, \\ \text{sp}(\mathbf{W}\mathbf{V}_2\mathbf{W}\mathbf{V}_2) &= 1,7347, \\ \mathbf{y}'\mathbf{W}\mathbf{V}_1\mathbf{W}\mathbf{y} &= 2,0153, \quad \mathbf{y}'\mathbf{W}\mathbf{V}_2\mathbf{W}\mathbf{y} = 3,2168, \end{aligned}$$

so daß \mathbf{S} wegen $\det \mathbf{S} \neq 0$ regulär ist und aus (362.9) und (362.11) folgt

$$\begin{aligned}(\hat{\sigma}_1^2)_1 &= 0,8889, \quad V((\hat{\sigma}_1^2)_1) = 0,9942, \\(\hat{\sigma}_2^2)_1 &= 1,7917, \quad V((\hat{\sigma}_2^2)_1) = 1,1579.\end{aligned}$$

Mit $(\alpha_1^2)_2 = (\hat{\sigma}_1^2)_1 \alpha_1^2$ und $(\alpha_2^2)_2 = (\hat{\sigma}_2^2)_1 \alpha_2^2$ als Näherungswerten wird erneut eine Schätzung der Varianzkomponenten durchgeführt. Man erhält

$$\mathbf{V}_1 = 0,8889 \cdot 10^{-4} \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{vmatrix},$$

$$\mathbf{V}_2 = 7,1667 \cdot 10^{-4} \begin{vmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix}$$

und damit

$$\mathbf{W} = 10^4 \begin{vmatrix} 0,7786 & -0,3464 & -0,3464 & -0,04296 & -0,04296 \\ & 0,7786 & -0,3464 & -0,04296 & -0,04296 \\ & & 0,7786 & -0,04296 & -0,04296 \\ & & & 0,13421 & -0,00533 \\ & & & & 0,13421 \end{vmatrix}$$

und weiter

$$\begin{aligned}\text{sp}(\mathbf{WV}_1 \mathbf{WV}_1) &= 2,0058, \quad \text{sp}(\mathbf{WV}_1 \mathbf{WV}_2) = 0,0705, \\ \text{sp}(\mathbf{WV}_2 \mathbf{WV}_2) &= 1,8531, \\ \mathbf{y}' \mathbf{WV}_1 \mathbf{Wy} &= 2,2549, \quad \mathbf{y}' \mathbf{WV}_2 \mathbf{Wy} = 1,8037,\end{aligned}$$

so daß sich aus (362.9) und (362.11) ergibt

$$\begin{aligned}(\hat{\sigma}_1^2)_2 &= 1,0914, \quad V((\hat{\sigma}_1^2)_2) = 0,9984, \\(\hat{\sigma}_2^2)_2 &= 0,9318, \quad V((\hat{\sigma}_2^2)_2) = 1,0808\end{aligned}$$

und somit

$$\begin{aligned}(\alpha_1^2)_3 &= (\hat{\sigma}_1^2)_2 (\alpha_1^2)_2 = 0,9701 \cdot 10^{-4}, \\(\alpha_2^2)_3 &= (\hat{\sigma}_2^2)_2 (\alpha_2^2)_2 = 6,6779 \cdot 10^{-4}.\end{aligned}$$

Die dritte Iteration ergibt bereits Schätzwerte für die Varianzkomponenten, die nahe bei Eins liegen. Mit (363.2) folgen dann

$$\begin{aligned}(\hat{\sigma}_1^2)_3 &= 0,9892, \quad \hat{V}((\hat{\sigma}_1^2)_3) = 0,9979, \\(\hat{\sigma}_2^2)_3 &= 1,0127, \quad \hat{V}((\hat{\sigma}_2^2)_3) = 1,0942\end{aligned}$$

und die Varianzkomponenten der iterierten Schätzung

$$(\hat{\sigma}_1^2)_3(\alpha_1^2)_3 = 0,9596 \cdot 10^{-4}, \quad (\hat{\sigma}_2^2)_3(\alpha_2^2)_3 = 6,7627 \cdot 10^{-4}$$

mit ihren Standardabweichungen aus (363.3)

$$\begin{aligned}[\hat{V}((\hat{\sigma}_1^2)_3(\alpha_1^2)_3)]^{1/2} &= 0,9586 \cdot 10^{-4}, \\[\hat{V}((\hat{\sigma}_2^2)_3(\alpha_2^2)_3)]^{1/2} &= 7,0741 \cdot 10^{-4}.\end{aligned}$$

Die iterierten Schätzungen der Varianzkomponenten werden also für das Beispiel rasch erhalten.

364 Beste erwartungstreue Schätzung der Varianz der Gewichtseinheit

Die Schätzung $\hat{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit im Gauß–Markoff–Modell $\mathbf{X}\boldsymbol{\beta} = \mathbf{E}(\mathbf{y})$ mit $D(\mathbf{y}) = \sigma^2\boldsymbol{\Sigma}_0$ und nicht vollem Rang soll nun unter der Annahme der Normalverteilung für die Beobachtungen als beste invariante quadratische erwartungstreue Schätzung abgeleitet werden. Mit $k = 1$, $\mathbf{V}_1 = \boldsymbol{\Sigma}_0$ und $\text{rg}\mathbf{X} = q$ ergibt sich wegen (152.3), (153.4) und (362.7)

$$\begin{aligned}\text{sp}(\mathbf{W}\boldsymbol{\Sigma}_0\mathbf{W}\boldsymbol{\Sigma}_0) &= \text{sp}(\mathbf{W}\boldsymbol{\Sigma}_0) \\ &= n - \text{sp}((\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X}) = n - q.\end{aligned}\quad (364.1)$$

Hiermit erhält man aus (362.9) in Übereinstimmung mit (331.12)

$$\hat{\sigma}^2 = \Omega/(n - q) \quad \text{mit} \quad \Omega = \mathbf{y}'\mathbf{W}\mathbf{y}, \quad (364.2)$$

da, wie bereits bei (362.5) erwähnt, \mathbf{W} mit der Matrix der quadratischen Form für Ω im Modell $\mathbf{X}\boldsymbol{\beta} = \mathbf{E}(\mathbf{y})$ mit $D(\mathbf{y}) = \sigma^2\boldsymbol{\Sigma}_0$ identisch ist. Mit $\hat{\sigma}^2\boldsymbol{\Sigma}_0$ anstelle von $\boldsymbol{\Sigma}_0$ in (361.9) folgt unabhängig von einem Näherungsfaktor in $\boldsymbol{\Sigma}_0$ die Schätzung $(\hat{\sigma}^2)_1 = 1$ schon nach der ersten Iteration, so daß eine gleichmäßig beste Schätzung vorliegt.

Die Varianz der Schätzung $(\hat{\sigma}^2)_1 = 1$ ergibt sich mit (362.11) und (363.2) zu $\hat{V}((\hat{\sigma}^2)_1) = 2/(n - q)$. Für das Produkt $\hat{\sigma}^2(\hat{\sigma}^2)_1 = \hat{\sigma}^2$ folgt dann aus (363.3) die mit der Schätzung $\hat{\sigma}^2$ berechnete Varianz $\hat{V}(\hat{\sigma}^2)$ von $\hat{\sigma}^2$

$$\hat{V}(\hat{\sigma}^2) = 2(\hat{\sigma}^2)^2/(n - q). \quad (364.3)$$

Diese Beziehung läßt sich auch wie folgt ableiten. Mit (271.2) ergibt sich für die Varianz von Ω in (364.2) $V(\Omega) = 2(\sigma^2)^2 \text{sp}(\mathbf{W}\Sigma_0\mathbf{W}\Sigma_0) + 4\sigma^2\boldsymbol{\beta}'\mathbf{X}'\mathbf{W}\Sigma_0\mathbf{W}\mathbf{X}\boldsymbol{\beta}$ und mit (364.1) und $\mathbf{W}\mathbf{X} = \mathbf{0}$, wie bei (362.2) erwähnt, $V(\Omega) = 2(\sigma^2)^2(n - q)$. Mit (233.2) und (364.2) folgt dann

$$V(\hat{\sigma}^2) = 2(\sigma^2)^2/(n - q) . \quad (364.4)$$

Hieraus ergibt sich die mit der geschätzten Varianz $\hat{\sigma}^2$ berechnete Varianz $\hat{V}(\hat{\sigma}^2)$ in Übereinstimmung mit (364.3).

Zusammenfassend gilt der

Satz: Unter der Voraussetzung normalverteilter Beobachtungen sind die gleichmäßig beste invariante quadratische erwartungstreue Schätzung $\hat{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit und ihre mit $\hat{\sigma}^2$ berechnete Varianz $\hat{V}(\hat{\sigma}^2)$ in dem Gauß–Markoff–Modell mit nicht vollem Rang $\mathbf{X}\boldsymbol{\beta} = \mathbf{E}(\mathbf{y})$ und $D(\mathbf{y}) = \sigma^2\Sigma_0$ durch (364.2) und (364.3) gegeben. (364.5)

37 Multivariate Parameterschätzung

371 Multivariates Gauß–Markoff–Modell

Sind für eine Koeffizientenmatrix in einem Gauß–Markoff–Modell anstelle eines Beobachtungsvektors mehrere Beobachtungsvektoren mit identischen Gewichtsmatrizen gegeben und die entsprechende Anzahl von Parametervektoren gesucht, spricht man von einem *multivariaten Gauß–Markoff–Modell*. Im Gegensatz hierzu bezeichnet man (321.1) oder (321.5) auch als *univariates Gauß–Markoff–Modell*. Die Varianzanalyse wird häufig in multivariaten Modellen vorgenommen und zwar dann, wenn die Effekte der Faktoren sich nicht nur durch ein Merkmal, sondern durch mehrere Merkmale erklären lassen. Man spricht dann von *multivariater Varianzanalyse*. Sollen beispielsweise die Effekte verschiedener Regionen auf die Entwicklung von Tierarten untersucht werden, kann als ein Merkmal das Gewicht und als weiteres Merkmal die Größe der zu untersuchenden Tierarten dienen.

Multivariate Modelle lassen sich ebenfalls aufstellen, wenn zur Erfassung eines Geschehens Beobachtungen in zeitlichen Abständen wiederholt werden. Laufen die Wiederholungsmessungen, beispielsweise zur Überwachung von Veränderungen an Bauwerken, bei identischen Gewichtsmatrizen jeweils unter dem gleichen Meßprogramm ab, bleibt die Koeffizientenmatrix im Gauß–Markoff–Modell für jede Wiederholungsmessung unverändert, und jeweils eine Wiederholungsmessung entspricht einem Merkmal.

Definition: Es sei \mathbf{X} eine $n \times u$ Matrix gegebener Koeffizienten mit $\text{rg } \mathbf{X} = q \leq u$, β_i mit $i \in \{1, \dots, p\}$ seien $u \times 1$ Vektoren fester, unbekannter Parameter, \mathbf{y}_i die $n \times 1$ Zufallsvektoren der Beobachtungen von p Merkmalen mit $p \leq n - q$, und es gelte $C(\mathbf{y}_i, \mathbf{y}_j) = \sigma_{ij} \mathbf{I}_n$, wobei die Kovarianzmatrix Σ mit $\Sigma = (\sigma_{ij})$ unbekannt und positiv definit sei, dann bezeichnet man

$$\mathbf{X}\beta_i = E(\mathbf{y}_i) \quad \text{mit} \quad C(\mathbf{y}_i, \mathbf{y}_j) = \sigma_{ij} \mathbf{I}_n \quad \text{für} \quad i, j \in \{1, \dots, p\}$$

als *multivariates Gauß–Markoff–Modell*. (371.1)

Mit $D(\mathbf{y}_i) = C(\mathbf{y}_i, \mathbf{y}_i) = \sigma_i^2 \mathbf{I}_n$ folgt, daß die Elemente der Beobachtungsvektoren \mathbf{y}_i unkorreliert sind und gleiche Varianzen besitzen. Trifft diese Voraussetzung nicht zu, kann man von dem (321.1) entsprechenden multivariaten Modell mit identischen Gewichtsmatrizen für die Beobachtungsvektoren ausgehen

$$\bar{\mathbf{X}}\bar{\beta}_i = E(\bar{\mathbf{y}}_i) \quad \text{mit} \quad C(\bar{\mathbf{y}}_i, \bar{\mathbf{y}}_j) = \sigma_{ij} \mathbf{P}^{-1} \\ \text{für} \quad i, j \in \{1, \dots, p\}, \quad (371.2)$$

das sich durch die Transformation (321.4) auf das Modell (371.1) zurückführen läßt.

Faßt man die p Vektoren β_i zur $u \times p$ Matrix \mathbf{B} der unbekannt Parameter und die p Vektoren \mathbf{y}_i zur $n \times p$ Matrix \mathbf{Y} der Beobachtungen zusammen, kann das Modell (371.1) wegen (131.22) und (137.5) auch wie folgt formuliert werden

$$\mathbf{X}\mathbf{B} = E(\mathbf{Y}) \quad \text{mit} \quad D(\text{vec } \mathbf{Y}) = \Sigma \otimes \mathbf{I}_n, \quad (371.3)$$

worin die Kovarianzmatrix $\Sigma \otimes \mathbf{I}_n$ wegen (143.2) positiv definit ist, denn ihre Eigenwerte ergeben sich aus dem Produkt der Eigenwerte der Matrizen Σ und \mathbf{I}_n (ANDERSON 1958, S.348), die positiv definit sind und daher positive Eigenwerte besitzen.

Die Zeilen der Matrix \mathbf{Y} seien durch die $1 \times p$ Vektoren \mathbf{z}'_k mit $k \in \{1, \dots, n\}$ gebildet, die jeweils die Beobachtungen der p Merkmale enthalten und daher als *Merkmalsvektoren* bezeichnet werden. Mit $D(\text{vec } \mathbf{Y}) = \Sigma \otimes \mathbf{I}_n$ folgt dann

$$D(\mathbf{z}_k) = \Sigma \quad \text{und} \quad C(\mathbf{z}_k, \mathbf{z}_l) = \mathbf{0} \\ \text{für} \quad k \neq l \quad \text{und} \quad k, l \in \{1, \dots, n\}. \quad (371.4)$$

Im Modell (371.1) wird also vorausgesetzt, daß die Elemente der Spalten der Matrix \mathbf{Y} , die die Beobachtungen eines Merkmals enthalten, unkorreliert sind, während die Elemente der Zeilen der Matrix \mathbf{Y} , die die Beobachtungen der

verschiedenen Merkmale enthalten, um gleiche Beträge in den verschiedenen Zeilen miteinander korreliert sind.

Nimmt man für die Beobachtungen die Normalverteilung an, ergibt sich mit (251.1) und (371.3)

$$\text{vec}\mathbf{Y} \sim N(\text{vec}(\mathbf{X}\mathbf{B}), \mathbf{\Sigma} \otimes \mathbf{I}_n). \quad (371.5)$$

Für die Beobachtungsvektoren \mathbf{y}_i folgt dann nach (253.1)

$$\mathbf{y}_i \sim N(\mathbf{X}\boldsymbol{\beta}_i, \sigma_i^2 \mathbf{I}_n) \quad \text{für } i \in \{1, \dots, p\} \quad (371.6)$$

und für die Merkmalsvektoren \mathbf{z}_k

$$\mathbf{z}_k \sim N(\mathbf{B}'\mathbf{x}_k, \mathbf{\Sigma}) \quad \text{für } k \in \{1, \dots, n\}, \quad (371.7)$$

denn der Erwartungswertvektor von \mathbf{z}_k berechnet sich aus (371.3) mit $\mathbf{X}' = |\mathbf{x}_1, \dots, \mathbf{x}_n|$ zu $\mathbf{B}'\mathbf{x}_k$. Wegen (254.1) und (371.4) sind die Merkmalsvektoren \mathbf{z}_k und \mathbf{z}_l mit $k \neq l$ voneinander unabhängig.

Im multivariaten Gauß–Markoff–Modell sind sowohl die Parametervektoren $\boldsymbol{\beta}_i$ als auch die Kovarianzen σ_{ij} der Kovarianzmatrix $\mathbf{\Sigma}$ unbekannt. Ihre Schätzung wird in den beiden folgenden Kapiteln behandelt.

372 Schätzung der Parametervektoren

Trotz der gemeinsamen Kovarianzmatrix $D(\text{vec}\mathbf{Y})$ für die Beobachtungsvektoren \mathbf{y}_i mit $i \in \{1, \dots, p\}$ kann man sich für die Schätzung der unbekannt Parametervektoren $\boldsymbol{\beta}_i$ das multivariate Gauß–Markoff–Modell (371.1) in p univariate Modelle zerfallen denken, so daß auch für das multivariate Modell die Sätze (324.4), (331.4), (332.7) und (333.11) gelten.

Die Richtigkeit dieser Aussagen läßt sich durch die Umschreibung des Modells (371.3) mittels (137.6) in das folgende univariate Modell beweisen

$$(\mathbf{I}_p \otimes \mathbf{X})\text{vec}\mathbf{B} = E(\text{vec}\mathbf{Y}) \quad \text{mit} \quad D(\text{vec}\mathbf{Y}) = \mathbf{\Sigma} \otimes \mathbf{I}_n. \quad (372.1)$$

Da \mathbf{X} beliebigen Rang besitzt, ergibt sich aus (331.3) entsprechend (322.9), falls $\text{vec}\bar{\mathbf{B}}$ die Schätzung von $\text{vec}\mathbf{B}$ bedeutet,

$$\text{vec}\bar{\mathbf{B}} = [(\mathbf{I}_p \otimes \mathbf{X})'(\mathbf{\Sigma} \otimes \mathbf{I}_n)^{-1}(\mathbf{I}_p \otimes \mathbf{X})]^{-1}(\mathbf{I}_p \otimes \mathbf{X})'(\mathbf{\Sigma} \otimes \mathbf{I}_n)^{-1}\text{vec}\mathbf{Y}.$$

Mit (131.23), (131.26) und (131.27) folgt weiter $\text{vec}\bar{\mathbf{B}} = (\mathbf{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X})^{-1}(\mathbf{\Sigma}^{-1} \otimes \mathbf{X}')\text{vec}\mathbf{Y}$. Beachtet man, daß wegen (131.26) und (153.1)

$$(\mathbf{\Sigma}^{-1} \otimes \mathbf{X}'\mathbf{X})^{-1} = \mathbf{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1} \quad (372.2)$$

gilt, ergibt sich schließlich

$$\text{vec}\bar{\mathbf{B}} = (\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\text{vec}\mathbf{Y} \quad (372.3)$$

und nach (233.2) die Kovarianzmatrix $D(\text{vec}\bar{\mathbf{B}}) = (\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)(\mathbf{I}_p \otimes \mathbf{X}((\mathbf{X}'\mathbf{X})^{-1})')$ der Schätzung, für die mit (131.26) und (333.2) folgt

$$D(\text{vec}\bar{\mathbf{B}}) = \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}. \quad (372.4)$$

Mit $\text{vec}\bar{\mathbf{B}} = |\bar{\beta}_1, \dots, \bar{\beta}_p|$ erhält man wegen (137.6)

$$\bar{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad \text{oder} \quad \bar{\beta}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i \quad (372.5)$$

in Übereinstimmung mit (331.3). Bezeichnet man mit $\hat{\mathbf{B}}$ die beste lineare erwartungstreue Schätzung der Matrix \mathbf{B} der Parametervektoren, dann gilt für Modelle mit vollem Rang wegen (153.22) und (322.9) $\bar{\mathbf{B}} = \hat{\mathbf{B}}$ mit $\hat{\mathbf{B}} = |\hat{\beta}_1, \dots, \hat{\beta}_p|$.

Die Schätzung $\bar{\mathbf{B}}$ aus (372.5) bedeutet wegen (331.4) die Schätzung nach der Methode (313.1) der kleinsten Quadrate. Folglich wird die quadratische Form $S(\text{vec}\mathbf{B})$, die sich (323.1) entsprechend aus (371.3) zu $S(\text{vec}\mathbf{B}) = (\text{vec}(\mathbf{Y} - \mathbf{X}\mathbf{B}))'(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}\text{vec}(\mathbf{Y} - \mathbf{X}\mathbf{B})$ und mit (131.27) und (137.8) zu

$$S(\text{vec}\mathbf{B}) = \text{sp}[(\mathbf{Y} - \mathbf{X}\mathbf{B})\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{B})'] \quad (372.6)$$

ergibt, für $\bar{\mathbf{B}} = \mathbf{B}$ minimal.

Sind die Beobachtungen normalverteilt, so daß für $\text{vec}\mathbf{Y}$ die Verteilung (371.5) gilt, ergibt sich aus (314.1) und (251.1) die Likelihoodfunktion $L(\text{vec}\mathbf{Y}; \mathbf{B}, \boldsymbol{\Sigma})$ mit den unbekanntem Matrizen \mathbf{B} und $\boldsymbol{\Sigma}$

$$L(\text{vec}\mathbf{Y}; \mathbf{B}, \boldsymbol{\Sigma}) = (2\pi)^{-np/2} \det(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1/2} \exp\{-[\text{vec}(\mathbf{Y} - \mathbf{X}\mathbf{B})]'(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}[\text{vec}(\mathbf{Y} - \mathbf{X}\mathbf{B})]/2\}.$$

Mit $\det(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{1/2} = \det \boldsymbol{\Sigma}^{n/2}$ (ANDERSON 1958, S.348) und (137.8) folgt dann die Likelihoodfunktion

$$L(\text{vec}\mathbf{Y}; \mathbf{B}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}(\det \boldsymbol{\Sigma})^{n/2}} \exp \text{sp}\left[-\frac{1}{2}(\mathbf{Y} - \mathbf{X}\mathbf{B})\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{B})'\right]. \quad (372.7)$$

Zur Maximum-Likelihood-Schätzung (314.2) von \mathbf{B} ist abgesehen vom Faktor 1/2 die mit (372.6) identische quadratische Form zu minimieren, so daß identische Schätzwerte $\bar{\mathbf{B}}$ für \mathbf{B} folgen.

373 Schätzung der Kovarianzmatrix

Wie im univariaten Modell erhält man die Schätzung $\hat{\sigma}_i^2$ der Varianz σ_i^2 der Gewichtseinheit nach (331.12) zu

$$\hat{\sigma}_i^2 = \frac{1}{n-q} (\mathbf{X}\bar{\boldsymbol{\beta}}_i - \mathbf{y}_i)' (\mathbf{X}\bar{\boldsymbol{\beta}}_i - \mathbf{y}_i) \quad \text{für } i \in \{1, \dots, p\}. \quad (373.1)$$

Entsprechend läßt sich auch die Kovarianz σ_{ij} schätzen, so daß man erhält

$$\hat{\sigma}_{ij} = \frac{1}{n-q} (\mathbf{X}\bar{\boldsymbol{\beta}}_i - \mathbf{y}_i)' (\mathbf{X}\bar{\boldsymbol{\beta}}_j - \mathbf{y}_j) \quad \text{für } i, j \in \{1, \dots, p\} \quad (373.2)$$

oder mit $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ij})$ und der $p \times p$ Matrix $\boldsymbol{\Omega}$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-q} \boldsymbol{\Omega} \quad \text{mit } \boldsymbol{\Omega} = (\mathbf{X}\bar{\mathbf{B}} - \mathbf{Y})' (\mathbf{X}\bar{\mathbf{B}} - \mathbf{Y}). \quad (373.3)$$

Somit wird die Kovarianzmatrix $\boldsymbol{\Sigma}$ mit Hilfe der quadratischen und bilinearen Formen der Residuen geschätzt.

Die Schätzung (373.3) ist erwartungstreu, denn mit $\boldsymbol{\Omega} = (\omega_{ij})$ und (331.9) folgt

$$\omega_{ij} = \mathbf{y}_i' (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \mathbf{y}_j \quad (373.4)$$

und mit (271.1) und (371.1)

$$E(\omega_{ij}) = \sigma_{ij} \text{sp}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') + \beta_i' \mathbf{X}' (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \mathbf{X} \beta_j$$

sowie schließlich wegen (153.5) und (331.7)

$$E(\omega_{ij}) = \sigma_{ij} (n - q). \quad (373.5)$$

Eine nicht erwartungstreue Schätzung der Kovarianzmatrix $\boldsymbol{\Sigma}$ aus den quadratischen und bilinearen Formen der Residuen, die der Schätzung (331.8) im univariaten Modell entspricht, erhält man mit der Maximum-Likelihood-Methode (314.2). Wie in (324.3) wird $\ln L(\text{vec} \mathbf{Y}; \mathbf{B}, \boldsymbol{\Sigma})$ anstelle von $L(\text{vec} \mathbf{Y}; \mathbf{B}, \boldsymbol{\Sigma})$ aus (372.7) nach $\boldsymbol{\Sigma}$ differenziert. Mit (136.14), (137.3) und $\boldsymbol{\Sigma}^{-1} = \mathbf{S}$ erhält man

$$\begin{aligned} \ln L(\text{vec} \mathbf{Y}; \mathbf{B}, \mathbf{S}^{-1}) &= -n p \ln(2\pi) / 2 + n \ln \det \mathbf{S} / 2 \\ &\quad - \text{sp}[\mathbf{S}(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})] / 2 \end{aligned}$$

sowie mit (172.3) und (172.4)

$$\partial \ln L(\text{vec} \mathbf{Y}; \mathbf{B}, \mathbf{S}^{-1}) / \partial \mathbf{S} = \mathbf{F} - \text{diag} \mathbf{F} / 2 = \mathbf{0}$$

mit

$$\mathbf{F} = n\mathbf{S}^{-1} - (\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB}),$$

so daß $\mathbf{F} = \mathbf{0}$ folgt. Hieraus ergibt sich mit (372.5) und (373.3) die nicht erwartungstreue Schätzung $\bar{\Sigma}$ von Σ

$$\bar{\Sigma} = \frac{1}{n}\mathbf{\Omega}. \quad (373.6)$$

Die Matrizen $\mathbf{\Omega}$ und $\bar{\Sigma}$ sowie $\hat{\Sigma}$ aus (373.3) können wegen $p \leq n - q$ als positiv definit angenommen werden, da unter der Voraussetzung normalverteilter Beobachtungen die Matrix $\mathbf{\Omega}$ mit der Wahrscheinlichkeit Null positiv semidefinit wird, wie im Zusammenhang mit (412.5) gezeigt wird.

Es soll noch bewiesen werden, daß die Schätzung $\hat{\Sigma}$ nach (373.3) der Kovarianzmatrix Σ auch als Varianz- und Kovarianzkomponentenschätzung nach (362.12) abgeleitet werden kann, so daß $\hat{\Sigma}$ nicht nur eine erwartungstreue Schätzung, wie bereits mit (373.5) gezeigt wurde, und eine invariante Schätzung, was aus (373.4) mit (153.5) und (361.6) folgt, sondern unter der Voraussetzung der Normalverteilung (371.5) für die Beobachtungen $\text{vec}\mathbf{Y}$ auch eine beste Schätzung darstellt. Diese Eigenschaft war für die Schätzung der Varianz der Gewichtseinheit im univariaten Modell mit (364.5) gezeigt worden. Zum Beweis wird das multivariate Modell (371.3) in das univariate Modell (372.1) umgeschrieben und die Kovarianzmatrix Σ für das Modell (361.3) der Varianz- und Kovarianzkomponentenschätzung dargestellt durch

$$\Sigma = \sigma_1^2 \alpha_1^2 \mathbf{e}_1 \mathbf{e}_1' + \sigma_{12} \alpha_{12} (\mathbf{e}_1 \mathbf{e}_2' + \mathbf{e}_2 \mathbf{e}_1') + \dots + \sigma_p^2 \alpha_p^2 \mathbf{e}_p \mathbf{e}_p', \quad (373.7)$$

wobei die $p \times 1$ Vektoren \mathbf{e}_i mit $\mathbf{e}_i = |0, \dots, 0, 1, 0, \dots, 0|'$ als i -te Komponente den Wert Eins erhalten. Für die Kovarianzmatrix $D(\text{vec}\mathbf{Y})$ in (372.1) ergibt sich dann wegen (131.24) entsprechend (361.3)

$$\begin{aligned} D(\text{vec}\mathbf{Y}) &= \Sigma \otimes \mathbf{I}_n = \sigma_1^2 \alpha_1^2 \mathbf{T}_1 + \sigma_{12} \alpha_{12} \mathbf{T}_2 + \dots + \sigma_p^2 \alpha_p^2 \mathbf{T}_k \\ &= \sigma_1^2 \mathbf{V}_1 + \sigma_{12} \mathbf{V}_2 + \dots + \sigma_p^2 \mathbf{V}_k \end{aligned} \quad (373.8)$$

mit $k = p(p+1)/2$ sowie $\mathbf{V}_l = \alpha_{ij} (\mathbf{e}_i \mathbf{e}_j') \otimes \mathbf{I}_n$ für $i = j$ und $\alpha_i^2 = \alpha_{ii}$ sowie $\mathbf{V}_l = \alpha_{ij} (\mathbf{e}_i \mathbf{e}_j' + \mathbf{e}_j \mathbf{e}_i') \otimes \mathbf{I}_n$ für $i \neq j$ oder

$$\begin{aligned} \mathbf{V}_l &= \alpha_{ij} \mathbf{E}_{ij} \otimes \mathbf{I}_n \quad \text{mit} \quad \mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}_j' \quad \text{für} \quad i = j \\ &\quad \text{und} \quad \mathbf{E}_{ij} = \mathbf{e}_i \mathbf{e}_j' + \mathbf{e}_j \mathbf{e}_i' \quad \text{für} \quad i \neq j. \end{aligned} \quad (373.9)$$

Bei der Schätzung der Varianz- und Kovarianzkomponenten wird $\sigma_i^2 = \sigma_{ij} = 1$ gesetzt, so daß mit (361.9) folgt

$$\Sigma_0 \otimes \mathbf{I}_n = \sum_{l=1}^k \mathbf{V}_l. \quad (373.10)$$

Die Matrix \mathbf{W} in (362.3) ergibt sich dann mit (372.1) zu $\mathbf{W} = (\boldsymbol{\Sigma}_0 \otimes \mathbf{I}_n)^{-1} - (\boldsymbol{\Sigma}_0 \otimes \mathbf{I}_n)^{-1}(\mathbf{I}_p \otimes \mathbf{X})[(\mathbf{I}_p \otimes \mathbf{X})'(\boldsymbol{\Sigma}_0 \otimes \mathbf{I}_n)^{-1}(\mathbf{I}_p \otimes \mathbf{X})]^{-1}(\mathbf{I}_p \otimes \mathbf{X})'(\boldsymbol{\Sigma}_0 \otimes \mathbf{I}_n)^{-1}$ oder mit (131.23) bis (131.27) und mit (372.2)

$$\mathbf{W} = \boldsymbol{\Sigma}_0^{-1} \otimes \mathbf{R} \quad \text{mit} \quad \mathbf{R} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' . \quad (373.11)$$

Hiermit folgt $\mathbf{WV}_l\mathbf{W} = \alpha_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1} \otimes \mathbf{R}$ und damit die l -te Komponente q_l des Vektors \mathbf{q} in (362.5)

$$q_l = (\text{vec}\mathbf{Y})'(\alpha_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1} \otimes \mathbf{R})\text{vec}\mathbf{Y} . \quad (373.12)$$

Weiter erhält man ein Element s_{lm} der Matrix \mathbf{S} in (362.8) zu $\text{sp}(\mathbf{WV}_l\mathbf{WV}_m) = \text{sp}(\alpha_{ij}\alpha_{oq}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{oq} \otimes \mathbf{R})$ oder mit (331.7)

$$\text{sp}(\mathbf{WV}_l\mathbf{WV}_m) = (n - q)\text{sp}(\alpha_{ij}\alpha_{oq}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{oq}) . \quad (373.13)$$

Die Elemente $\hat{\sigma}_{ij}$ der Matrix $\hat{\boldsymbol{\Sigma}}$ ergeben sich mit (373.3), (373.4) und (373.11) aus $\mathbf{y}'_i\mathbf{R}\mathbf{y}_j/(n - q)$. Die Varianz- und Kovarianzkomponenten werden jedoch wegen der in (373.7) eingeführten Näherungswerte α_{ij} aus $\hat{\sigma}_{ij}\alpha_{ij}$ geschätzt, so daß $\hat{\sigma}_{ij} = \mathbf{y}'_i\mathbf{R}\mathbf{y}_j/(\alpha_{ij}(n - q))$ zu setzen ist. Mit der in (373.9) eingeführten Matrix \mathbf{E}_{ij} ergibt sich dann wegen (131.22)

$$\hat{\sigma}_{ij} = \frac{1}{n - q}(\text{vec}\mathbf{Y})' \left(\frac{\mathbf{E}_{ij}}{(2 - \delta_{ij})\alpha_{ij}} \otimes \mathbf{R} \right) \text{vec}\mathbf{Y} \quad \text{mit} \quad \delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j . \end{cases} \quad (373.14)$$

Faßt man die Schätzwerte $\hat{\sigma}_{ij}$ im Vektor $\hat{\boldsymbol{\sigma}}$ zusammen, so soll jetzt gezeigt werden, daß sie die aus (362.8) folgende Gleichung $\mathbf{S}\hat{\boldsymbol{\sigma}} = \mathbf{q}$ erfüllen. Aus der Multiplikation von $\hat{\boldsymbol{\sigma}}$ mit der l -ten Zeile von \mathbf{S} , die durch die Indizes (i, j) in (373.13) bestimmt ist, ergibt sich mit (137.3), (137.8) und (373.12)

$$\begin{aligned} & \sum_{\substack{o=1 \\ o \leq q}}^p \sum_{\substack{q=1 \\ q \leq p}}^p \text{sp}(\alpha_{ij}\alpha_{oq}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{oq}) [(\text{vec}\mathbf{Y})' \left(\frac{\mathbf{E}_{oq}}{(2 - \delta_{oq})\alpha_{oq}} \otimes \mathbf{R} \right) \text{vec}\mathbf{Y}] \\ &= \sum_{o \leq q} \sum_{q \leq p} \text{sp}(\alpha_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{oq}) \text{sp} \left(\mathbf{Y}'\mathbf{R}\mathbf{Y} \frac{\mathbf{E}_{oq}}{2 - \delta_{oq}} \right) \\ &= \sum_{o \leq q} \sum_{q \leq p} \alpha_{ij}(2 - \delta_{oq})(\mathbf{e}'_q\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{e}_o)(\mathbf{e}'_q\mathbf{Y}'\mathbf{R}\mathbf{Y}\mathbf{e}_o) \\ &= \text{sp}(\alpha_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1} \sum_{o=1}^p \sum_{q=1}^p \mathbf{e}_o\mathbf{e}'_q\mathbf{Y}'\mathbf{R}\mathbf{Y}\mathbf{e}_o\mathbf{e}'_q) \\ &= \text{sp}(\alpha_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{Y}'\mathbf{R}\mathbf{Y}) \\ &= (\text{vec}\mathbf{Y})'(\alpha_{ij}\boldsymbol{\Sigma}_0^{-1}\mathbf{E}_{ij}\boldsymbol{\Sigma}_0^{-1} \otimes \mathbf{R})\text{vec}\mathbf{Y} = q_l . \end{aligned}$$

Die Gleichung $\mathbf{S}\hat{\boldsymbol{\sigma}} = \mathbf{q}$ ist unabhängig von den in (373.14) eingeführten Näherungswerten α_{ij} erfüllt und die Gleichung (362.4) mit entsprechenden p_j für alle Werte λ_i , da $\hat{\boldsymbol{\sigma}}$ wegen (373.5) erwartungstreu geschätzt wird, so daß $\hat{\boldsymbol{\Sigma}}$ aus (373.3) eine gleichmäßig beste invariante quadratische erwartungstreue Schätzung darstellt.

Abschließend sollen noch unter der Annahme der Normalverteilung (371.5) für die Beobachtungen $\text{vec}\mathbf{Y}$ die Varianzen der Schätzwerte $\hat{\sigma}_i^2$ und $\hat{\sigma}_{ij}$ berechnet werden. Mit (271.2) und (371.6) folgt für die Varianz von ω_{ii} aus (373.4) $V(\omega_{ii}) = V(\mathbf{y}'_i \mathbf{R} \mathbf{y}_i) = 2(\sigma_i^2)^2 \text{sp} \mathbf{R}^2 + 4\sigma_i^2 \boldsymbol{\beta}'_i \mathbf{X}' \mathbf{R}^2 \mathbf{X} \boldsymbol{\beta}_i$ und wegen (331.7) $V(\omega_{ii}) = 2(\sigma_i^2)^2(n - q)$. Mit (233.2) und (373.3) ergibt sich dann

$$V(\hat{\sigma}_i^2) = 2(\sigma_i^2)^2 / (n - q). \quad (373.15)$$

Für die mit der geschätzten Varianz $\hat{\sigma}_i^2$ berechnete Varianz $\hat{V}(\hat{\sigma}_i^2)$ gilt

$$\hat{V}(\hat{\sigma}_i^2) = 2(\hat{\sigma}_i^2)^2 / (n - q). \quad (373.16)$$

Zur Bestimmung der Varianz von $\hat{\sigma}_{ij}$ ist nach (373.4) $V(\omega_{ij}) = V(\mathbf{y}'_i \mathbf{R} \mathbf{y}_j)$ zu berechnen. Mit (373.5) folgt wegen (232.5) $V(\omega_{ij}) = E[(\omega_{ij} - \sigma_{ij}(n - q))^2]$ und daher

$$V(\omega_{ij}) = E(\omega_{ij}^2) - \sigma_{ij}^2(n - q)^2. \quad (373.17)$$

Es muß jetzt $E(\omega_{ij}^2)$ bestimmt werden, wozu die Vektoren

$$\mathbf{e} = |\mathbf{e}'_i, \mathbf{e}'_j|' \quad \text{mit} \quad \mathbf{e}_i = \mathbf{X}\boldsymbol{\beta}_i - \mathbf{y}_i \quad \text{und} \quad \mathbf{e}_j = \mathbf{X}\boldsymbol{\beta}_j - \mathbf{y}_j \quad (373.18)$$

definiert werden. Dann gilt $E(\omega_{ij}^2) = E[(\mathbf{X}\boldsymbol{\beta}_i - \mathbf{e}_i)' \mathbf{R} (\mathbf{X}\boldsymbol{\beta}_j - \mathbf{e}_j)]^2 = E(\mathbf{e}'_i \mathbf{R} \mathbf{e}_j \mathbf{e}'_i \mathbf{R} \mathbf{e}_j)$ wegen (153.5). Mit $\mathbf{R} = (r_{kl})$, $\mathbf{e}_i = (e_{ki})$ und $\mathbf{e}_j = (e_{lj})$ ergibt sich hierfür

$$E(\omega_{ij}^2) = \sum_{k=1}^n \sum_{l=1}^n \sum_{o=1}^n \sum_{p=1}^n r_{kl} r_{op} E(e_{ki} e_{lj} e_{oi} e_{pj}). \quad (373.19)$$

Den Erwartungswert $E(e_{ki} e_{lj} e_{oi} e_{pj})$ erhält man wie im Beispiel des Kapitels 252. Wegen (371.5) folgt mit (253.1) für die Verteilung des Vektors $|\mathbf{y}'_i, \mathbf{y}'_j|'$

$$\begin{vmatrix} \mathbf{y}_i \\ \mathbf{y}_j \end{vmatrix} \sim N \left(\begin{vmatrix} \mathbf{X}\boldsymbol{\beta}_i \\ \mathbf{X}\boldsymbol{\beta}_j \end{vmatrix}, \begin{vmatrix} \sigma_i^2 \mathbf{V} & \sigma_{ij} \mathbf{V} \\ \sigma_{ji} \mathbf{V} & \sigma_j^2 \mathbf{V} \end{vmatrix} \right),$$

in der zur Vereinfachung der folgenden Ableitung anstelle der Einheitsmatrix \mathbf{I}_n die $n \times n$ positiv definite Matrix \mathbf{V} eingeführt wurde. Für den Vektor \mathbf{e} in (373.18) gilt dann wegen (255.1)

$$\mathbf{e} = \begin{vmatrix} \mathbf{e}_i \\ \mathbf{e}_j \end{vmatrix} \sim N \left(\begin{vmatrix} \mathbf{0} \\ \mathbf{0} \end{vmatrix}, \begin{vmatrix} \sigma_i^2 \mathbf{V} & \sigma_{ij} \mathbf{V} \\ \sigma_{ji} \mathbf{V} & \sigma_j^2 \mathbf{V} \end{vmatrix} \right),$$

so daß mit den beiden $n \times 1$ Vektoren \mathbf{t}_i und \mathbf{t}_j und mit $\mathbf{t} = |\mathbf{t}'_i, \mathbf{t}'_j|'$ die momenterzeugende Funktion von \mathbf{e} sich aus (252.1) ergibt zu

$$\begin{aligned} M_{\mathbf{e}}(\mathbf{t}) &= \exp\left(\frac{1}{2}|\mathbf{t}'_i, \mathbf{t}'_j| \begin{vmatrix} \sigma_i^2 \mathbf{V} & \sigma_{ij} \mathbf{V} \\ \sigma_{ji} \mathbf{V} & \sigma_j^2 \mathbf{V} \end{vmatrix} \begin{vmatrix} \mathbf{t}_i \\ \mathbf{t}_j \end{vmatrix}\right) \\ &= \exp\left(\frac{1}{2}(\sigma_i^2 \mathbf{t}'_i \mathbf{V} \mathbf{t}_i + 2\sigma_{ij} \mathbf{t}'_i \mathbf{V} \mathbf{t}_j + \sigma_j^2 \mathbf{t}'_j \mathbf{V} \mathbf{t}_j)\right). \end{aligned}$$

Mit $\mathbf{t}_i = (t_{ki})$, $\mathbf{t}_j = (t_{lj})$ und $\mathbf{V} = (v_{ij})$ folgt durch Differentiation von $M_{\mathbf{e}}(\mathbf{t})$ nach t_{ki} , t_{lj} , t_{oi} und t_{pj} entsprechend der Herleitung für (252.8)

$$E(e_{ki}e_{lj}e_{oi}e_{pj}) = \sigma_{ij}^2 v_{kl}v_{op} + \sigma_i^2 \sigma_j^2 v_{ko}v_{lp} + \sigma_{ij}^2 v_{lo}v_{kp}. \quad (373.20)$$

Dieses Ergebnis in (373.19) eingesetzt ergibt wie im Beweis zu (271.2)

$$E(\omega_{ij}^2) = \sigma_{ij}^2 \text{sp}(\mathbf{R}\mathbf{V})\text{sp}(\mathbf{R}\mathbf{V}) + \sigma_i^2 \sigma_j^2 \text{sp}(\mathbf{R}\mathbf{V}\mathbf{R}\mathbf{V}) + \sigma_{ij}^2 \text{sp}(\mathbf{R}\mathbf{V}\mathbf{R}\mathbf{V})$$

und mit $\mathbf{V} = \mathbf{I}_n$ sowie (331.7) $E(\omega_{ij}^2) = \sigma_{ij}^2(n-q)^2 + (\sigma_i^2 \sigma_j^2 + \sigma_{ij}^2)(n-q)$. Anstelle von (373.17) folgt dann

$$V(\omega_{ij}) = (\sigma_i^2 \sigma_j^2 + \sigma_{ij}^2)(n-q). \quad (373.21)$$

Mit $\hat{\sigma}_{ij} = \omega_{ij}/(n-q)$ wegen (373.3) erhält man schließlich mit (233.2)

$$V(\hat{\sigma}_{ij}) = (\sigma_i^2 \sigma_j^2 + \sigma_{ij}^2)/(n-q) \quad (373.22)$$

oder die mit den Schätzwerten der Varianzen und Kovarianzen berechnete Varianz $\hat{V}(\hat{\sigma}_{ij})$

$$\hat{V}(\hat{\sigma}_{ij}) = (\hat{\sigma}_i^2 \hat{\sigma}_j^2 + \hat{\sigma}_{ij}^2)/(n-q). \quad (373.23)$$

Zusammenfassend gilt der

Satz: Unter der Voraussetzung normalverteilter Beobachtungen sind die gleichmäßig beste invariante quadratische erwartungstreue Schätzung $\hat{\sigma}_i^2$ und $\hat{\sigma}_{ij}$ der Varianzen σ_i^2 und Kovarianzen σ_{ij} im multivariaten Modell (371.1) durch (373.3) und ihre mit $\hat{\sigma}_i^2$ und $\hat{\sigma}_{ij}$ berechneten Varianzen $\hat{V}(\hat{\sigma}_i^2)$ und $\hat{V}(\hat{\sigma}_{ij})$ durch (373.16) und (373.23) gegeben. (373.24)

374 Numerische Berechnung der Schätzwerte und unvollständige multivariate Modelle

a) Numerische Berechnung der Schätzwerte

Für Modelle mit vollem Rang braucht zur Berechnung der Schätzwerte $\hat{\mathbf{B}}$ der Parametermatrix \mathbf{B} nach (372.5) die Normalgleichungsmatrix $\mathbf{X}'\mathbf{X}$ nur einmal reduziert zu werden, falls bei der Gaußschen Elimination (133.16) mit anschließender Rückrechnung (133.17) anstelle eines Vektors die Matrix $\mathbf{X}'\mathbf{Y}$ an die Normalgleichungsmatrix angehängt wird. Erweitert man die Normalgleichungsmatrix gleichzeitig um die Matrix $\mathbf{Y}'\mathbf{Y}$, ergibt sich (326.3) entsprechend mit der Gaußschen Elimination die Matrix $\mathbf{\Omega}$ der quadratischen und bilinearen Formen der Residuen, aus der nach (373.3) $\hat{\mathbf{\Sigma}}$ folgt. Man erhält mit (134.8)

$$\left| \begin{array}{cc|cc} \mathbf{I} & \mathbf{0} & \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Y} \\ -\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{I} & \mathbf{Y}'\mathbf{X} & \mathbf{Y}'\mathbf{Y} \end{array} \right| = \left| \begin{array}{cc|c} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Y} & \\ \mathbf{0} & \mathbf{\Omega} & \end{array} \right|, \quad (374.1)$$

denn es gilt $\mathbf{\Omega} = \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ wegen (373.4). Die Matrix $\hat{\mathbf{B}}$ und die zur Berechnung der Kovarianzmatrizen $D(\hat{\beta}_i)$ und $D(\text{vec}\hat{\mathbf{B}})$ benötigte inverse Normalgleichungsmatrix $(\mathbf{X}'\mathbf{X})^{-1}$ lassen sich (326.4) entsprechend ebenfalls durch die Gaußsche Elimination gewinnen

$$\left| \begin{array}{ccc|ccc} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Y} & \mathbf{I} \\ -\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{I} & \mathbf{0} & \mathbf{Y}'\mathbf{X} & \mathbf{Y}'\mathbf{Y} & \mathbf{0} \\ -(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{0} & \mathbf{I} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{array} \right| = \left| \begin{array}{ccc|cc} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Y} & \mathbf{I} & & \\ \mathbf{0} & \mathbf{\Omega} & -\hat{\mathbf{B}}' & & \\ \mathbf{0} & -\hat{\mathbf{B}} & -(\mathbf{X}'\mathbf{X})^{-1} & & \end{array} \right|. \quad (374.2)$$

Für Modelle mit nicht vollem Rang kann die Normalgleichungsmatrix entsprechend (333.7) oder (333.8) erweitert und (374.1) und (374.2) angewendet werden, sofern man die Zeilen und Spalten derart anordnet, daß keine Nullen auf der Diagonalen bei der Gaußschen Elimination auftreten, wie bereits bei (133.11) erläutert wurde. Das folgende Beispiel wird nach dieser Methode berechnet. Soll das Umordnen der Zeilen und Spalten unterbleiben, ist mit (155.17), (155.19), (155.23) oder (155.25) zu arbeiten.

Beispiel: Um die Effekte einer ungünstigen und günstigen Geschäftslage auf die Preisgestaltung von Waren zu untersuchen, werden in drei Geschäften in ungünstiger Lage und in drei Geschäften in günstiger Lage die Preise von in den sechs Geschäften jeweils nahezu identischen Waren 1 und Waren 2 ermittelt. Die Waren 1 und Waren 2 seien repräsentativ für das allgemeine Warenangebot der Geschäfte. Damit liegt eine multivariate Einwegklassifikation vor. Der Faktor A , der die Geschäftslage repräsentiert, kommt in den beiden Stufen ungünstig und günstig vor. Die Preise der Waren 1 entsprechen

dem 1. Merkmal und die der Waren 2 dem 2. Merkmal. Man erhält mit (342.1)

$$\bar{\mathbf{X}} = \begin{vmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix}, \mathbf{B} = \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \mu_1 & \mu_2 \end{vmatrix}, \bar{\mathbf{y}}_1 = \begin{vmatrix} 1,40 \\ 1,50 \\ 1,55 \\ 2,60 \\ 2,40 \\ 2,50 \end{vmatrix}, \bar{\mathbf{y}}_2 = \begin{vmatrix} 1,50 \\ 1,60 \\ 1,45 \\ 2,60 \\ 2,30 \\ 2,45 \end{vmatrix},$$

worin $\bar{\mathbf{X}}$ die Versuchsplanmatrix, \mathbf{B} die Matrix der unbekannt Parameter $\alpha_{1i}, \alpha_{2i}, \mu_i$ mit $i \in \{1, 2\}$, $\bar{\mathbf{y}}_1$ die Preise der Waren 1 und $\bar{\mathbf{y}}_2$ die der Waren 2 in DM angeben. Die Streuung der Preise in der Dimension DM^2 infolge geringfügiger Abweichungen innerhalb der Waren 1 und der Waren 2 wird durch die folgende Kovarianzmatrix ausgedrückt

$$C(\bar{\mathbf{y}}_i, \bar{\mathbf{y}}_j) = \sigma_{ij} \begin{vmatrix} 0,01 & 0 & 0 & 0 & 0 & 0 \\ & 0,01 & 0 & 0 & 0 & 0 \\ & & 0,01 & 0 & 0 & 0 \\ & & & 0,04 & 0 & 0 \\ & & & & 0,04 & 0 \\ & & & & & 0,04 \end{vmatrix}$$

für $i, j \in \{1, 2\}$.

Für die Parameterschätzung liegt damit das Modell (371.2) vor, das durch die Transformation (321.4) auf das Modell (371.1) zurückgeführt wird, wobei \mathbf{G}' in (321.4) und damit \mathbf{X} , \mathbf{y}_1 und \mathbf{y}_2 sich ergeben zu

$$\mathbf{G}' = \begin{vmatrix} 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{vmatrix}, \mathbf{X} = \begin{vmatrix} 10 & 0 & 10 \\ 10 & 0 & 10 \\ 10 & 0 & 10 \\ 0 & 5 & 5 \\ 0 & 5 & 5 \\ 0 & 5 & 5 \end{vmatrix},$$

$$\mathbf{y}_1 = \begin{vmatrix} 14,0 \\ 15,0 \\ 15,5 \\ 13,0 \\ 12,0 \\ 12,5 \end{vmatrix}, \mathbf{y}_2 = \begin{vmatrix} 15,0 \\ 16,0 \\ 14,5 \\ 13,0 \\ 11,5 \\ 12,25 \end{vmatrix}.$$

Da die Versuchsplanmatrix \mathbf{X} keinen vollen Spaltenrang besitzt, wird die Parameterschätzung, wie im Kapitel 343 erläutert, mittels der Restriktionen (343.5) vorgenommen, die in der Matrix \mathbf{B}^* zusammengefaßt seien. Aus

(343.5) folgt $\alpha_{1i} + \alpha_{2i} = 0$ für $i \in \{1, 2\}$, so daß man erhält

$$\mathbf{B}^* = |1, 1, 0|.$$

Die aus (333.7) sich ergebenden Normalgleichungen sollen nach (374.2) gelöst werden, so daß zur Vermeidung von Nullen auf der Diagonalen bei der Gaußschen Elimination der Multiplikator k_i , der nach (333.9) für die Restriktion $\alpha_{1i} + \alpha_{2i} = 0$ erscheint, im Anschluß an die Parameter α_{1i} und α_{2i} in den Normalgleichungen aufgeführt wird. Man erhält

α_{1i}	α_{2i}	k_i	μ_i	1	2	\mathbf{I}
300	0	1	300	445	455	1 0 0 0
	75	1	75	187,5	183,75	0 1 0 0
		0	0	0	0	0 0 1 0
			375	632,5	638,75	0 0 0 1
				1130,5	1134,875	0 0 0 0
					1142,562	0 0 0 0
						0 0 0 0
						0 0 0 0
						0 0
						0

Nach vier Eliminationsschritten, die wie in (326.7) ablaufen, ergibt sich

1	2	\mathbf{I}			
1,667	0,583	0,5083	-0,5083	0,0	-1,9917
	2,291	0,4667	-0,4667	0,0	-1,9833
		-0,00417	0,00417	-0,5	0,00250
			-0,00417	-0,5	-0,00250
				0,0	0,50000
					-0,00417

Nach (374.2) folgt dann mit (373.3) und $n - q = 4$

$$\hat{\mathbf{B}} = \begin{vmatrix} -0,51 & -0,47 \\ 0,51 & 0,47 \\ 1,99 & 1,98 \end{vmatrix}, \hat{\mathbf{\Omega}} = \begin{vmatrix} 1,667 & 0,583 \\ 2,291 & \end{vmatrix}, \hat{\mathbf{\Sigma}} = \frac{1}{4} \hat{\mathbf{\Omega}}$$

und daher ergeben sich

$$\hat{\sigma}_1^2 = 0,417; \hat{\sigma}_{12} = 0,146; \hat{\sigma}_2^2 = 0,573$$

sowie mit (373.16) und (373.23) die Standardabweichungen

$$[\hat{V}(\hat{\sigma}_1^2)]^{1/2} = 0,29; [\hat{V}(\hat{\sigma}_{12})]^{1/2} = 0,26; [\hat{V}(\hat{\sigma}_2^2)]^{1/2} = 0,41.$$

Die Effekte der ungünstigen oder günstigen Lage bewirken also bei den Waren 1 eine Erniedrigung oder Erhöhung um DM 0,51 des Mittelwertes von DM 1,99 und bei den Waren 2 um DM 0,47 des Mittelwertes von DM 1,98. Weiter ergibt sich nach (374.2) wegen (333.4)

$$(\mathbf{X}'\mathbf{X})_{rs}^- = \begin{vmatrix} 0,00417 & -0,00417 & -0,00250 \\ & 0,00417 & 0,00250 \\ & & 0,00417 \end{vmatrix},$$

so daß (325.7) entsprechend die Kovarianzmatrizen $\hat{D}(\hat{\beta}_i) = \hat{\sigma}_i^2(\mathbf{X}'\mathbf{X})_{rs}^-$ und mit (372.4) $\hat{C}(\hat{\beta}_i, \hat{\beta}_j) = \hat{\sigma}_{ij}(\mathbf{X}'\mathbf{X})_{rs}^-$ angebbbar sind.

b) Unvollständige multivariate Modelle

Im multivariaten Modell (371.1) ist die Koeffizientenmatrix \mathbf{X} für die p Beobachtungsvektoren \mathbf{y}_i identisch, wobei die Vektoren \mathbf{y}_i vollständig gegeben sein müssen. Sind zum Beispiel durch Änderungen im Beobachtungsprogramm bei Wiederholungsmessungen oder durch den Ausfall von Messungen diese Voraussetzungen nicht erfüllt, liegt ein *unvollständiges multivariates Modell* vor. Auch dann wird man in manchen Fällen durch eine Transformation der ursprünglichen Beobachtungen in abgeleitete Beobachtungen ein Modell (371.1) erhalten können (SRIVASTAVA 1966). Unterscheiden sich sämtliche Koeffizientenmatrizen voneinander, besitzen aber die p Beobachtungsvektoren \mathbf{y}_i gleiche Dimensionen, läßt sich die Kovarianzmatrix Σ zusammen mit den Parametervektoren β_i iterativ schätzen (PRESS 1972, S.219).

Allgemein sei ein unvollständiges multivariates Modell derart gegeben, daß weder die Dimensionen der Koeffizientenmatrizen noch die der Parameter- und Beobachtungsvektoren übereinstimmen, also

$$\mathbf{X}_i\beta_i = E(\mathbf{y}_i) \quad \text{für } i \in \{1, \dots, p\}, \quad (374.3)$$

worin \mathbf{X}_i die $n_i \times u_i$ Matrizen gegebener Koeffizienten, β_i die $u_i \times 1$ Vektoren fester, unbekannter Parameter und \mathbf{y}_i die $n_i \times 1$ Zufallsvektoren der Beobachtungen bedeuten. Es seien k_{ij} Komponenten der Vektoren \mathbf{y}_i und \mathbf{y}_j für $i \neq j$ miteinander korreliert. Werden diese Komponenten durch Umordnen an den Anfang gestellt, so daß die Vektoren $\bar{\mathbf{y}}_i = |\bar{\mathbf{y}}'_{i1}, \bar{\mathbf{y}}'_{i2}|'$ und $\bar{\mathbf{y}}_j = |\bar{\mathbf{y}}'_{j1}, \bar{\mathbf{y}}'_{j2}|'$ erhalten werden, in denen $\bar{\mathbf{y}}_{i1}$ und $\bar{\mathbf{y}}_{j1}$ die korrelierten $k_{ij} \times 1$ Vektoren bedeuten, dann sei die Kovarianzmatrix $C(\bar{\mathbf{y}}_i, \bar{\mathbf{y}}_j)$ dem multivariaten Modell (371.1) entsprechend gegeben durch

$$C(\bar{\mathbf{y}}_i, \bar{\mathbf{y}}_j) = C\left(\begin{vmatrix} \bar{\mathbf{y}}_{i1} \\ \bar{\mathbf{y}}_{i2} \end{vmatrix}, \begin{vmatrix} \bar{\mathbf{y}}_{j1} \\ \bar{\mathbf{y}}_{j2} \end{vmatrix}\right) = \begin{vmatrix} \sigma_{ij}\mathbf{I}_{k_{ij}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{vmatrix}. \quad (374.4)$$

Für die Kovarianzmatrix $D(\mathbf{y}_i)$ gelte wie in (371.1)

$$D(\mathbf{y}_i) = \sigma_i^2\mathbf{I}_{n_i}. \quad (374.5)$$

Anstelle der Einheitsmatrizen in (374.4) und (374.5) lassen sich dem Modell (371.2) entsprechend auch positiv definite Matrizen einführen.

Das multivariate Modell (374.3) wird in das univariate Modell umgeschrieben

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad \mathbf{X} = \text{diag}(\mathbf{X}_1, \dots, \mathbf{X}_p), \\ \boldsymbol{\beta} = |\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_p|', \quad \mathbf{y} = |\mathbf{y}'_1, \dots, \mathbf{y}'_p|' \quad (374.6)$$

mit der Kovarianzmatrix $D(\mathbf{y})$, die nach (374.4) und (374.5) aufzubauen ist. In diesem Modell sind neben den Parametern $\boldsymbol{\beta}$ noch die Varianzen σ_i^2 und die Kovarianzen σ_{ij} unbekannt, die als Varianz- und Kovarianzkomponenten zu interpretieren sind, so daß das Modell (361.3) vorliegt. Damit ist das unvollständige multivariate Modell (374.3) bis (374.5) auf das Gauß-Markoff-Modell (361.3) mit unbekanntem Varianz- und Kovarianzkomponenten zurückgeführt worden. Eine Verallgemeinerung des Modells (374.3) bis (374.5) befindet sich in (SCHAFFRIN 1983, S.101).

375 Spezielles Modell zur Schätzung von Kovarianzmatrizen und Schätzung von Kovarianzen für stochastische Prozesse

Soll das multivariate Modell lediglich dazu dienen, die Kovarianzmatrix $\boldsymbol{\Sigma}$ in (371.4) der Merkmalsvektoren \mathbf{z}_k zu schätzen, wählt man zweckmäßig das folgende spezielle Modell. Die Parametervektoren $\boldsymbol{\beta}_i$ mit $i \in \{1, \dots, p\}$ besitzen jeweils nur eine Komponente, so daß $\boldsymbol{\beta}_i = \beta_i$ folgt und die Matrix \mathbf{X} mit $\mathbf{X} = |1, \dots, 1|'$ zu einem $n \times 1$ Vektor wird, also mit $\mathbf{y}_i = (y_{ki})$ für $k \in \{1, \dots, n\}$

$$\begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} \begin{pmatrix} | \\ | \\ \dots \\ | \end{pmatrix} |\beta_1, \dots, \beta_p| = E \left(\begin{pmatrix} y_{11} & y_{12} & \dots & y_{1p} \\ y_{21} & y_{22} & \dots & y_{2p} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{np} \end{pmatrix} \right).$$

Die Schätzwerte $\hat{\beta}_i$ von β_i ergeben sich dann aus (372.5) als Mittelwerte der Beobachtungsvektoren \mathbf{y}_i zu

$$\hat{\beta}_i = \frac{1}{n} \sum_{k=1}^n y_{ki} \quad \text{für} \quad i \in \{1, \dots, p\} \quad (375.1)$$

und die Schätzung $\hat{\boldsymbol{\Sigma}}$ der Kovarianzmatrix $\boldsymbol{\Sigma}$ mit $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ij})$ aus (373.2) zu

$$\hat{\sigma}_{ij} = \frac{1}{n-1} \sum_{k=1}^n (\hat{\beta}_i - y_{ki})(\hat{\beta}_j - y_{kj}) \quad \text{für} \quad i, j \in \{1, \dots, p\}. \quad (375.2)$$

Bei diesem speziellen Modell enthalten die Beobachtungsvektoren \mathbf{y}_i insgesamt n Beobachtungen einer Messungsgröße oder eines Merkmals. Diese Beobachtungen müssen unkorreliert sein, oder alle Vektoren \mathbf{y}_i müssen nach (371.2) die gleiche Gewichtsmatrix besitzen. Die Vektoren \mathbf{y}_i und \mathbf{y}_j mit $i \neq j$ enthalten jeweils n Beobachtungen verschiedener Merkmale. Die Kovarianzen σ_{ij} ihrer Kovarianzmatrix $C(\mathbf{y}_i, \mathbf{y}_j) = \sigma_{ij} \mathbf{I}_n$ aus (371.1) werden mit (375.2) geschätzt. Ob Kovarianzen verschiedener Merkmale erwartet werden können, richtet sich nach dem Experiment oder dem physikalischen Meßvorgang. Häufig sind beispielsweise Entfernungen, die in kurzen Zeitabständen elektro-optisch gemessen werden, miteinander korreliert. Werden p Strecken nacheinander bestimmt, bilden sie die Beobachtungen y_{ki} mit $i \in \{1, \dots, p\}$. Insgesamt n Beobachtungsserien y_{ki} mit $k \in \{1, \dots, n\}$ sind durchzuführen, um die Kovarianzen nach (375.2) zu schätzen.

Für (356.7) und (356.15) wurden bereits Zufallsbeobachtungen erläutert, die Funktionen der Zeit beziehungsweise eines Ortes sind. Sie stellen somit Realisierungen eines eindimensionalen beziehungsweise eines mehrdimensionalen *stochastischen Prozesses* oder *Zufallsprozesses* dar. Liegen nicht Beobachtungen sondern Beobachtungsvektoren vor, erhält man *vektorielle Zufallsprozesse* (KOCH und SCHMIDT 1994, S.162; MEIER und KELLER 1990, S.85). Zufallsprozesse bezeichnet man als *stationär* im Falle der Zeitabhängigkeit und *homogen* im Falle der Ortsabhängigkeit, wenn die Verteilungsfunktionen der Zufallsvariablen nicht von der Zeit oder vom Ort abhängen. Im Falle der Zeitabhängigkeit sind dann die Kovarianzen der Beobachtungen lediglich Funktionen der Zeitdifferenzen beziehungsweise Funktionen der Differenzen von Ortsvektoren im Falle der Ortsabhängigkeit. Ist ein von einem Ortsvektor abhängiger Prozeß homogen und *isotrop*, das heißt richtungsunabhängig (JAGLOM 1959, S.66; KOCH und SCHMIDT 1994, S.251), so sind die Kovarianzen der Beobachtungen lediglich Funktionen der Längen der Ortsvektordifferenzen. Bei vektoriellen homogenen isotropen Prozessen, die von einem Ortsvektor abhängen, ergibt sich die *Taylor-Karman-Struktur* der Kovarianzmatrix (GRAFAREND und SCHAFFRIN 1979; MONIN und YAGLOM 1975, S.39). Ist schließlich der stochastische Prozeß noch *ergodisch* (KOCH und SCHMIDT 1994, S.209; PAPOULIS 1965, S.327), lassen sich die Werte seiner Realisierung als Wiederholungsmessungen oder als Beobachtungen unterschiedlicher Merkmale interpretieren, so daß die Kovarianzen der Beobachtungen nach (375.2) geschätzt werden können.

Im Falle eines von der Zeit abhängigen Prozesses $y(t_1), y(t_2), \dots, y(t_p)$ entspricht dann der Zeitpunkt t_i dem Index i der Beobachtung y_{ki} in (375.2), der Zeitpunkt t_j dem Index j von y_{kj} , und der Index k in y_{ki} und y_{kj} bezeichnet die Beobachtungspaare, über die nach (375.2) zu summieren ist und deren Zeitdifferenz durch $t_i - t_j = l\Delta t$ gegeben ist, wobei $l \in \{0, 1, 2, \dots, p-1\}$ und

Δt die Zeitdifferenz bedeutet, mit der die Beobachtungen bestimmt wurden. Liegen die Beobachtungen nicht in gleichen Zeitabständen vor, läßt sich die Zuordnung zu gleichen Zeitdifferenzen $t_i - t_j$ nur näherungsweise vornehmen. Die Schätzungen $\hat{\sigma}_{ij}$ ergeben sich nun als Funktionen $\hat{\sigma}(l\Delta t)$ der Zeitdifferenz $l\Delta t$ zu

$$\hat{\sigma}_{ij} = \hat{\sigma}(l\Delta t) = \frac{1}{p-1} \sum_{k=1}^{p-l} (\hat{\beta} - y(t_k))(\hat{\beta} - y(t_k + l\Delta t)) \quad (375.3)$$

mit

$$\hat{\beta} = \frac{1}{p} \sum_{k=1}^p y(t_k), \quad \hat{\sigma}_{ii} = \hat{\sigma}_{jj} = \hat{\sigma}(0), \quad l \in \{0, 1, \dots, p-1\}.$$

Mit wachsendem l nimmt die Anzahl der zu bildenden Beobachtungspaare ab, was in der Summation in (375.3) berücksichtigt wird, nicht dagegen im Faktor $1/(p-1)$. Dies widerspricht der Schätzung nach (375.2), in der die Anzahl n der Produkte, über die summiert wird, konstant ist, so daß (375.3) eine nicht erwartungstreue Schätzung darstellt. Auf die Vorzüge dieser Schätzung gegenüber einer erwartungstreuen Schätzung wird in (JENKINS und WATTS 1968, S.174; KOCH und SCHMIDT 1994, S.218) eingegangen.

Bei einem homogenen isotropen Prozeß, der von einem Ortsvektor abhängt, beispielsweise bei einem durch die Grauwerte einer Photographie definierten Prozeß, erhält man im Falle der Ergodizität die Schätzungen der Kovarianzen nach (375.3) als Funktion $\hat{\sigma}(l\Delta s)$ des Abstandes Δs zwischen zwei Beobachtungspunkten.

Die nach (375.3) an diskreten Stellen geschätzten Kovarianzen lassen sich durch eine analytische Funktion, die *Kovarianzfunktion*, annähern, beispielsweise durch (HIRVONEN 1962; KOCH 1973; MEIER und KELLER 1990, S.123)

$$\hat{\sigma}(f\Delta t) = \frac{\hat{\sigma}(0)}{1 + (f\Delta t/a)^2} \quad \text{mit } f \in \mathbb{R} \quad \text{und } 0 \leq f < \infty, \quad (375.4)$$

worin a eine Konstante bedeutet, durch deren Variation die Anpassung an (375.3) erfolgt. Die Größe $f_a = a/\Delta t$ heißt Halbwertsbreite der Kovarianzfunktion, da $\hat{\sigma}(f_a\Delta t) = \hat{\sigma}(0)/2$ gilt. Mit $f\Delta s$ in (375.4) erhält man die entsprechende Funktion für einen ergodischen Prozeß, der von einem Ortsvektor abhängt.

Aus (375.4) lassen sich nun für beliebige Zeitdifferenzen $f\Delta t$ oder Entfernungsdifferenzen $f\Delta s$ die Kovarianzen berechnen, die für die im Kapitel 356 behandelte Prädiktion benötigt werden. Die mit (375.4) erzeugten Kovarianzmatrizen sind positiv definit, denn $\hat{\sigma}(f\Delta t)$ ist eine positiv definite Funktion

(GNEDENKO 1957, S.212), da (375.4) auf einen Spezialfall der charakteristischen Funktion der Laplace-Verteilung zurückgeführt werden kann (BÄHR und RICHTER 1975; WENTZEL und OWTSCHAROW 1975, S.243). Ein Hypothesentest für die Annäherung von (375.3) durch (375.4) wird im Kapitel 426 behandelt.

376 Multivariates Modell mit Restriktionen

Wie im univariaten Modell mit (327.1) oder mit (334.1) und (334.2) erhält man für das multivariate Modell die

Definition: Gelten im multivariaten Gauß-Markoff-Modell (371.1) für die Parametervektoren β_i zusätzlich die Restriktionen $\mathbf{H}\beta_i = \mathbf{w}_i$, wobei \mathbf{H} eine $r \times u$ Matrix bekannter Koeffizienten mit $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{H}$ sowie $\text{rg}\mathbf{H} = r \leq q$ und \mathbf{w}_i bekannte $r \times 1$ Vektoren bedeuten, so bezeichnet man

$$\mathbf{X}\beta_i = E(\mathbf{y}_i) \quad \text{mit} \quad \mathbf{H}\beta_i = \mathbf{w}_i \quad \text{und} \quad C(\mathbf{y}_i, \mathbf{y}_j) = \sigma_{ij}\mathbf{I}_n \\ \text{für } i, j \in \{1, \dots, p\}$$

als multivariates Gauß-Markoff-Modell mit *Restriktionen*. (376.1)

Faßt man die p Vektoren \mathbf{w}_i in der $r \times p$ Matrix \mathbf{W} zusammen, erhält man (371.3) entsprechend die folgende Formulierung des Modells (376.1)

$$\mathbf{X}\mathbf{B} = E(\mathbf{Y}) \quad \text{mit} \quad \mathbf{H}\mathbf{B} = \mathbf{W} \quad \text{und} \quad D(\text{vec}\mathbf{Y}) = \boldsymbol{\Sigma} \otimes \mathbf{I}_n. \quad (376.2)$$

Wie schon im multivariaten Modell (371.1) unterscheidet sich die Schätzung der Parametervektoren β_i in (376.1) nicht von der Schätzung im univariaten Modell (327.1), (327.2) oder (334.1), so daß die Sätze (327.8), (327.13), (334.3) und (334.16) gelten. Um das zu zeigen, wird wie in (372.1) das multivariate Modell (376.2) in das folgende univariate Modell umgeschrieben

$$(\mathbf{I}_p \otimes \mathbf{X})\text{vec}\mathbf{B} = E(\text{vec}\mathbf{Y}) \quad \text{mit} \quad (\mathbf{I}_p \otimes \mathbf{H})\text{vec}\mathbf{B} = \text{vec}\mathbf{W} \\ \text{und} \quad D(\text{vec}\mathbf{Y}) = \boldsymbol{\Sigma} \otimes \mathbf{I}_n. \quad (376.3)$$

Besitzt \mathbf{X} beliebigen Rang, ergibt sich mit (131.23), (131.26), (131.27) und (372.2) aus (334.4), falls $\text{vec}\tilde{\mathbf{B}}$ die Schätzung von $\text{vec}\mathbf{B}$ bedeutet

$$\text{vec}\tilde{\mathbf{B}} = (\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1})\{(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}')\text{vec}\mathbf{Y} \\ + (\mathbf{I}_p \otimes \mathbf{H}')[(\mathbf{I}_p \otimes \mathbf{H})(\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1})(\mathbf{I}_p \otimes \mathbf{H}')]^{-1} \\ [\text{vec}\mathbf{W} - (\mathbf{I}_p \otimes \mathbf{H})(\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1})(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}')\text{vec}\mathbf{Y}]\}$$

und folglich

$$\begin{aligned} \text{vec}\tilde{\mathbf{B}} &= (\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\text{vec}\mathbf{Y} \\ &+ [\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}]\text{vec}\mathbf{W} \\ &- [\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\text{vec}\mathbf{Y} . \end{aligned} \quad (376.4)$$

Mit $\tilde{\mathbf{B}} = |\tilde{\beta}_1, \dots, \tilde{\beta}_p|$ ergibt sich hieraus

$$\tilde{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'\mathbf{Y} + \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{W} - \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y})] \quad (376.5)$$

in Übereinstimmung mit (334.4). Nach (334.9) bedeutet daher $\tilde{\mathbf{B}}$ die Schätzung nach der Methode der kleinsten Quadrate, die Maximum-Likelihood-Schätzung im Falle normalverteilter Beobachtungen und für Modelle mit vollem Rang mit $\tilde{\mathbf{B}} = \tilde{\mathbf{B}}$ und $\tilde{\mathbf{B}} = |\tilde{\beta}_1, \dots, \tilde{\beta}_p|$ wegen (153.22) und (327.8) die beste lineare erwartungstreue Schätzung $\tilde{\mathbf{B}}$ von \mathbf{B} .

Die $p \times p$ Matrix Ω_H der quadratischen und bilinearen Formen der Residuen ergibt sich (334.8) entsprechend zu

$$\Omega_H = (\mathbf{X}\tilde{\mathbf{B}} - \mathbf{Y})'(\mathbf{X}\tilde{\mathbf{B}} - \mathbf{Y}) \quad (376.6)$$

und (334.11) entsprechend mit (372.5) und (373.3) zu

$$\begin{aligned} \Omega_H &= \Omega + \mathbf{R} \quad \text{mit} \\ \mathbf{R} &= (\mathbf{H}\tilde{\mathbf{B}} - \mathbf{W})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\tilde{\mathbf{B}} - \mathbf{W}) . \end{aligned} \quad (376.7)$$

Diese Beziehung gibt die Änderung der Matrix Ω infolge der Einführung der Restriktionen $\mathbf{H}\mathbf{B} = \mathbf{W}$ an.

Mit Überlegungen, die denen entsprechen, die zu (334.15) und (373.5) führen, ergibt sich als erwartungstreue Schätzung $\tilde{\Sigma}$ der Kovarianzmatrix Σ in (376.2)

$$\tilde{\Sigma} = \frac{1}{n - q + r} \Omega_H . \quad (376.8)$$

Für die Maximum-Likelihood-Schätzung $\bar{\Sigma}$ der Kovarianzmatrix Σ folgt (373.6) entsprechend

$$\bar{\Sigma} = \frac{1}{n} \Omega_H . \quad (376.9)$$

Wie im Zusammenhang mit (412.8) gezeigt wird, ist unter der Voraussetzung normalverteilter Beobachtungen die Wahrscheinlichkeit gleich Null, daß die Matrix Ω_H positiv semidefinit wird.

38 Robuste Parameterschätzung

381 Wahl der Zielfunktion

Wie im Kapitel 311 erwähnt, bezeichnet man eine Schätzfunktion als *robust*, wenn ihre Wahrscheinlichkeitsverteilung gegenüber kleinen Änderungen der Verteilung der Stichprobe unempfindlich ist. Im Kapitel 345 wurde darauf hingewiesen, daß Beobachtungen durch grobe Fehler, also durch *Ausreißer*, verfälscht sein können. Sie bewirken eine Änderung der Verteilung der Beobachtungen. Verhält sich die Schätzfunktion unempfindlich gegenüber dieser Änderung, bezeichnet man sie als *robust*. Unter *robuster Parameterschätzung* versteht man daher häufig eine Schätzung, die gegenüber groben Fehlern in den Beobachtungen unempfindlich ist. In diesem eingeschränkten Sinn soll die robuste Parameterschätzung im folgenden behandelt werden.

Zur Parameterschätzung wurden bislang die im Kapitel 31 vorgestellte beste erwartungstreue Schätzung, die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode angewendet. Sie liefern nach (324.4) im Falle normalverteilter Beobachtungen identische Ergebnisse, so daß sie nach (323.5) die Eigenschaft besitzen, daß die Quadratsumme der Residuen minimal wird. Dies bewirkt eine Verschmierung der Ausreißer, so daß die beste erwartungstreue Schätzung, die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode nicht zu den robusten Schätzverfahren gehören, wie im folgenden noch gezeigt wird. Trotzdem lassen sich einige wenige Ausreißer in sich gegenseitig kontrollierenden Beobachtungen durch diese Schätzungen aufdecken. Hierzu dienen Ausreißertests, auf die im Kapitel 44 eingegangen wird.

Genügen Ausreißertests nicht, sind robuste Parameterschätzungen anzuwenden, deren Entwicklung auf HUBER (1964) zurückgeht. Robuste Verfahren ersetzen in der zu minimierenden Quadratsumme der Residuen die Quadrate oder die Summe durch geeignetere Funktionen. Sie lassen sich, wie im folgenden gezeigt, bis auf die im Kapitel 387 behandelte LMS-Methode als Maximum-Likelihood-Schätzungen, sogenannte M-Schätzungen ableiten. Sie sind daher auch als Bayes-Schätzungen interpretierbar, so daß Hypothesentests vorzunehmen und Konfidenzbereiche anzugeben sind (KOCH 1990, S.144; KOCH und YANG 1998; KOCH 2000, S.196).

Werden durch Ausreißertests oder robuste Parameterschätzungen Beobachtungen identifiziert, die Ausreißer enthalten können, sollte untersucht werden, ob dies tatsächlich der Fall ist. Die Beobachtungen mit Ausreißern werden eliminiert, um dann die Parameter nach den im Kapitel 31 vorgestellten Methoden zu schätzen.

382 Robuste M-Schätzung

Das Modell (331.2) mit nicht vollem Rang und voneinander unabhängigen Beobachtungen wird vorausgesetzt, so daß unter Einführung des Fehlervektors \mathbf{e} gilt

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) = \mathbf{y} + \mathbf{e} \quad \text{mit} \quad \text{rg}\mathbf{X} = q < u \quad \text{und} \quad D(\mathbf{y}) = \sigma^2 \mathbf{I}. \quad (382.1)$$

Die unbekannt Parameter $\boldsymbol{\beta}$ sind nach robusten Verfahren zu schätzen. Besitzen die Beobachtungen y_i mit $\mathbf{y} = (y_i)$ unterschiedliche Gewichte p_i , wird das Modell (382.1) mit

$$\mathbf{G} = \text{diag}(\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_n}) \quad (382.2)$$

durch Homogenisierung des Ausgangsmodells erhalten, wie im 1. Beispiel des Kapitels 321 erläutert. Durch die Homogenisierung werden lediglich die einzelnen Beobachtungen modifiziert, aber keine Linearkombinationen von Beobachtungen gebildet, so daß die Transformation mit (382.2) für die Ausreißersuche unschädlich ist. Vorschläge zur robusten Schätzung abhängiger Beobachtungen findet man zum Beispiel bei (XU 1989; YANG 1994).

Mit $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]'$ und $\mathbf{e} = (e_i)$ erhält man aus (382.1) die Beobachtungsgleichung für die Beobachtung y_i zu

$$\mathbf{x}'_i \boldsymbol{\beta} = E(y_i) = y_i + e_i \quad \text{mit} \quad V(y_i) = \sigma^2, \quad i \in \{1, \dots, n\}. \quad (382.3)$$

Die unbekannt Parameter $\boldsymbol{\beta}$ werden durch die Maximum-Likelihood-Methode (314.2) geschätzt. Die Wahrscheinlichkeitsdichte $p(y_i)$ der Beobachtung y_i sei proportional zu $f(y_i, \boldsymbol{\beta})$, also

$$p(y_i) \propto f(y_i, \boldsymbol{\beta}). \quad (382.4)$$

Die Dichte $p(\mathbf{y})$ des Beobachtungsvektors \mathbf{y} folgt dann wegen der Unabhängigkeit der y_i nach (228.2) zu

$$p(\mathbf{y}) \propto \prod_{i=1}^n f(y_i, \boldsymbol{\beta}). \quad (382.5)$$

Die Likelihoodfunktion $L(\mathbf{y}; \boldsymbol{\beta})$ ergibt sich mit $p(\mathbf{y})$. Führt man, wie im Zusammenhang mit (324.2) erläutert, $\ln L(\mathbf{y}; \boldsymbol{\beta})$ ein, ist für die Maximum-Likelihood-Schätzung der $\boldsymbol{\beta}$

$$\ln \prod_{i=1}^n f(y_i, \boldsymbol{\beta}) \quad (382.6)$$

zu maximieren oder

$$\sum_{i=1}^n -\ln f(y_i, \boldsymbol{\beta}) \quad (382.7)$$

zu minimieren. Falls die Beobachtungen normalverteilt sind, ergeben sich nach (324.4) und (331.4) die Schätzwerte der besten erwartungstreuen Schätzung und der Methode der kleinsten Quadrate.

Zur Ableitung robuster Schätzverfahren ist die zu minimierende Zielfunktion (382.7) durch eine geeignetere Zielfunktion zu ersetzen, die mit $\rho(y_i, g_i(\boldsymbol{\beta}))$ bezeichnet werde und in der $g_i(\boldsymbol{\beta})$ eine Funktion der unbekannt Parameter $\boldsymbol{\beta}$ bezeichne. Nach (HUBER 1981, S.43) erhält man eine Schätzung vom Typ Maximum-Likelihood-Schätzung, abgekürzt *M-Schätzung* genannt, durch Minimierung von

$$\sum_{i=1}^n \rho(y_i, g_i(\boldsymbol{\beta})) . \quad (382.8)$$

Mit der Ableitung $\psi(y_i, \boldsymbol{\beta})$ der Zielfunktion $\rho(y_i, g_i(\boldsymbol{\beta}))$

$$\psi(y_i, \boldsymbol{\beta}) = \frac{\partial}{\partial g_i(\boldsymbol{\beta})} \rho(y_i, g_i(\boldsymbol{\beta})) \quad (382.9)$$

folgt also nach (171.4) die Schätzung $\hat{\boldsymbol{\beta}}$ von $\boldsymbol{\beta}$ mit $\hat{\boldsymbol{\beta}} = (\hat{\beta}_k)$ aus

$$\sum_{i=1}^n \psi(y_i, \hat{\boldsymbol{\beta}}) \frac{\partial g_i(\hat{\boldsymbol{\beta}})}{\partial \hat{\beta}_k} = 0 \quad \text{für } k \in \{1, \dots, u\} . \quad (382.10)$$

Robuste M-Schätzungen werden mit der Funktion $\psi(y_i, \boldsymbol{\beta})$ gefunden, der Ableitung der Zielfunktion nach der Funktion der unbekannt Parameter. Genau dann ergibt sich eine robuste Schätzung, wenn $\psi(y_i, \boldsymbol{\beta})$ beschränkt ist. Die Ableitung ist nämlich proportional zu der von HAMPEL für die Beurteilung robuster Schätzungen eingeführten *Einflußfunktion* (HAMPEL et al. 1986, S.40, 101), die den Einfluß einer zusätzlichen Beobachtung auf die Schätzung ergibt.

Die Beobachtung y_i und die unbekannt Parameter $\boldsymbol{\beta}$ sind durch die Beobachtungsgleichung (382.3) verbunden, so daß $g_i(\hat{\boldsymbol{\beta}}) = \mathbf{x}'_i \hat{\boldsymbol{\beta}}$ gilt. In (382.3) ist außerdem die unbekannt Varianz σ^2 der Gewichtseinheit enthalten. Sie soll im folgenden als bekannt vorausgesetzt und nur die robuste Schätzung von $\boldsymbol{\beta}$ behandelt werden. Zur robusten Schätzung von σ^2 siehe zum Beispiel (HUBER 1981, S.179). Um in der Lösung von (382.10) die als Maßstab wirkende Standardabweichung σ zu berücksichtigen, wird mit

$$x = \frac{\mathbf{x}'_i \boldsymbol{\beta} - y_i}{\sigma} = \frac{e_i}{\sigma} \quad (382.11)$$

der standardisierte Fehler x eingeführt und als Zielfunktion

$$\rho(y_i, g_i(\boldsymbol{\beta})) = \rho(x) \quad (382.12)$$

erhalten. Aus (382.10) folgt dann mit $\mathbf{x}_i = (x_{ik})$

$$\sum_{i=1}^n \psi\left(\frac{\mathbf{x}'_i \hat{\boldsymbol{\beta}} - y_i}{\sigma}\right) \frac{x_{ik}}{\sigma} = 0 \quad \text{für } k \in \{1, \dots, u\}. \quad (382.13)$$

Mit dem Gewicht w_i und dem mit $\hat{e} = (\hat{e}_i)$ sowie (323.4) definierten Residuum \hat{e}_i

$$w_i = \psi\left(\frac{\hat{e}_i}{\sigma}\right) / \left(\frac{\hat{e}_i}{\sigma}\right), \quad (382.14)$$

erhält man weiter

$$\frac{1}{\sigma^2} \sum_{i=1}^n w_i \hat{e}_i x_{ik} = 0 \quad \text{für } k \in \{1, \dots, u\} \quad (382.15)$$

oder mit der Diagonalmatrix \mathbf{W} der Gewichte w_i

$$\mathbf{W} = \text{diag}(w_1, \dots, w_n) \quad (382.16)$$

die u Gleichungen

$$\mathbf{X}' \mathbf{W} \hat{e} = \mathbf{0} \quad (382.17)$$

und schließlich mit (323.4)

$$\mathbf{X}' \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}' \mathbf{W} \mathbf{y}. \quad (382.18)$$

Wegen (382.14) sind die Gewichte w_i abhängig von \hat{e}_i und damit von $\hat{\boldsymbol{\beta}}$. Die Schätzwerte sind daher iterativ zu berechnen. Als erste Näherung $\hat{\boldsymbol{\beta}}^{(0)}$ wird (333.4) gewählt, da projizierte Parameter im Modell mit nicht vollem Rang sich bequem schätzen lassen,

$$\hat{\boldsymbol{\beta}}^{(0)} = (\mathbf{X}' \mathbf{X})_{rs}^- \mathbf{X}' \mathbf{y}. \quad (382.19)$$

Die $(m+1)$ -te Iteration liefert

$$\hat{\boldsymbol{\beta}}^{(m+1)} = (\mathbf{X}' \mathbf{W}^{(m)} \mathbf{X})_{rs}^- \mathbf{X}' \mathbf{W}^{(m)} \mathbf{y} \quad (382.20)$$

mit

$$\mathbf{W}^{(m)} = \text{diag}(w_1^{(m)}, \dots, w_n^{(m)}), \quad w_i^{(m)} = \psi\left(\frac{\hat{e}_i^{(m)}}{\sigma}\right) / \left(\frac{\hat{e}_i^{(m)}}{\sigma}\right) \quad (382.21)$$

und

$$\hat{\mathbf{e}}^{(m)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(m)} - \mathbf{y}. \quad (382.22)$$

Die Substitution von (382.22) in (382.20) führt mit (153.11) auf den Differenzvektor

$$\hat{\boldsymbol{\beta}}^{(m+1)} - \hat{\boldsymbol{\beta}}^{(m)} = -(\mathbf{X}'\mathbf{W}^{(m)}\mathbf{X})_{rs}^{-1}\mathbf{X}'\mathbf{W}^{(m)}\hat{\mathbf{e}}^{(m)}, \quad (382.23)$$

der rechentechnisch günstiger zu berechnen ist. HUBER (1981, S.184) beweist die Konvergenz der nach (382.23) und damit nach (382.20) berechneten Schätzwerte.

383 M-Schätzung nach Huber

Nimmt man die Beobachtungen y_i als normalverteilt an, erhält man mit (241.1) und (382.11), da in (382.4) Konstanten nicht berücksichtigt werden,

$$f(x) = e^{-\frac{x^2}{2}} \quad (383.1)$$

und

$$\rho(x) = \frac{x^2}{2}, \quad (383.2)$$

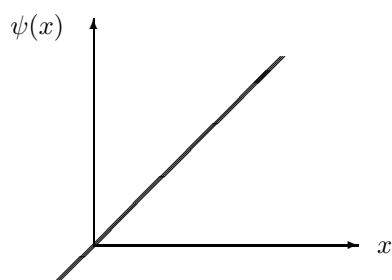
so daß anstelle von (382.8)

$$\frac{1}{2\sigma^2} \sum_{i=1}^n e_i^2 \quad (383.3)$$

folgt. Durch Minimierung erhält man die identischen Ergebnisse der besten erwartungstreuen Schätzung, der Methode der kleinsten Quadrate und der Maximum-Likelihood-Methode. Diese Schätzverfahren sind nicht robust, denn die Funktion $\psi(x)$ folgt aus (382.9) mit (383.2) zu

$$\psi(x) = x. \quad (383.4)$$

Wie aus Abbildung 383-1 ersichtlich, wächst $\psi(x)$ mit wachsendem x unbeschränkt. Die Schätzwerte ergeben sich ohne Iterationen aus (382.19), da

Abb. 383-1: $\psi(x)$ für die Parameterschätzung

das Gewicht w_i aus (382.14) mit

$$w_i = 1 \quad (383.5)$$

folgt.

Um ein robustes Schätzverfahren zu erhalten, wird für die Beobachtungen eine Wahrscheinlichkeitsdichte gewählt, die im mittleren Teil aus der Normalverteilung besteht und an den Enden aus der Laplaceverteilung, auch Exponentialverteilung genannt. Dadurch befindet sich mehr Wahrscheinlichkeitsmasse an den Enden der Verteilung als bei der Normalverteilung, so daß Ausreißer berücksichtigt werden können. Es gilt (HUBER 1981, S.71)

$$f(x) = e^{-\frac{x^2}{2}} \quad \text{für } |x| \leq c \quad (383.6)$$

$$f(x) = e^{-c|x|+c^2/2} \quad \text{für } |x| > c, \quad (383.7)$$

worin c eine Konstante bedeutet, deren Wert vom Anteil der Ausreißer in den Daten abhängt (HUBER 1981, S.87). Beträgt ihr Anteil etwa 4%, ist $c = 1,5$ zu setzen, bei etwas unter 1% gilt $c = 2,0$. Sinnvoll ist, c in diesem Bereich zu wählen. Für die Zielfunktion in (382.8) gilt

$$\rho(x) = \frac{x^2}{2} \quad \text{für } |x| \leq c \quad (383.8)$$

$$\rho(x) = c|x| - c^2/2 \quad \text{für } |x| > c \quad (383.9)$$

und für die Funktion $\psi(x)$ aus (382.9)

$$\psi(x) = x \quad \text{für } |x| \leq c \quad (383.10)$$

$$\psi(x) = c \frac{x}{|x|} \quad \text{für } |x| > c. \quad (383.11)$$

Die Plausibilität der Ableitung der nur stückweise stetig differenzierbaren Funktion $|x|$ in (383.9) kann durch

$$\frac{\partial}{\partial x} \sqrt{x^2} = \frac{x}{|x|} \quad (383.12)$$

überprüft werden. Die Funktion $\psi(x)$ aus (383.10) und (383.11) ist, wie die Abbildung 383-2 zeigt, beschränkt, denn für $|x| > c$ bleibt $\psi(x)$ konstant. Die M-Schätzung nach HUBER liefert also ein robustes Schätzverfahren.

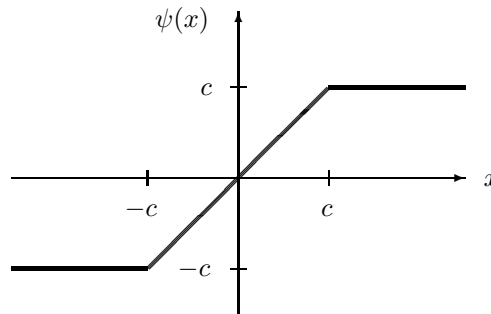


Abb. 383-2: $\psi(x)$ für die robuste M-Schätzung nach HUBER

Das Gewicht $w_i^{(m)}$ folgt mit (382.11), (382.14), (382.21), (383.10) und (383.11) zu

$$w_i^{(m)} = 1 \quad \text{für } |\hat{e}_i^{(m)}| \leq c\sigma \quad (383.13)$$

$$w_i^{(m)} = \frac{c\sigma}{|\hat{e}_i^{(m)}|} \quad \text{für } |\hat{e}_i^{(m)}| > c\sigma. \quad (383.14)$$

Die robuste M-Schätzung nach HUBER berechnet sich also durch iterative Regewichtung nach (382.20) mit den Gewichten $w_i^{(m)}$ aus (383.13) und (383.14). Ein Beispiel befindet sich im Kapitel 386.

Anstelle der von HUBER vorgeschlagenen Gewichte findet man bei der häufig angewendeten dänischen Methode (KRARUP et al. 1980; KUBIK et al. 1985)

$$w_i^{(m)} = 1 \quad \text{für } |\hat{e}_i^{(m)}| \leq 2\sigma \quad (383.15)$$

$$w_i^{(m)} \propto e^{-k\hat{e}_i^2} \quad \text{für } |\hat{e}_i^{(m)}| > 2\sigma, \quad (383.16)$$

in der k eine Konstante bedeutet. Weitere Gewichtsfestsetzungen sind bei (BENNING 1995; SOMOGYI und ZÁVOTI 1993; YANG 1994) angegeben.

Bei der M-Schätzung von HAMPEL (HAMPEL et al. 1986, S.150) ist (383.11) entsprechend $\psi(x)$ für $|x| > c$ zunächst konstant und geht dann gegen Null, so daß extreme Beobachtungen ganz vernachlässigt werden. Die zugehörige Zielfunktion ist in (HOAGLIN et al. 1983, S.366) angegeben.

384 L_p -Norm Schätzung

Wählt man die folgende Dichte für die Beobachtungen, in der wegen (382.4) wieder keine Konstante berücksichtigt wird,

$$f(x) = e^{-|x|^p} \quad \text{für } p \geq 1, \quad (384.1)$$

erhält man für $p = 1$ die Laplace-Verteilung und für $p = 2$ die Normalverteilung. Die Zielfunktion (382.8) folgt mit

$$\rho(x) = |x|^p. \quad (384.2)$$

Ohne Berücksichtigung einer Konstanten ergibt sich die Funktion $\psi(x)$ aus (382.9) zu

$$\psi(x) = |x|^{p-2} x, \quad (384.3)$$

siehe (383.12), und das Gewicht w_i aus (382.14) zu

$$w_i = |x|^{p-2}. \quad (384.4)$$

Die Dichte (384.1) führt auf die L_p -Norm-Schätzung mit der Zielfunktion

$$\|\mathbf{e}\|_p = \left(\sum_{i=1}^n |e_i|^p \right)^{1/p}, \quad (384.5)$$

die zu minimieren ist. Als Verallgemeinerung von (123.3) bedeutet $\|\mathbf{e}\|_p$ die L_p -Norm des Fehlervektors \mathbf{e} . Um zu zeigen, daß die L_p -Norm-Schätzung folgt, wird die Ableitung

$$\frac{\partial \|\mathbf{e}\|_p}{\partial \beta_k} = \left(\sum_{i=1}^n |e_i|^p \right)^{(1/p)-1} \left(\sum_{i=1}^n |e_i|^{p-2} e_i x_{ik} \right) \quad (384.6)$$

zu Null gesetzt. Da der erste Faktor auf der rechten Seite ungleich Null ist, folgt

$$\sum_{i=1}^n |\hat{e}_i|^{p-2} \hat{e}_i x_{ik} = 0 \quad \text{für } k \in \{1, \dots, u\}. \quad (384.7)$$

Das ist aber genau die Gleichung, die sich aus (382.10) durch Substitution von (384.3) ergibt.

Mit $p = 1$ folgt aus (384.5) die L_1 -Norm-Schätzung, mit der die Summe $\sum_{i=1}^n |e_i|$ der Absolutbeträge der Fehler minimiert wird und mit $p = \infty$ die Tschebyscheff-Approximation, bei der $\max |e_i|$ für $i \in \{1, \dots, n\}$ minimal wird. Die L_2 -Norm-Schätzung entspricht der Methode der kleinsten Quadrate (313.1).

Wie aus (384.3) ersichtlich, wächst für $1 \leq p < 2$ die Funktion $\psi(x)$ langsamer als x . Für $p = 1$ gilt

$$\psi(x) = x/|x|, \quad (384.8)$$

so daß $\psi(x)$ konstant ist, siehe Abbildung 384-1. Die L_1 -Norm-Schätzung ist

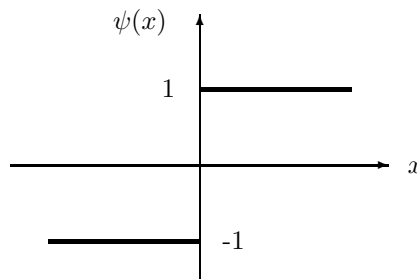


Abb. 384-1: $\psi(x)$ für die L_1 -Norm-Schätzung

also eine robuste Schätzung. Für sie ergibt sich aus (382.21) das Gewicht $w_i^{(m)}$ mit (384.4) zu

$$w_i^{(m)} = 1/|\hat{e}_i^{(m)}|, \quad (384.9)$$

in dem die Konstante σ wegen (382.18) nicht berücksichtigt zu werden braucht. Die Ausgangsgleichung für die L_1 -Norm-Schätzung $\hat{\beta}$ folgt mit $p = 1$ aus (384.7) zu

$$\sum_{i=1}^n \frac{\mathbf{x}_i' \hat{\beta} - y_i}{|\mathbf{x}_i' \hat{\beta} - y_i|} x_{ik} = 0 \quad \text{für } k \in \{1, \dots, u\}. \quad (384.10)$$

Die L_1 -Norm-Schätzung ist nicht immer eindeutig (SPÄTH 1987, S.23). Berechnet man sie aber mit dem Simplex-Algorithmus der linearen Programmierung (SPÄTH 1987, S.58), werden nicht eindeutige Lösungen entdeckt. Ein solcher Hinweis fehlt bei der iterativen Regewichtung nach (382.20) mit

den Gewichten $w_i^{(m)}$ aus (384.9). Dieses Verfahren geht auf SCHLOSSMACHER (1973) zurück.

Beispiel: Aus n voneinander unabhängigen Beobachtungen y_i soll der unbekannte Parameter β durch die L_1 -Norm-Schätzung bestimmt werden. Dann gilt in (384.10) $\hat{\beta} = \hat{\beta}, \mathbf{x}'_i = 1$ und daher

$$\sum_{i=1}^n \frac{\hat{\beta} - y_i}{|\hat{\beta} - y_i|} = 0. \quad (384.11)$$

Die Summe ist identisch Null, falls die Differenz $\hat{\beta} - y_i$ ebenso häufig positiv wie negativ wird. Ordnet man die Daten nach aufsteigender Größe

$$y_1 \leq y_2 \leq \dots \leq y_n, \quad (384.12)$$

dann ergibt sich die L_1 -Norm-Schätzung $\hat{\beta}$, die als *Median* bezeichnet wird, mit

$$\hat{\beta} = y_{(n+1)/2}, \quad (384.13)$$

falls n ungerade ist, und

$$y_{n/2} < \hat{\beta} < y_{n/2+1}, \quad (384.14)$$

falls n gerade ist. In diesem Fall ist die Schätzung nicht eindeutig, so daß häufig

$$\hat{\beta} = \frac{1}{2}(y_{n/2} + y_{n/2+1}) \quad (384.15)$$

gewählt wird, falls n gerade ist.

Durch den Median kann in 3 Beobachtungen 1 grober Fehler, bei 5 Beobachtungen 2 grobe Fehler, in 7 Beobachtungen 3 grobe Fehler und so fort entdeckt werden. Bei 4 Beobachtungen läßt sich 1 Ausreißer, bei 6 Beobachtungen 2 Ausreißer, bei 8 Beobachtungen 3 Ausreißer und so weiter identifizieren. Die Anzahl der aufzudeckenden Fehler geht bei n Beobachtungen gegen die maximal mögliche Anzahl von $n/2$, denn bei einer größeren Anzahl von Ausreißern könnte man nicht mehr zwischen Beobachtungen und Ausreißern unterscheiden. Der Median ist also äußerst robust.

385 Hebelpunkte

Ein besonderes Problem bei der robusten Parameterschätzung bilden die Hebelpunkte. Befindet sich beispielsweise bei der Anpassung einer Geraden an

Meßpunkte ein Punkt weit entfernt von den übrigen Punkten, besitzt er einen erheblichen Einfluß auf die Parameterschätzung und wird als Hebelpunkt bezeichnet. Allgemein versteht man unter *Hebelpunkten* die Daten y_i , die aufgrund ihrer Geometrie, die sich im Vektor \mathbf{x}_i der Beobachtungsgleichung (382.3) ausdrückt, weit entfernt von der Masse der übrigen Daten liegen. Zur Berechnung des Abstandes kann die Mahalanobis-Distanz (KOCH 1990, S.137) zusammen mit einer robusten Schätzung dienen (ROUSSEEUW und LEROY 1987, S.258). Ausreißer in Hebelpunkten können, wie im folgenden gezeigt wird, die Parameterschätzung erheblich beeinflussen. HAMPEL führte daher den *Bruchpunkt* einer Parameterschätzung ein (HAMPEL et al. 1986, S.12). Der Bruchpunkt gibt den Grenzwert des kleinsten Prozentsatzes von Ausreißern an, der eine Schätzung dadurch zusammenbrechen läßt, daß sie falsche Ergebnisse liefert. Der Bruchpunkt der besten erwartungstreuen Schätzung, der Methode der kleinsten Quadrate und der Maximum-Likelihood-Schätzung beträgt 0%, aber auch der Bruchpunkt der robusten M-Schätzung nach HUBER und der robusten L_1 -Norm-Schätzung beträgt 0%, da Ausreißer in Hebelpunkten nicht aufgedeckt werden (ROUSSEEUW und LEROY 1987, S.10, 149). Lediglich der Median besitzt, wie im Beispiel des vorangegangenen Kapitels gezeigt, den höchst möglichen Bruchpunkt von 50%.

Um Hebelpunkte zu charakterisieren, wird die Matrix \mathbf{H} der linearen Transformation betrachtet, die den Beobachtungsvektor \mathbf{y} in den Vektor $\hat{\mathbf{y}}$ der geschätzten Erwartungswerte überführt und die für das Gauß-Markoff-Modell mit vollem Rang mit dem Projektionsoperator \mathbf{H} in (323.6) identisch ist. Sie berechnet sich mit (323.4) und (333.4) zu

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \quad \text{mit} \quad \mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})_{rs}^{-1}\mathbf{X}' . \quad (385.1)$$

Nach (345.4) taucht die Matrix \mathbf{H} auch bei der Transformation des Beobachtungsvektors \mathbf{y} in den Vektor $\hat{\mathbf{e}}$ der Residuen auf

$$\hat{\mathbf{e}} = -(\mathbf{I} - \mathbf{H})\mathbf{y} , \quad (385.2)$$

so daß mit $\mathbf{H} = (h_{ij})$ und $i, j \in \{1, \dots, n\}$

$$\hat{e}_i = -(1 - h_{ii})y_i + \sum_{j \neq i} h_{ij}y_j \quad (385.3)$$

erhalten wird. Die Kovarianzmatrix $D(\hat{\mathbf{e}})$ des Vektors $\hat{\mathbf{e}}$ folgt mit (345.5) zu

$$D(\hat{\mathbf{e}}) = \sigma^2(\mathbf{I} - \mathbf{H}) , \quad (385.4)$$

so daß für die Varianz $V(\hat{e}_i)$ des Residuums \hat{e}_i gilt

$$V(\hat{e}_i) = \sigma^2(1 - h_{ii}) . \quad (385.5)$$

Wie sich mit (153.5) zeigen läßt, ist die Matrix \mathbf{H} idempotent und außerdem symmetrisch, daher gilt

$$h_{ii} = (\mathbf{H}\mathbf{H})_{ii} = \sum_{j=1}^n h_{ij}h_{ji} = \sum_{j=1}^n h_{ij}h_{ij} \quad (385.6)$$

oder

$$h_{ii} = h_{ii}^2 + \sum_{j \neq i} h_{ij}^2 \quad \text{für } i \in \{1, \dots, n\}. \quad (385.7)$$

Hieraus folgen

$$0 \leq h_{ii} \leq 1 \quad \text{und} \quad |h_{ij}| < 1. \quad (385.8)$$

Die Spur der Matrix \mathbf{H} ergibt sich aus (137.3), (152.3) und (153.4)

$$\text{sp}\mathbf{H} = \text{sp}((\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}) = \text{rg}((\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}) = q \quad (385.9)$$

und somit

$$\sum_{i=1}^n h_{ii} = q \quad \text{sowie} \quad \sum_{i=1}^n (1 - h_{ii}) = n - q. \quad (385.10)$$

Die Summe der Diagonalelemente r_i der Matrix $\mathbf{I} - \mathbf{H}$ in (385.2)

$$r_i = (\mathbf{I} - \mathbf{H})_{ii} \quad (385.11)$$

beträgt also $n - q$, so daß man r_i als *Teilredundanz* bezeichnet. Sie gibt den Anteil der Beobachtung y_i an der Gesamtredundanz $n - q$ an. Liegt r_i in der Nähe von Null oder h_{ii} in der Nähe von Eins, dann braucht ein Ausreißer in der Beobachtung y_i sich wegen (385.3) nicht im Residuum \hat{e}_i niederzuschlagen, sondern beispielsweise im Residuum \hat{e}_j , falls h_{ji} vergleichsweise groß ist. Macht aber ein Ausreißer in y_i sich nicht im Residuum \hat{e}_i bemerkbar, muß er wegen $\hat{e}_i = \mathbf{x}'_i \hat{\boldsymbol{\beta}} - y_i$ die Parameterschätzung erheblich verfälschen. Liegt umgekehrt r_i in der Nähe von Eins oder h_{ii} in der Nähe von Null, erkennt man den Ausreißer im Residuum, er besitzt nur einen geringen Einfluß auf die Parameterschätzung. Beobachtungen mit großen Teilredundanzen kontrollieren sich also gegenseitig.

Beobachtungen mit kleinen Teilredundanzen r_i bilden die Hebelpunkte. Umgekehrt müssen aber Hebelpunkte nicht kleine Werte für r_i bedingen. Der Grund liegt in folgendem. Befindet sich nur ein Hebelpunkt in einer Parameterschätzung mit n Beobachtungen, bewirkt er einen kleinen Wert für r_i .

Wandern weitere Beobachtungen in die Nähe des Hebelpunktes, so daß eine Gruppe von Hebelpunkten entsteht, müssen wegen (385.10) die Werte für die Teilredundanzen r_i der Hebelpunkte wachsen, da $n - q$ konstant bleibt. Die Wirkung der Hebelpunkte wird maskiert (ROUSSEEUW und LEROY 1987, S.222). Wie schon zu Beginn dieses Abschnitts erwähnt, decken auch robuste Parameterschätzungen Ausreißer in Hebelpunkten nicht auf, insbesondere wenn ihre Wirkung maskiert ist. Es sind daher spezielle Verfahren anzuwenden, die in den beiden folgenden Kapiteln behandelt werden.

386 Modifizierte M-Schätzung nach Huber

Um Ausreißer in Hebelpunkten zu entdecken, sollte die M-Schätzung von HUBER wie folgt modifiziert werden. Zum einen ist die Grenze c aus (383.6) in Abhängigkeit vom Vektor \mathbf{x}_i in (382.3) derart zu variieren, daß sie für Hebelpunkte kleiner wird. Zum anderen sind für die Beobachtungen y_i Gewichte $w(\mathbf{x}_i)$ einzuführen, die den Einfluß von Hebelpunkten verringern. Ein auf SCHWEPPE zurückgehender Vorschlag (ROUSSEEUW und LEROY 1987, S.13; HUBER 1981, S.192) sieht daher vor, die Schätzwerte nicht aus (382.13), sondern aus

$$\sum_{i=1}^n w(\mathbf{x}_i) \psi\left(\frac{\mathbf{x}_i' \hat{\boldsymbol{\beta}} - y_i}{\sigma g(\mathbf{x}_i)}\right) \frac{x_{ik}}{\sigma} = 0 \quad \text{für } k \in \{1, \dots, u\} \quad (386.1)$$

zu berechnen, wo $w(\mathbf{x}_i)$ und $g(\mathbf{x}_i)$ Funktionen von \mathbf{x}_i bedeuten. Mit (382.21), (383.10) und (383.11) folgen dann die Gewichte aus

$$w_i^{(m)} = w(\mathbf{x}_i) \quad \text{für } |\hat{e}_i^{(m)}| \leq c\sigma g(\mathbf{x}_i) \quad (386.2)$$

$$w_i^{(m)} = \frac{c\sigma g(\mathbf{x}_i) w(\mathbf{x}_i)}{|\hat{e}_i^{(m)}|} \quad \text{für } |\hat{e}_i^{(m)}| > c\sigma g(\mathbf{x}_i). \quad (386.3)$$

HUBER (1981, S.193) wählt

$$g(\mathbf{x}_i) = w(\mathbf{x}_i) = \sqrt{1 - h_{ii}} = \sqrt{r_i}, \quad (386.4)$$

so daß die Grenze c und das Gewicht für y_i wegen (385.5) durch die Standardabweichung von \hat{e}_i gesteuert wird. Eine Grenze c , die durch die Standardabweichung von \hat{e}_i variiert wird, hat ebenfalls WICKI (1992) vorgeschlagen, siehe auch (CAROSIO 1995).

Diese Vorschläge setzen voraus, daß die Teilredundanzen r_i für Hebelpunkte klein sind. Wie bereits im vorangegangenen Kapitel erläutert, braucht das aber wegen der Maskierungseffekte nicht der Fall zu sein. Nimmt man daher an, daß die Teilredundanzen r_i für Hebelpunkte, die in Gruppen auftreten,

sich nur wenig von den übrigen r_i unterscheiden, dann muß die Grenze c und das Gewicht für y_i in Abhängigkeit von r_i sehr viel drastischer gesenkt werden. Gewählt wird (KOCH 1996)

$$g(\mathbf{x}_i) = w(\mathbf{x}_i) = r_i^{p/2}/r_D \quad \text{mit} \quad r_D = \frac{1}{n} \sum_{i=1}^n r_i^{p/2}, \quad (386.5)$$

so daß aus (386.2) und (386.3) die Gewichte folgen

$$w_i^{(m)} = r_i^{p/2}/r_D \quad \text{für} \quad |\hat{e}_i^{(m)}| \leq c\sigma r_i^{p/2}/r_D \quad (386.6)$$

$$w_i^{(m)} = \frac{c\sigma r_i^p}{r_D^2 |\hat{e}_i^{(m)}|} \quad \text{für} \quad |\hat{e}_i^{(m)}| > c\sigma r_i^{p/2}/r_D \quad (386.7)$$

mit beispielsweise $p/2 = 8$. Für $p/2$ sollte aber ein kleinerer Wert gewählt werden, wenn in der Parameterschätzung kleine Werte für die Teilredundanzen r_i auftreten. Der Durchschnittswert r_D wird eingeführt, damit für $r_i^{p/2} = r_D$ in (386.6) und (386.7) als Grenze $c\sigma$ in Übereinstimmung mit (383.13) und (383.14) folgt.

Die drastische Herabsetzung der Grenze c und die Herabgewichtung der Beobachtungen y_i mit kleinen Teilredundanzen r_i ist mit einer Erhöhung der Grenze und einer Heraufgewichtung der Beobachtungen mit großen Teilredundanzen gekoppelt. Letztere Beobachtungen beeinflussen daher die Parameterschätzung erheblich, so daß Ausreißer in diesen Beobachtungen unentdeckt bleiben können. Dies wird in dem folgenden Beispiel bestätigt.

Um Ausreißer auch in Hebelpunkten aufzudecken, ist daher wie folgt vorzugehen. Begonnen wird mit einer robusten Parameterschätzung nach HUBER mit den Gewichten (383.13) und (383.14), um Ausreißer in den Beobachtungen zu identifizieren, die sich gegenseitig kontrollieren und daher große Teilredundanzen r_i besitzen. Anschließend ist die modifizierte M-Schätzung nach HUBER mit den Gewichten (386.6) und (386.7) vorzunehmen, um Ausreißer in Beobachtungen mit kleinen Teilredundanzen zu entdecken, wo Hebelpunkte vorliegen können. Befinden sich allerdings Ausreißer sowohl in Hebelpunkten als auch in Beobachtungen mit großen Teilredundanzen, wird man sie auch durch diese Schätzung in zwei Schritten nicht entdecken.

Beispiel: Für neun Punkte mit den folgenden vorgegebenen Abszissenwerten x_i wurden voneinander unabhängig mit gleichen Varianzen die folgenden Ordinaten y_i gemessen.

Punkt	1	2	3	4	5	6	7	8	9
x_i	1	2	3	21	22	23	41	42	43
y_i	2	3	4	10	10	10	42	43	44

Die Ordinaten sollen eine Gerade repräsentieren, deren Parameter β_1 und β_2 zu schätzen sind, so daß die Beobachtungsgleichungen

$$\beta_1 + x_i\beta_2 = E(y_i) = y_i + e_i \quad \text{mit} \quad V(y_i) = \sigma^2, \quad i \in \{1, \dots, 9\}$$

erhalten werden. Die nach (385.11) ermittelten Teilredundanzen r_i für die beobachteten Ordinaten y_i besitzen folgende Werte:

Punkt	1	2	3	4	5	6	7	8	9
r_i	0,706	0,723	0,739	0,888	0,889	0,888	0,739	0,723	0,706

Wie aus dem Datenmaterial ersichtlich, befinden sich in den Punkten 4, 5 und 6 mit großen Teilredundanzen r_i Ausreißer mit den Beträgen -12, -13 und -14. Die unbekannt Parameter β_1 und β_2 werden nach (382.20) zunächst durch die M-Schätzung mit den Gewichten (383.13) und (383.14) und dann durch die modifizierte M-Schätzung mit den Gewichten (386.6) und (386.7) geschätzt, wobei $c = 1, 5; \sigma = 0, 5; p/2 = 8$ gesetzt wurde. Erhalten werden die folgenden Residuen \hat{e}_i :

Punkt	1	2	3	4	5	6	7	8	9
M-Schätzg.	-0,4	-0,4	-0,4	11,6	12,6	13,6	-0,4	-0,4	-0,4
mod.M-Sch.	0,0	-0,6	-1,2	-0,3	0,1	0,5	-24,5	-25,1	-25,8

Während die M-Schätzung die Ausreißer in den Punkten 4, 5 und 6 aufdeckt, verlegt die modifizierte M-Schätzung die Ausreißer in die Punkte 7, 8 und 9. Wie bereits im Zusammenhang mit (386.6) und (386.7) erwähnt, war dieses Ergebnis zu erwarten.

Die Ausreißer in den Punkten 4, 5 und 6 werden jetzt beseitigt und mit gleichen Beträgen in die Punkte 7, 8 und 9 mit kleinen Teilredundanzen r_i verlegt. Diese Punkte sind Hebelpunkte, deren Wirkung maskiert ist. Folgende Meßdaten werden also jetzt angenommen:

Punkt	1	2	3	4	5	6	7	8	9
x_i	1	2	3	21	22	23	41	42	43
y_i	2	3	4	22	23	24	30	30	30

Es wird wieder die M-Schätzung und die modifizierte M-Schätzung angewendet. Folgende Residuen werden erhalten:

Punkt	1	2	3	4	5	6	7	8	9
M-Schätzg.	0,7	0,4	0,1	-5,7	-6,0	-6,4	-0,2	0,5	1,2
mod.M-Sch.	0,4	0,4	0,4	-0,1	-0,1	-0,1	11,4	12,4	13,3

Da die Ausreißer jetzt in Hebelpunkten liegen, deckt sie die M-Schätzung nicht auf, während sie von der modifizierten M-Schätzung erkannt werden. Weitere Beispiele befinden sich in (KOCH 1996).

387 Verfahren von Rousseeuw

Als robuste Parameterschätzung entwickelte ROUSSEEUW (1984) die Methode des kleinsten Medians der Quadrate, LMS (least median of squares)-Schätzung genannt. Sie ersetzt die Summe in der Methode der kleinsten Quadrate durch den Median. Aus den n Beobachtungen werden wiederholt u Beobachtungen herausgegriffen. Eine solche Teilmenge von u Beobachtungen, die mit dem Index $I = \{i_1, \dots, i_u\}$ versehen sei, bestimmt eindeutig die u unbekannt Parameter. Das Ergebnis sei mit β_I bezeichnet. Für jedes Resultat β_I wird die Zielfunktion der LMS-Methode berechnet

$$\text{med}_{i=1, \dots, n} (y_i - \mathbf{x}'_i \beta_I)^2 . \quad (387.1)$$

Als Schätzwert folgt der Wert für β_I , für den (387.1) minimal wird.

Der Bruchpunkt der LMS-Schätzung erreicht 50% (ROUSSEEUW und LEROY 1987, S.120), Ausreißer in Hebelpunkten werden also entdeckt. Die LMS-Methode wäre für die Ausreißersuche bestens geeignet, wäre sie nicht so rechenintensiv. Für n Beobachtungen muß bei u unbekannt Parametern nach (136.2) eine Kombination u -ter Ordnung bestimmt werden, es sind also $\binom{n}{u}$ Vektoren β_I zu berechnen, beispielsweise $\binom{40}{4} = 91390$. ROUSSEEUW UND LEROY (1987, S.198) schlagen daher vor, die Teilmenge von u Beobachtungen zufällig auszuwählen. In Abhängigkeit vom Anteil der Ausreißer in den Daten

und einer vorgegebenen Wahrscheinlichkeit von 95% oder 99% bestimmen sie die Anzahl der notwendigen Parameterschätzungen. Auch diese Zahl ist noch sehr hoch. Ein Näherungsverfahren zur Einschränkung des Rechenaufwandes der LMS-Schätzung in Triangulationsnetzen schlägt HUANG (1995) vor.

Um eine robuste Parameterschätzung zu erhalten, die auch Ausreißer in Hebelpunkten aufdeckt, schlägt KAMPMANN (1993; 1994) die balancierte Ausgleichung vor. Die Idee der Balancierung besteht darin, daß die Beobachtungen y_i für die Parameterschätzung derart mit Gewichten versehen werden, daß die Teilredundanzen r_i in (385.11) identisch werden. Unter Berücksichtigung der Balancierungsgewichte werden dann die Parameter durch Minimierung der L_1 -Norm geschätzt. Die balancierte Ausgleichung besitzt ähnliche Eigenschaften wie die modifizierte M-Schätzung mit den Gewichten (386.6) und (386.7) (KOCH 1996). Da wegen der aufwendigen Berechnung der Balancierungsgewichte die modifizierte M-Schätzung sehr viel einfacher anzuwenden ist, wird die balancierte Ausgleichung hier nicht behandelt.

4 Hypothesenprüfung, Bereichsschätzung und Ausreißertest

Die im Abschnitt 3 behandelte Parameterschätzung liefert Schätzwerte für die unbekannt Parameter und mit Hilfe ihrer Varianzen und Kovarianzen Angaben über die Streuung der Schätzwerte um ihre Erwartungswerte und über die Abhängigkeiten zwischen den Schätzwerten. Im Hinblick auf die Varianzen und Kovarianzen sollen im folgenden noch zusätzliche Angaben gemacht werden, indem für mehrere Parameter der Bereich und für einen Parameter das Intervall bestimmt wird, in dem die unbekannt Parameter bei einer vorgegebenen Wahrscheinlichkeit liegen. Man bezeichnet dieses Problem als Bereichsschätzung.

Zuvor wird jedoch die Prüfung von Hypothesen behandelt, da sich optimale Eigenschaften der Hypothesenprüfung auf die Bereichsschätzung übertragen lassen. Die Hypothesenprüfung dient dazu, vorhandene Information über die unbekannt Parameter zu testen. Diese Information kann aus vorausgegangenen oder zusätzlichen Messungen stammen oder sich auf Vermutungen stützen. Die Hypothesen werden als lineare Funktionen der Parameter formuliert, da bereits die Parameterschätzung in linearen Modellen erfolgte. Das Ergebnis der Hypothesenprüfung besteht unter Vorgabe einer Fehlerwahrscheinlichkeit in der Annahme oder Ablehnung der Hypothese.

Schließlich soll in diesem Abschnitt, wie bereits im Kapitel 381 erwähnt, das Problem gelöst werden, aus den Beobachtungen zur Parameterschätzung die grob verfälschten Werte, die sogenannten Ausreißer, in Abhängigkeit von einer vorgegebenen Fehlerwahrscheinlichkeit auszusortieren.

Die Hypothesenprüfung, die Bereichsschätzung und der Ausreißertest lassen sich bequem durchführen, wenn die Beobachtungen normalverteilt sind. Aufgrund des im Kapitel 241 erläuterten zentralen Grenzwertsatzes wird daher für die Beobachtungen in Übereinstimmung mit (324.1) die Normalverteilung angenommen, so daß im folgenden zunächst die hieraus sich ergebenden Verteilungen abgeleitet werden. Verteilungsfreie Testverfahren, die ohne die Annahme bestimmter Verteilungen auskommen, für die aber die Prüfung allgemeiner linearer Hypothesen nicht möglich ist, werden in (HOLLANDER und WOLFE 1973; SACHS 1978, S.224; WITTING und NÖLLE 1970, S.97) behandelt.

Mit Hilfe des χ^2 -Anpassungstests und des Kolmogoroff–Smirnow–Tests für die Güte der Anpassung (SACHS 1978, S.251,256; WITTING und NÖLLE 1970, S.87,167) lassen sich empirische oder angenommene univariate Verteilungen mit theoretischen oder hypothetischen univariaten Verteilungen, beispielsweise der univariaten Normalverteilung vergleichen. Verallgemeinerungen auf multivariate Normalverteilungen befinden sich in (ANDREWS et al. 1973; WITTING und NÖLLE 1970, S.89). Häufig wird mit diesen Tests die Annahme der Normalverteilung für die Beobachtungen bei Parameterschätzungen geprüft. Diese Vorgehensweise ist dann problematisch, wenn die Schätzwerte der Parameter dazu benutzt werden, um die theoretische Normalverteilung zu berechnen. Sind die Schätzwerte durch die im Kapitel 329 behandelten Modellabweichungen verfälscht, kann sich schon aus diesem Grund eine Abweichung von der Normalverteilung ergeben.

Die Hypothesenprüfung, die Bereichsschätzung und der Ausreißertest bleiben auf das Gauß–Markoff–Modell beschränkt, doch lassen sich wegen (352.4) der Ausreißertest und für die festen Parameter β die Hypothesenprüfung und Bereichsschätzung auch im gemischten Modell (352.1) anwenden. Da man diese Verfahren häufig im Gauß–Markoff–Modell mit nicht vollem Rang einsetzt (HECK et al. 1977; HEIN 1978; KOCH 1978; MIERLO 1981; PAPO und PERELMUTER 1983; PELZER 1971), soll im folgenden dieses Modell angenommen werden.

41 Verteilungen aufgrund normalverteilter Beobachtungen

411 Verteilungen von Funktionen der Residuen im univariaten Modell

Die Parameterschätzung sei im Gauß–Markoff–Modell (331.2) mit nicht vollem Rang vorgenommen worden, so daß mit der Annahme eines normalverteilten Beobachtungsvektors \mathbf{y} aus (251.1) folgt

$$\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I}) \quad \text{mit} \quad \text{rg}\mathbf{X} = q < u. \quad (411.1)$$

Durch die Transformation (321.4) erhält man wieder die Ergebnisse für das Modell mit $D(\mathbf{y}) = \sigma^2\mathbf{P}^{-1}$.

Für die Hypothesenprüfung und Bereichsschätzung benötigt man die Verteilungen der Quadratsummen der Residuen im Gauß–Markoff–Modell mit und ohne Restriktionen.

a) Verteilung der Quadratsumme Ω der Residuen

Nach (331.9) gilt für die Quadratsumme Ω der Residuen $\Omega = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$, worin $(\mathbf{X}'\mathbf{X})^{-1}$ als generalisierte Inverse von $\mathbf{X}'\mathbf{X}$ wegen (153.8) durch $(\mathbf{X}'\mathbf{X})_{rs}^{-}$ aus (333.4), durch $(\mathbf{X}'\mathbf{X})^+$ aus (333.6) oder durch $(\mathbf{X}'\mathbf{X})^{-1}$ wegen (153.22) für Modelle mit vollem Rang ersetzt werden kann. Die Matrix $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ ist, wie bei (331.7) gezeigt, idempotent, so daß auch das Produkt $(1/\sigma^2)(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\sigma^2\mathbf{I}$ idempotent ist und nach (331.7) $\text{rg}[(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')/\sigma^2] = n - q$ gilt. Mit (272.1) folgt dann für Ω/σ^2 die nichtzentrale χ^2 -Verteilung $\Omega/\sigma^2 \sim \chi'^2(n - q, \lambda)$ mit $n - q$ Freiheitsgraden und dem Nichtzentralitätsparameter wegen (153.5) $\lambda = \boldsymbol{\beta}'\mathbf{X}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X}\boldsymbol{\beta}/\sigma^2 = 0$. Hiermit folgt dann aus (262.1) die χ^2 -Verteilung (261.1)

$$\Omega/\sigma^2 \sim \chi^2(n - q). \quad (411.2)$$

b) Verteilung des Schätzwertes $\hat{\sigma}^2$ der Varianz σ^2 der Gewichtseinheit

Aus (331.12) und (411.2) folgt für $\hat{\sigma}^2$

$$(n - q)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n - q). \quad (411.3)$$

Es soll noch erwähnt werden, daß die Schätzung $\hat{\sigma}^2$, die als quadratische Form des normalverteilten Beobachtungsvektors \mathbf{y} erhalten wird, unabhängig von der Schätzung $\hat{\alpha}$ einer schätzbaren Funktion $\alpha = \mathbf{a}'\boldsymbol{\beta}$ ist, die nach (332.7) als lineare Funktion von \mathbf{y} gegeben ist, denn es gilt wegen (153.5) $\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{0}$, so daß (274.1) wegen (152.8) anwendbar ist.

c) Verteilung der quadratischen Form R

Die quadratische Form R , um die die Quadratsumme Ω der Residuen durch die Einführung von Restriktionen im Gauß-Markoff-Modell anwächst, erhält man aus (334.11) zu $R = (\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w})$, worin die Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'$, wie für (334.3) bewiesen, positiv definit ist. Da $\mathbf{H}\bar{\boldsymbol{\beta}}$ nach (334.2) schätzbare Funktionen darstellen, ergibt sich aus (255.1) mit (332.7) die Verteilung von $\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}$ in R zu

$$\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w} \sim N(\mathbf{H}\boldsymbol{\beta} - \mathbf{w}, \sigma^2\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'). \quad (411.4)$$

Die Matrix $(1/\sigma^2)(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}\sigma^2\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}' = \mathbf{I}$ ist idempotent, und es gilt $\text{rg}[(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')/\sigma^2] = r$, so daß man mit (272.1) für die Verteilung von R/σ^2 erhält

$$R/\sigma^2 \sim \chi'^2(r, \lambda) \quad (411.5)$$

mit dem Nichtzentralitätsparameter

$$\lambda = \frac{1}{\sigma^2} (\mathbf{H}\boldsymbol{\beta} - \mathbf{w})' (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1} (\mathbf{H}\boldsymbol{\beta} - \mathbf{w}). \quad (411.6)$$

d) Verteilung des Verhältnisses von R und Ω

Mit Hilfe von (273.1) soll nun gezeigt werden, daß die quadratischen Formen R und Ω voneinander unabhängig sind, so daß nach (264.1) die Verteilung ihres Verhältnisses angegeben werden kann. Setzt man (331.3) in (334.11) ein, ergibt sich $R = (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{w})' (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1} (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{w})$. Da \mathbf{H} vollen Zeilenrang besitzt, existiert nach (143.8) $(\mathbf{H}\mathbf{H}')^{-1}$, so daß $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{w} = \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{w})$ wegen (334.2) gilt. Hiermit erhält man für R

$$R = (\mathbf{y} - \mathbf{X}\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{w})' \mathbf{X}((\mathbf{X}'\mathbf{X})^{-1})' \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1} \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{w}). \quad (411.7)$$

Andererseits läßt sich wegen (153.5) die quadratische Form Ω aus (331.9) umformen in

$$\Omega = (\mathbf{y} - \mathbf{X}\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{w})' (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') (\mathbf{y} - \mathbf{X}\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{w}). \quad (411.8)$$

Damit sind R und Ω als quadratische Formen des Vektors $\mathbf{y} - \mathbf{X}\mathbf{H}'(\mathbf{H}\mathbf{H}')^{-1}\mathbf{w}$ gegeben, der nach (255.1) und (411.1) mit der Kovarianzmatrix $\sigma^2\mathbf{I}$ normalverteilt ist. Da wegen (143.7) die Matrix der ersten quadratischen Form und wegen (152.8) die Matrix der zweiten quadratischen Form positiv semidefinit sind und mit (153.5)

$$\sigma^2 \mathbf{X}((\mathbf{X}'\mathbf{X})^{-1})' \mathbf{H}'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1} \mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') = \mathbf{0} \quad (411.9)$$

gilt, folgt nach (273.1) die Unabhängigkeit von R und Ω . Mit (411.2) und (411.5) ergibt sich dann aus (264.1) für das Verhältnis $(R/r)/(\Omega/(n-q))$ die nichtzentrale F -Verteilung mit r und $n-q$ Freiheitsgraden und dem Nichtzentralitätsparameter λ aus (411.6)

$$\frac{R/r}{\Omega/(n-q)} \sim F'(r, n-q, \lambda) \quad \text{mit} \quad \lambda = \frac{1}{\sigma^2} (\mathbf{H}\boldsymbol{\beta} - \mathbf{w})' (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1} (\mathbf{H}\boldsymbol{\beta} - \mathbf{w}). \quad (411.10)$$

412 Verteilungen der im multivariaten Modell geschätzten Kovarianzmatrizen

Die Beobachtungen im multivariaten Modell (371.1) seien normalverteilt, so daß (371.5) gilt. Damit folgt, wie bereits im Zusammenhang mit (371.7) erwähnt, die Unabhängigkeit der $p \times 1$ Merkmalsvektoren \mathbf{z}_k und \mathbf{z}_l für $k \neq l$ und

$$\mathbf{z}_k \sim N(\mathbf{B}'\mathbf{x}_k, \mathbf{\Sigma}) \quad \text{mit} \quad k \in \{1, \dots, n\}. \quad (412.1)$$

Zunächst soll die Verteilung der Matrix $\mathbf{\Omega}$ der quadratischen und bilinearen Formen der Residuen abgeleitet werden. Um hierzu (285.1) anwenden zu können, wird das multivariate Modell (371.3) durch $\boldsymbol{\beta} = \mathbf{B}\mathbf{b}$ und $\mathbf{y} = \mathbf{Y}\mathbf{b}$, worin \mathbf{b} einen beliebigen $p \times 1$ Vektor von Konstanten bezeichnet, auf das univariate Modell

$$\mathbf{X}\boldsymbol{\beta} = E(\mathbf{y}) \quad \text{mit} \quad D(\mathbf{y}) = (\mathbf{b}'\mathbf{\Sigma}\mathbf{b})\mathbf{I} = \sigma_b^2\mathbf{I} \quad (412.2)$$

zurückgeführt, denn mit (131.23), (131.26), (137.6), (233.2) und (371.3) erhält man

$$\begin{aligned} D(\mathbf{y}) &= D(\text{vec}\mathbf{Y}\mathbf{b}) = D((\mathbf{b}' \otimes \mathbf{I})\text{vec}\mathbf{Y}) = (\mathbf{b}' \otimes \mathbf{I})(\mathbf{\Sigma} \otimes \mathbf{I})(\mathbf{b}' \otimes \mathbf{I})' \\ &= (\mathbf{b}'\mathbf{\Sigma}\mathbf{b}) \otimes \mathbf{I} = (\mathbf{b}'\mathbf{\Sigma}\mathbf{b})\mathbf{I}. \end{aligned}$$

Die Restriktionen $\mathbf{H}\mathbf{B} = \mathbf{W}$ in (376.2) lassen sich mit dem Vektor \mathbf{b} ebenfalls auf die Form

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{H}\mathbf{B}\mathbf{b} = \mathbf{W}\mathbf{b} = \mathbf{w} \quad (412.3)$$

der Restriktionen im univariaten Modell bringen.

Mit $\mathbf{\Omega} = (\omega_{ij})$ ergeben sich die Quadratsummen der Residuen aus (373.3) zu $\omega_{ii} = (\mathbf{X}\bar{\boldsymbol{\beta}}_i - \mathbf{y}_i)'(\mathbf{X}\bar{\boldsymbol{\beta}}_i - \mathbf{y}_i)$ für $i \in \{1, \dots, p\}$, die wegen (331.6) minimal sind. Setzt man $\bar{\boldsymbol{\beta}} = \bar{\mathbf{B}}\mathbf{b}$, erhält man mit (373.3) für die Quadratsumme Ω der Residuen des univariaten Modells (412.2)

$$\Omega = (\mathbf{X}\bar{\boldsymbol{\beta}} - \mathbf{y})'(\mathbf{X}\bar{\boldsymbol{\beta}} - \mathbf{y}) = \mathbf{b}'(\mathbf{X}\bar{\mathbf{B}} - \mathbf{Y})'(\mathbf{X}\bar{\mathbf{B}} - \mathbf{Y})\mathbf{b} = \mathbf{b}'\mathbf{\Omega}\mathbf{b}, \quad (412.4)$$

die ebenfalls minimal ist, wie sich wie für (331.6) zeigen läßt. Mit (411.2) folgt für Ω die Verteilung $\Omega/\sigma_b^2 \sim \chi^2(n-q)$ und somit nach (285.1) für die Matrix $\mathbf{\Omega} = \mathbf{Y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}$ aus (373.4) die Wishart-Verteilung

$$\mathbf{\Omega} \sim W(n-q, \mathbf{\Sigma}), \quad (412.5)$$

da die voneinander unabhängigen Zeilen \mathbf{z}_k von \mathbf{Y} nach (412.1) mit der Kovarianzmatrix $\mathbf{\Sigma}$ normalverteilt sind. Dann besitzt nach (281.1) wegen $p \leq n-q$

aus (371.1) die Matrix $\mathbf{\Omega}$ dieselbe Verteilung wie eine Matrix, deren Dichte von Null verschieden ist, falls sie positiv definit ist, und deren Dichte gleich Null ist, falls sie positiv semidefinit ist. Somit ist die Wahrscheinlichkeit gleich Null, daß die Matrix $\mathbf{\Omega}$ positiv semidefinit ist.

Die Verteilung der Matrix $\mathbf{\Omega}_H$ aus (376.6) der quadratischen und bilinearen Formen der Residuen bei Einführung von Restriktionen für die Parameter läßt sich mit den gleichen Überlegungen herleiten. Mit $\tilde{\tilde{\boldsymbol{\beta}}} = \tilde{\tilde{\mathbf{B}}}\mathbf{b}$ ergibt sich aus (376.6) die Quadratsumme Ω_H der Residuen des unvariatischen Modells zu

$$\Omega_H = (\mathbf{X}\tilde{\tilde{\boldsymbol{\beta}}} - \mathbf{y})'(\mathbf{X}\tilde{\tilde{\boldsymbol{\beta}}} - \mathbf{y}) = \mathbf{b}'(\mathbf{X}\tilde{\tilde{\mathbf{B}}} - \mathbf{Y})'(\mathbf{X}\tilde{\tilde{\mathbf{B}}} - \mathbf{Y})\mathbf{b} = \mathbf{b}'\mathbf{\Omega}_H\mathbf{b}$$

oder mit (334.11) und (376.7)

$$\Omega_H = \Omega + R = \mathbf{b}'\mathbf{\Omega}\mathbf{b} + \mathbf{b}'\mathbf{R}\mathbf{b} . \quad (412.6)$$

Nach (411.5) gilt für R die Verteilung

$$R/\sigma_b^2 \sim \chi'^2(r, \lambda)$$

mit dem Nichtzentralitätsparameter

$$\lambda = \mathbf{b}'(\mathbf{H}\mathbf{B} - \mathbf{W})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\mathbf{B} - \mathbf{W})\mathbf{b}/\sigma_b^2 .$$

Dann erhält man für \mathbf{R} mit der (411.7) entsprechenden Darstellung wegen (285.1) die nichtzentrale Wishart-Verteilung

$$\mathbf{R} \sim W'(r, \mathbf{\Sigma}, \mathbf{\Lambda}) \quad (412.7)$$

mit der Matrix $\mathbf{\Lambda}$ der Nichtzentralitätsparameter

$$\mathbf{\Lambda} = (\mathbf{H}\mathbf{B} - \mathbf{W})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\mathbf{B} - \mathbf{W})\mathbf{\Sigma}^{-1}/2 .$$

Weiter sind wegen (411.9) Ω und R voneinander unabhängig, so daß nach (285.2) auch die Matrizen $\mathbf{\Omega}$ und \mathbf{R} voneinander unabhängig sind.

Für $\mathbf{H}\mathbf{B} = \mathbf{W}$ ist $\mathbf{\Lambda} = \mathbf{0}$ in (412.7), so daß für \mathbf{R} die zentrale Wishart-Verteilung folgt. Die Verteilung von $\mathbf{\Omega}_H$ ergibt sich dann aus (283.1) und (376.7) zu

$$\mathbf{\Omega}_H \sim W(n - q + r, \mathbf{\Sigma}) \quad \text{für} \quad \mathbf{H}\mathbf{B} = \mathbf{W} . \quad (412.8)$$

Ebenso wie für $\mathbf{\Omega}$ gilt daher, daß die Wahrscheinlichkeit für eine positiv semidefinite Matrix $\mathbf{\Omega}_H$ gleich Null ist.

Mit (412.5) und (412.8) ergeben sich schließlich die Verteilungen der Schätzwerte (373.3) und (376.8) der Kovarianzmatrix Σ im Modell (371.3) zu

$$(n - q)\hat{\Sigma} \sim W(n - q, \Sigma) \quad (412.9)$$

und im Modell (376.2) zu

$$(n - q + r)\tilde{\Sigma} \sim W(n - q + r, \Sigma) \quad \text{für } \mathbf{HB} = \mathbf{W}. \quad (412.10)$$

42 Test von Hypothesen

421 Methode der Hypothesenprüfung und Likelihood-Quotiententest

Wie im Kapitel 311 erläutert, spannen die unbekannt Parametervektoren β verschiedener Stichproben den Parameterraum B auf, also $\beta \in B$. Es sei b mit $b \subset B$ eine durch (111.2) definierte Teilmenge der Vektoren des Parameterraums B . Unter einer statistischen Hypothese versteht man dann die Annahme, daß ein Parametervektor β der Teilmenge b oder der mit (112.4) definierten komplementären Menge $B \setminus b$ angehört, also $\beta \in b$ oder $\beta \in B \setminus b$.

Man bezeichnet die Annahme $H_0 : \beta \in b$ als *Nullhypothese* und die Annahme $H_1 : \beta \in B \setminus b$ als *Alternativhypothese*. Aufgrund der vorliegenden Stichprobe, ausgedrückt durch den Beobachtungsvektor \mathbf{y} , wird entschieden, ob die Nullhypothese anzunehmen oder abzulehnen ist, so daß der Wahrscheinlichkeitsraum, aus dem die Stichprobe stammt, in die Teilmenge S_K , den *Ablehnungs- oder kritischen Bereich*, und in die Teilmenge S_A , den *Annahmebereich*, aufgeteilt wird. Eine solche Aufteilung nennt man *Testverfahren*, wozu häufig als Funktion der Beobachtungen \mathbf{y} die *Testgröße* oder *Teststatistik* $t(\mathbf{y})$ eingeführt wird, so daß man H_0 ablehnt, falls $t(\mathbf{y}) \in S_K^*$ gilt, und im anderen Fall akzeptiert, wobei S_K^* den Bereich der Testgröße $t(\mathbf{y})$ bezeichnet, der dem kritischen Bereich S_K entspricht.

Mit der Annahme und der Ablehnung einer Hypothese sind zwei mögliche Fehler verbunden.

Definition: Ein *Fehler 1. Art* tritt ein, wenn $\beta \in b$ ist, aber $\mathbf{y} \in S_K$ gilt, so daß die Nullhypothese H_0 abgelehnt wird, obwohl sie wahr ist. Ein *Fehler 2. Art* tritt ein, wenn $\beta \in B \setminus b$ ist, aber $\mathbf{y} \in S_A$ gilt, so daß die Nullhypothese H_0 angenommen wird, obwohl sie falsch ist. (421.1)

Gegeben sei nun ein Testverfahren T , mit dem die Entscheidung über die Annahme oder das Ablehnen einer Hypothese getroffen werden soll. Die Entscheidungen sind, da sie auf den Beobachtungen \mathbf{y} basieren, zufälliger Art,

so daß auch die in (421.1) definierten Fehlentscheidungen zufällige Ereignisse darstellen. Um ihre Wahrscheinlichkeiten angeben zu können, wird die *Güte* $\Pi_T(\beta)$ des Tests eingeführt, die die Wahrscheinlichkeit bezeichnet, die Nullhypothese H_0 in Abhängigkeit von β abzulehnen. Die Wahrscheinlichkeit des Fehlers 1. Art ergibt sich daher zu $\Pi_T(\beta)$ für $\beta \in b$. Für die Wahrscheinlichkeit, eine falsche Nullhypothese abzulehnen, gilt $\Pi_T(\beta)$ für $\beta \in B \setminus b$. Dann ergibt sich die Wahrscheinlichkeit des Fehlers 2. Art wegen (213.4) zu $1 - \Pi_T(\beta)$ für $\beta \in B \setminus b$. Die Wahrscheinlichkeit $\Pi_T(\beta)$ für $\beta \in B \setminus b$ nennt man die *Macht* oder *Trennschärfe* des Tests.

Das Testverfahren, also die Festlegung des kritischen Bereiches, sollte derart gewählt werden, daß sowohl die Wahrscheinlichkeit des Fehlers 1. Art als auch die des Fehlers 2. Art minimal wird, beziehungsweise daß $\Pi_T(\beta)$ für $\beta \in b$ minimal und für $\beta \in B \setminus b$ maximal wird. Da sich beide Wahrscheinlichkeiten nicht gleichzeitig voneinander unabhängig minimieren und maximieren lassen, gibt man eine obere Schranke α für die Wahrscheinlichkeit des Fehlers 1. Art vor. In der Klasse der Testverfahren, die die Bedingung $\sup \Pi_T(\beta) \leq \alpha$ für alle $\beta \in b$ erfüllen, wählt man dann einen *besten* Test derart aus, daß $\Pi_T(\beta)$ für alle $\beta \in B \setminus b$ maximal wird. Die vorgegebene Schranke α bezeichnet man als *Fehlerwahrscheinlichkeit* oder *Signifikanzniveau*. Sein Wert wird meistens klein gewählt, beispielsweise $\alpha = 0, 1$; $\alpha = 0, 05$ oder $\alpha = 0, 01$.

Ein Testverfahren sollte *unverzerrt* sein, das heißt, es sollte $\Pi_T(\beta) \geq \sup_{\beta \in b} \Pi_T(\beta)$ für alle $\beta \in B \setminus b$ gelten, so daß die Wahrscheinlichkeit, eine falsche Nullhypothese abzulehnen, größer oder gleich ist der Wahrscheinlichkeit, eine richtige abzulehnen. Das Supremum $\sup_{\beta \in b} \Pi_T(\beta)$, für das der Wert α eingeführt wurde, bezeichnet man als den *Umfang* des Tests. Für einen unverzerrten Test, bei dem die Trennschärfe maximal wird, gilt die

Definition: Ein Test T der Hypothese H_0 gegen die Alternativhypothese H_1 bezeichnet man als *gleichmäßig besten unverzerrten* Test oder *trennscharfen unverzerrten* Test, wenn

- 1) $\sup_{\beta \in b} \Pi_T(\beta) = \alpha$ gilt, T also den Umfang α besitzt,
- 2) $\Pi_T(\beta) \geq \alpha$ für alle $\beta \in B \setminus b$ gilt, T also unverzerrt ist,
- 3) $\Pi_T(\beta) \geq \Pi_{T^*}(\beta)$ für alle $\beta \in B \setminus b$ und für jeden Test T^* gilt, der 1) und 2) erfüllt. (421.2)

Die Bedingungen für einen gleichmäßig besten Test, der die Forderungen 1) und 3) erfüllt, werden durch das Lemma von Neyman–Pearson angegeben (LEHMANN 1959, S.63; MOOD et al. 1974, S.411; RAO 1973, S.446). Für die Konstruktion von Testverfahren benutzt man häufig den im folgenden definierten Likelihood–Quotiententest, der in vielen Fällen auf gleichmäßig beste

oder auch auf gleichmäßig beste unverzerrte Tests führt (HUMAK 1977, S.187; WITTING und NÖLLE 1970, S.92).

Der Likelihood-Quotiententest ergibt sich mit der in (314.1) eingeführten Likelihoodfunktion $L(\mathbf{y}; \boldsymbol{\beta})$ der Beobachtungen \mathbf{y} und der unbekanntem Parameter $\boldsymbol{\beta}$.

Definition: Es sei $L(\mathbf{y}; \boldsymbol{\beta})$ mit $\boldsymbol{\beta} \in B$ die Likelihoodfunktion der Beobachtungen \mathbf{y} . Dann ist die Testgröße $\lambda(\mathbf{y})$ des *Likelihood-Quotiententests* der Hypothese $H_0 : \boldsymbol{\beta} \in b$ gegen die Alternativhypothese $H_1 : \boldsymbol{\beta} \in B \setminus b$ gegeben durch

$$\lambda(\mathbf{y}) = \frac{\sup_{\boldsymbol{\beta} \in b} L(\mathbf{y}; \boldsymbol{\beta})}{\sup_{\boldsymbol{\beta} \in B} L(\mathbf{y}; \boldsymbol{\beta})}.$$

Mit einem Signifikanzniveau von α wird H_0 abgelehnt, falls $\lambda(\mathbf{y}) < \lambda_\alpha$ gilt, wobei λ_α durch die Wahrscheinlichkeit $\sup_{\boldsymbol{\beta} \in b} \Pi_\lambda(\boldsymbol{\beta}) = P(\lambda(\mathbf{y}) < \lambda_\alpha) = \alpha$ definiert ist. (421.3)

Der Likelihood-Quotiententest liefert die Testgröße $\lambda(\mathbf{y})$, die bezüglich der zu testenden Parameter optimale Eigenschaften besitzt, da sie aus dem Quotienten der Likelihood-Maxima mit und ohne Annahme der Nullhypothese H_0 gebildet wird. Da die Likelihoodfunktion nach (314.1) eine Dichte angibt und $b \subset B$ ist, gilt $0 \leq \lambda(\mathbf{y}) \leq 1$. Je näher $\lambda(\mathbf{y})$ bei Eins liegt, desto eher muß erwartet werden, daß H_0 wahr ist. Liegt andererseits $\lambda(\mathbf{y})$ nahe bei Null, kann gefolgert werden, daß H_0 falsch und daher abzulehnen ist.

Der Ablehnungsbereich des Likelihood-Quotiententests ist durch das Intervall $0 \leq \lambda(\mathbf{y}) < \lambda_\alpha$ und der Annahmebereich durch das Intervall $\lambda_\alpha \leq \lambda(\mathbf{y}) \leq 1$ gegeben. Bezeichnet man mit $g(\lambda, H_0)$ die Verteilung der Testgröße $\lambda(\mathbf{y})$, falls H_0 wahr ist, und mit $g(\lambda, H_1)$ die Verteilung, falls H_1 richtig ist, ergibt sich daher aus (421.1) und (421.3) als Wahrscheinlichkeit des Fehlers 1. Art, also des Ablehnens einer wahren Nullhypothese

$$P(\text{Fehler 1. Art}) = P(\lambda(\mathbf{y}) < \lambda_\alpha) = \int_0^{\lambda_\alpha} g(\lambda, H_0) d\lambda = \alpha \quad (421.4)$$

und die Wahrscheinlichkeit $1 - \beta$ des Fehlers 2. Art, also die Wahrscheinlichkeit, die Nullhypothese H_0 anzunehmen, obwohl die Alternativhypothese H_1 richtig ist,

$$P(\text{Fehler 2. Art}) = P(\lambda(\mathbf{y}) > \lambda_\alpha) = \int_{\lambda_\alpha}^1 g(\lambda, H_1) d\lambda = 1 - \beta. \quad (421.5)$$

Die Wahrscheinlichkeit des zur Annahme einer falschen Nullhypothese komplementären Ereignisses, also die Wahrscheinlichkeit der Ablehnung einer falschen Nullhypothese, wenn die Alternativhypothese richtig ist, ergibt die

Trennschärfe des Tests, so daß mit (213.4) und (223.10) folgt

$$\text{Trennschärfe} = \int_0^{\lambda_\alpha} g(\lambda, H_1) d\lambda = \beta. \quad (421.6)$$

422 Test einer allgemeinen Hypothese im univariaten Gauß–Markoff–Modell

In dem linearen Gauß–Markoff–Modell werden auch die Hypothesen für die Parameter $\boldsymbol{\beta}$ als lineare Funktionen eingeführt. Ein allgemeiner Hypothesentest ergibt sich dann für das univariate Modell mit

$$H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{w} \quad \text{gegen} \quad H_1 : \mathbf{H}\boldsymbol{\beta} = \bar{\mathbf{w}} \neq \mathbf{w}, \quad (422.1)$$

worin die $r \times u$ Matrix \mathbf{H} mit $r \leq q$ vollen Zeilenrang besitze und \mathbf{w} sowie $\bar{\mathbf{w}}$ $r \times 1$ Vektoren bezeichnen. Die Nullhypothese $\mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ entspricht damit den Restriktionen in den Modellen (327.2) und (334.1). Bei letzterem Modell sollen wie in (334.2) $\mathbf{H}\boldsymbol{\beta}$ schätzbare Funktionen darstellen.

Für die Hypothese (422.1) bedeutet in der Definition (421.3) des Likelihood–Quotiententests b die Teilmenge der Parameter in B , für die die Restriktionen $\mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ gelten, und B die Menge der Parameter, die keinerlei Restriktionen unterworfen sind. Maximalwerte der Likelihoodfunktion für normalverteilte Beobachtungen \mathbf{y} im Gauß–Markoff–Modell (334.1) mit nicht vollem Rang und den Restriktionen $\mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ ergeben sich nach (334.9) und (334.10) mit $\tilde{\boldsymbol{\beta}}$ sowie $\bar{\sigma}^2$ und im Gauß–Markoff–Modell (331.2) ohne Restriktionen nach (331.4) und (331.8) mit $\boldsymbol{\beta}$ und $\bar{\sigma}^2$. Damit folgt die Testgröße $\lambda(\mathbf{y})$ aus (324.2) und (421.3) zu

$$\lambda(\mathbf{y}) = \frac{(2\pi\bar{\sigma}^2)^{n/2} \exp(-(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})/2\bar{\sigma}^2)}{(2\pi\bar{\sigma}^2)^{n/2} \exp(-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/2\bar{\sigma}^2)} = \frac{(\bar{\sigma}^2)^{n/2}}{(\bar{\sigma}^2)^{n/2}} \quad (422.2)$$

und mit (334.11)

$$\lambda(\mathbf{y}) = (\Omega/\Omega_H)^{n/2} = [1/(1 + R/\Omega)]^{n/2}. \quad (422.3)$$

Die Testgröße $\lambda(\mathbf{y})$ ist eine Funktion des Quotienten R/Ω der quadratischen Formen der Residuen, wobei $\lambda(\mathbf{y})$ mit monoton wachsendem R/Ω monoton fällt. Anstelle von $\lambda(\mathbf{y})$ kann daher auch die Testgröße $T = (R/r)/(\Omega/(n-q))$ benutzt werden, für die wegen (331.12) und (334.11) gilt

$$T = \frac{R/r}{\Omega/(n-q)} = \frac{1}{r\bar{\sigma}^2} (\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w})' (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1} (\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}). \quad (422.4)$$

Die Verteilung der Teststatistik T ergibt sich aus (411.10) zu

$$T \sim F(r, n - q), \quad (422.5)$$

falls die Nullhypothese $H_0 : \mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ gilt, und

$$T \sim F'(r, n - q, \lambda) \quad \text{mit} \quad \lambda = \frac{1}{\sigma^2}(\bar{\mathbf{w}} - \mathbf{w})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\bar{\mathbf{w}} - \mathbf{w}), \quad (422.6)$$

falls die Alternativhypothese $H_1 : \mathbf{H}\boldsymbol{\beta} = \bar{\mathbf{w}} \neq \mathbf{w}$ gilt.

Da die Testgröße T wächst, wenn $\lambda(\mathbf{y})$ fällt, erhält man anstelle des Ablehnungsbereiches $0 \leq \lambda(\mathbf{y}) < \lambda_\alpha$ aus (421.3) das Intervall $F_{1-\alpha; r, n-q} < T < \infty$, so daß die Nullhypothese H_0 mit einem Signifikanzniveau α abgelehnt wird, falls

$$T > F_{1-\alpha; r, n-q} \quad (422.7)$$

gilt, falls also die Teststatistik T größer als das $(1 - \alpha)$ -Fraktil der F -Verteilung mit r und $n - q$ Freiheitsgraden ist. Nach (263.5) gilt nämlich

$$\int_{F_{1-\alpha; r, n-q}}^{\infty} F(r, n - q) dT = \alpha, \quad (422.8)$$

so daß sich aus (421.4) die Wahrscheinlichkeit des Fehlers 1. Art ergibt zu

$$P(\text{Fehler 1. Art}) = P(T > F_{1-\alpha; r, n-q}) = \alpha. \quad (422.9)$$

Aus (421.5) und (422.6) folgt die Wahrscheinlichkeit des Fehlers 2. Art mit

$$P(\text{Fehler 2. Art}) = \int_0^{F_{1-\alpha; r, n-q}} F'(r, n - q, \lambda) dT = 1 - \beta \quad (422.10)$$

sowie aus (264.2) oder (264.3) und (421.6) die Trennschärfe β

$$\beta = 1 - F(F'; r, n - q, \lambda) \quad \text{mit} \quad F' = F_{1-\alpha; r, n-q}. \quad (422.11)$$

Der Annahme- und Ablehnungsbereich für den Test der allgemeinen Hypothese (422.1) sind in Abhängigkeit vom Signifikanzniveau α in der Abbildung 422-1 angegeben.

Die Wahrscheinlichkeit des Fehlers 2. Art sollte dann berechnet werden, wenn die Kosten, die eventuell mit diesem Fehler verbunden sind, die Kosten des Fehlers 1. Art übersteigen. Müssen zum Beispiel Bewegungen an Bauwerken durch Messungen aufgedeckt werden, so läßt sich die Nullhypothese, daß gewisse Punkte unverändert blieben, gegen die Alternativhypothese testen, daß

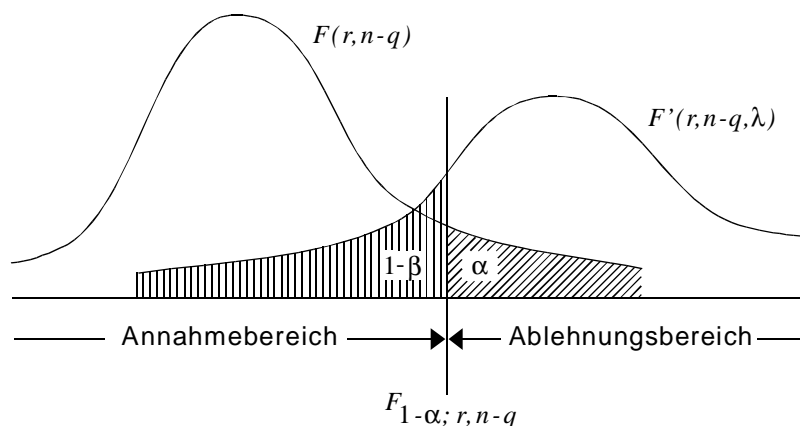


Abb. 422-1: Annahme- und Ablehnungsbereich

sich diese Punkte bewegt haben. Begeht man einen Fehler 1. Art, dann werden Bewegungen festgestellt, die nicht aufgetreten sind, so daß eventuell unnötige Verstärkungen des Bauwerkes vorgenommen werden. Bei einem Fehler 2. Art werden Bewegungen nicht entdeckt, die tatsächlich stattgefunden haben. Erreichen diese Bewegungen eine kritische Grenze, bricht das Bauwerk zusammen, so daß die Kosten des Fehlers 2. Art weit über denen des Fehlers 1. Art liegen.

Werte, für die die Kosten des Fehlers 2. Art kritisch werden können, sollten also benutzt werden, um die Alternativhypothese in (422.1) zu formulieren. Mit $\sigma^2 = \hat{\sigma}^2$ erhält man dann λ aus (422.6) und die Wahrscheinlichkeit $1 - \beta$ des Fehlers 2. Art aus (422.10). Ergibt sich $1 - \beta$ zu groß, zum Beispiel $1 - \beta > 0,10$, ist das Signifikanzniveau α des Tests zu erhöhen, von beispielsweise $\alpha = 0,05$ auf $\alpha = 0,10$, um auf diese Weise $1 - \beta$ zu verringern.

Anstatt die Ungleichung (422.7) für die Entscheidung über Annahme oder Ablehnung der Nullhypothese zu benutzen, läßt sich (422.8) entsprechend mit (263.2) die Wahrscheinlichkeit α_T ermitteln, daß die Teststatistik größer als der nach (422.4) berechnete Wert T ist

$$\alpha_T = \int_T^{\infty} F(r, n - q) dT = 1 - F(T; r, n - q). \quad (422.12)$$

Gilt für ein vorgegebenes Signifikanzniveau α

$$\alpha_T < \alpha, \quad (422.13)$$

wird die Nullhypothese abgelehnt. Interpretiert man α_T als Wahrscheinlichkeit des Fehlers 1. Art, ergibt sich die zugehörige Wahrscheinlichkeit des Fehlers 2. Art, wenn in (422.10) die obere Integrationsgrenze durch den Wert T der Teststatistik ersetzt wird.

Mit der allgemeinen Nullhypothese $\mathbf{H}\boldsymbol{\beta} = \mathbf{w}$ in (422.1) läßt sich in manchen Fällen nicht genau die Information oder die Annahme darstellen, die zu testen ist. Anstelle des Nullvektors für die Differenzen $\mathbf{H}\boldsymbol{\beta} - \mathbf{w}$ müßten nämlich kleine Intervalle oder Bereiche, die die Null einschließen, eingeführt werden. Bei der Aufdeckung von Bewegungen an Bauwerken zum Beispiel wird für die unveränderten Punkte die Hypothese getestet, daß die Differenzen ihrer Koordinaten aus zwei Meßepochen gleich Null sind. Wegen der Zentrierfehler oder der Restfehler bei der Berücksichtigung der Atmosphäre wäre es aber realistischer, die Hypothese in der Weise zu formulieren, daß für die Koordinatendifferenzen kleine Bereiche vorgesehen werden, in denen sie variieren können. Setzt man in dieser Situation dennoch $\mathbf{H}\boldsymbol{\beta} - \mathbf{w} = \mathbf{0}$, reagiert der Test zu empfindlich, das heißt, die Nullhypothese wird in Fällen abgelehnt, in denen aufgrund der Kenntnis der zu testenden Hypothese ihre Annahme hätte erwartet werden müssen.

Mit Hilfe der Bayes-Statistik lassen sich weniger empfindliche Tests ableiten, indem man für den zu testenden Differenzvektor $\mathbf{H}\boldsymbol{\beta} - \mathbf{w}$ einen kleinen Bereich in Form eines Hyperellipsoides vorgibt, der von der statistischen Beurteilung ausgeschlossen wird. Anstelle der Wahrscheinlichkeit α_T in (422.12) erhält man für einen weniger empfindlichen Test die Wahrscheinlichkeit α_{TU} zu (KOCH 1984; KOCH 1990, S.88; RIESMEIER 1984)

$$\alpha_{TU} = 1 - (1 - c)^{-1} \left(\int_0^T F(r, n - q) dT - c \right) \quad (422.14)$$

mit

$$\alpha_{TU} = 1 \quad \text{für} \quad T_R \geq T$$

und

$$c = \int_0^{T_R} F(r, n - q) dT$$

sowie

$$T_R = \frac{1}{r\hat{\sigma}^2} (\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}_R)' (\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1} (\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}_R).$$

Der Vektor $\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}_R$ enthält vorzugebende Werte, die über T_R die Größe des Hyperellipsoides bestimmen und die kleiner als die mit $\hat{\sigma}^2$ berechneten

Standardabweichungen für die Differenzen $\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}$ sein sollten, weil sonst der Test zu unempfindlich wird. Die Nullhypothese wird abgelehnt, falls

$$\alpha_{TU} < \alpha \quad (422.15)$$

gilt, wobei α das Signifikanzniveau bedeutet. Der Test mit (422.15) ist weniger empfindlich als der mit (422.12), da $\alpha_{TU} \geq \alpha_T$ gilt.

Die Entscheidung über die Annahme oder Ablehnung der Nullhypothese durch den weniger empfindlichen Test kann aber auch wie in (422.7) erfolgen, denn aus (422.14) erhält man für $\alpha_{TU} = \alpha$

$$(1 - c)^{-1} \left(\int_0^T F(r, n - q) dT - c \right) = 1 - \alpha$$

oder

$$\int_0^T F(r, n - q) dT = 1 - \alpha + c\alpha.$$

Aufgrund des weniger empfindlichen Tests ist also die Nullhypothese abzulehnen, falls

$$T > F_{1-\alpha+c\alpha; r, n-q} \quad (422.16)$$

gilt, wobei $F_{1-\alpha+c\alpha; r, n-q}$ das $(1 - \alpha + c\alpha)$ -Fraktile der F -Verteilung mit r und $n - q$ Freiheitsgraden bedeutet.

423 Spezielle Hypothesen

Aus der allgemeinen Hypothese (422.1) lassen sich spezielle Hypothesen entwickeln. Einige von ihnen, die häufig für Tests benutzt werden, sind im folgenden aufgeführt. Damit die Hypothesen, sofern sie einzelne Parameter betreffen, in Modellen mit nicht vollem Rang schätzbare Funktionen darstellen, werden nach (333.3) oder (333.5) projizierte Parameter vorausgesetzt.

a) Hypothese $H_0 : \beta_i = \beta_{oi}$ für ein festes $i \in \{1, \dots, u\}$ gegen $H_1 : \beta_i \neq \beta_{oi}$.
(423.1)

Mit $\boldsymbol{\beta}_b = (\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = (\beta_i)$ soll durch diese Hypothese getestet werden, ob der Parameter β_i einem vorgegebenen Wert β_{oi} gleicht. In (422.1) gilt dann $\mathbf{H} = |0, \dots, 0, 1, 0, \dots, 0|$, $\mathbf{w} = \beta_{oi}$ und $r = 1$, wobei der Wert Eins in der i -ten Spalte von \mathbf{H} steht. Die Testgröße T ergibt sich mit (333.4) und (422.4) zu

$$T = (\hat{\beta} - \beta_{oi})^2 / \hat{\sigma}_i^2, \quad (423.2)$$

sofern $\hat{\sigma}_i^2 \neq 0$ gilt, wobei $\hat{\sigma}_i^2 = \hat{\sigma}^2 \sigma_{ii}$ bedeutet mit

$$(\mathbf{X}'\mathbf{X})_{rs}^- = \boldsymbol{\Sigma}_\beta = (\sigma_{ij}) \quad (423.3)$$

und $(\mathbf{X}'\mathbf{X})_{rs}^-$ nach (333.5) durch $(\mathbf{X}'\mathbf{X})^+$ oder nach (153.22) durch $(\mathbf{X}'\mathbf{X})^{-1}$ für Modelle mit vollem Rang ersetzt werden kann. Für die Verteilung von T folgt, wenn die Nullhypothese gilt, aus (422.5) und (265.2)

$$T \sim F(1, n - q) \quad \text{oder} \quad \sqrt{T} \sim t(n - q). \quad (423.4)$$

$$\text{b) Hypothese } H_0 : \mathbf{b}'\boldsymbol{\beta} = w_0 \text{ gegen } H_1 : \mathbf{b}'\boldsymbol{\beta} \neq w_0. \quad (423.5)$$

Hiermit soll getestet werden, ob die lineare und schätzbare Funktion $\mathbf{b}'\boldsymbol{\beta}$ der unbekannt Parameter $\boldsymbol{\beta}$ den vorgegebenen Wert w_0 besitzt. In (422.1) gilt dann $\mathbf{H} = \mathbf{b}'$, $\mathbf{w} = w_0$ und $r = 1$, so daß sich die Testgröße T aus (422.4) mit $\hat{V}(\widehat{\mathbf{b}'\boldsymbol{\beta}}) = \hat{\sigma}^2 \mathbf{b}'(\mathbf{X}'\mathbf{X})^- \mathbf{b}$ aus (332.7) ergibt zu

$$T = \frac{(\widehat{\mathbf{b}'\boldsymbol{\beta}} - w_0)^2}{\hat{V}(\widehat{\mathbf{b}'\boldsymbol{\beta}})} \quad (423.6)$$

mit den Verteilungen wie in (423.4)

$$T \sim F(1, n - q) \quad \text{oder} \quad \sqrt{T} \sim t(n - q). \quad (423.7)$$

$$\text{c) Hypothese } H_0 : \beta_i = \beta_{oi} \text{ für alle } i \in \{j, j + 1, \dots, k\} \text{ gegen } H_1 : \beta_i \neq \beta_{oi} \text{ für wenigstens ein } i \in \{j, j + 1, \dots, k\}. \quad (423.8)$$

Für die Hypothese, daß die $k - j + 1$ Parameter β_j bis β_k den vorgegebenen Werten β_{oj} bis β_{ok} gleichen, ist die Matrix \mathbf{H} in (422.1) durch eine $(k - j + 1) \times u$ Matrix zu ersetzen, deren Spalten j bis k durch eine Einheitsmatrix gebildet werden und deren übrige Spalten nur aus Nullen bestehen. Mit $\boldsymbol{\beta}_{j\dots k} = \mathbf{H}\boldsymbol{\beta}_b = \mathbf{H}(\mathbf{X}'\mathbf{X})_{rs}^- \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = (\beta_i)$, $\boldsymbol{\beta}_{o,j\dots k} = (\beta_{oi})$, $i \in \{j, \dots, k\}$ und aus (423.3) mit $(\boldsymbol{\Sigma}_\beta)_{j\dots k} = (\sigma_{il})$ für $i, l \in \{j, \dots, k\}$ ergibt sich, da $(\boldsymbol{\Sigma}_\beta)_{j\dots k}$ bei Beachtung von (334.17) regulär ist, so daß $(\boldsymbol{\Sigma}_\beta)_{j\dots k}^{-1} = [(\boldsymbol{\Sigma}_\beta)_{j\dots k}]^{-1}$ existiert, für T und seine Verteilung im Falle der gültigen Nullhypothese

$$T = \frac{1}{(k - j + 1)\hat{\sigma}^2} (\hat{\boldsymbol{\beta}}_{j\dots k} - \boldsymbol{\beta}_{o,j\dots k})' (\boldsymbol{\Sigma}_\beta)_{j\dots k}^{-1} (\hat{\boldsymbol{\beta}}_{j\dots k} - \boldsymbol{\beta}_{o,j\dots k}) \quad \text{und} \quad T \sim F(k - j + 1, n - q). \quad (423.9)$$

$$\text{d) Hypothese } H_0 : \beta_i = 0 \text{ für alle } i \in \{j, j + 1, \dots, k\} \text{ gegen } H_1 : \beta_i \neq 0 \text{ für wenigstens ein } i \in \{j, j + 1, \dots, k\}. \quad (423.10)$$

Für die Hypothese, daß die $k - j + 1$ Parameter β_j bis β_k den Wert Null besitzen, erhält man (423.9) entsprechend

$$T = \frac{1}{(k - j + 1)\hat{\sigma}^2} \hat{\beta}'_{j\dots k} (\boldsymbol{\Sigma}_\beta)_{j\dots k}^{-1} \hat{\beta}_{j\dots k} \quad \text{und} \quad T \sim F(k - j + 1, n - q). \quad (423.11)$$

Dieser Test, angewendet auf die Effekte sämtlicher Stufen eines Faktors oder auf die Wechselwirkungen zwischen zwei Faktoren, stellt den wichtigsten Hypothesentest in der Varianzanalyse dar. Die Hypothese $H_0 : \beta_j = 0$ für $j \in \{1, \dots, q\}$ in (342.2) unterstellt beispielsweise, daß die Effekte des Faktors B ohne Einfluß auf das zu untersuchende Merkmal sind. Angewendet wird dieser Test im zweiten Beispiel des Kapitels 425.

Die Hypothese (423.10) kann auch dazu benutzt werden, den Grad zu bestimmen, bis zu dem die Polynommodelle (341.3) oder (341.4) zu entwickeln sind, indem beginnend mit dem höchsten Grad und absteigend zu den niedrigeren Graden die Identität der Parameter der Polynomentwicklung mit Null getestet wird, so daß für die Matrix \mathbf{H} in (422.1) $\mathbf{H} = |\mathbf{0}, \mathbf{I}|$ gilt. Bei diesen Tests ist es unerheblich, ob die j bis u Parameter $\beta_{j\dots u}$ der insgesamt u Parameter $\boldsymbol{\beta}$ der Polynome in (341.3) oder die j bis u Parameter $\gamma_{j\dots u}$ der orthogonalen Polynome in (341.4) der Hypothesenprüfung unterzogen werden, denn mit $\beta_{j\dots u} = \mathbf{H}\boldsymbol{\beta} = \mathbf{H}(\mathbf{G}')^{-1}\boldsymbol{\gamma}$ und $(\boldsymbol{\Sigma}_\beta)_{j\dots u}^{-1} = (\mathbf{H}(\mathbf{G}\mathbf{G}')^{-1}\mathbf{H}')^{-1}$ aus (341.5) und (422.4) erhält man in (423.11)

$$\hat{\beta}'_{j\dots u} (\boldsymbol{\Sigma}_\beta)_{j\dots u}^{-1} \hat{\beta}_{j\dots u} = \hat{\boldsymbol{\gamma}}' \mathbf{G}^{-1} \mathbf{H}' ((\mathbf{G}^{-1} \mathbf{H}')' \mathbf{G}^{-1} \mathbf{H}')^{-1} (\mathbf{G}^{-1} \mathbf{H}')' \hat{\boldsymbol{\gamma}}.$$

Mit $\mathbf{H} = |\mathbf{0}, \mathbf{I}|$ und einer entsprechenden Zerlegung der unteren Dreiecksmatrix \mathbf{G} in

$$\mathbf{G} = \begin{vmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{G}_2 & \mathbf{G}_3 \end{vmatrix} \quad \text{folgt} \\ \mathbf{G}^{-1} \mathbf{H}' = \begin{vmatrix} \mathbf{G}_1^{-1} & \mathbf{0} \\ -\mathbf{G}_3^{-1} \mathbf{G}_2 \mathbf{G}_1^{-1} & \mathbf{G}_3^{-1} \end{vmatrix} \begin{vmatrix} \mathbf{0} \\ \mathbf{I} \end{vmatrix} = \begin{vmatrix} \mathbf{0} \\ \mathbf{G}_3^{-1} \end{vmatrix}$$

und damit

$$\mathbf{G}^{-1} \mathbf{H}' ((\mathbf{G}^{-1} \mathbf{H}')' \mathbf{G}^{-1} \mathbf{H}')^{-1} (\mathbf{G}^{-1} \mathbf{H}')' \\ = \begin{vmatrix} \mathbf{0} \\ \mathbf{G}_3^{-1} \end{vmatrix} \left(((\mathbf{G}'_3)^{-1} \mathbf{G}_3^{-1})^{-1} \mathbf{0}, (\mathbf{G}'_3)^{-1} \right) = \begin{vmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{vmatrix} = \mathbf{H}' \mathbf{H}$$

sowie schließlich

$$\hat{\beta}'_{j\dots u} (\boldsymbol{\Sigma}_\beta)_{j\dots u}^{-1} \hat{\beta}_{j\dots u} = \hat{\boldsymbol{\gamma}}'_{j\dots u} \hat{\boldsymbol{\gamma}}_{j\dots u}.$$

Bei insgesamt u Parametern werden die Hypothesen $\beta_u = 0, \beta_u = \beta_{u-1} = 0, \beta_u = \beta_{u-1} = \beta_{u-2} = 0$ und so fort getestet, bis eine Hypothese abgelehnt werden muß. Dabei sollte das Signifikanzniveau α klein gewählt werden, denn die Wahrscheinlichkeit, bei einem Signifikanzniveau von α die Nullhypothese in k Tests anzunehmen, wenn sie wahr ist, berechnet sich nach (215.2) zu $\prod_{i=1}^k (1 - \alpha)$, falls Unabhängigkeit der Tests angenommen wird. Damit folgt die Wahrscheinlichkeit α^* des Fehlers 1. Art des gesamten Tests, also des Ablehnens zumindest einer wahren Nullhypothese zu

$$\alpha^* = 1 - \prod_{i=1}^k (1 - \alpha) . \quad (423.12)$$

Mit $k = 5$ und $\alpha = 0,01$ beispielsweise ergibt sich $\alpha^* = 0,05$. Ein Testverfahren für eine optimale Entscheidung über den Grad der Polynomentwicklung befindet sich bei (ANDERSON 1962).

Beispiel: Im Beispiel des Kapitels 341 soll die Hypothese

$$H_0 : \beta_2 = 0 \quad \text{gegen} \quad H_1 : \beta_2 = \hat{\beta}_2 \neq 0$$

getestet werden. Man erhält aus (423.2) die Testgröße T zu

$$T = \frac{\hat{\beta}_2^2}{\hat{\sigma}_2^2} = \frac{0,1242^2}{0,0250 \cdot 0,0171} = 36,1 .$$

Mit $r = 1, n - q = 3$ und einem Signifikanzniveau von $\alpha = 0,01$ beträgt das $(1 - \alpha)$ -Fraktile der F -Verteilung $F_{0,99;1,3} = 34,1$, so daß wegen (422.7) die Nullhypothese abzulehnen ist. Ist die Alternativhypothese H_1 richtig, und nimmt man zur Berechnung von λ aus (422.6) an, daß $\sigma^2 = \hat{\sigma}^2$ gilt, folgt aus dem Vergleich mit (422.4) $\lambda = rT$ oder $\lambda = 36,1$ und damit aus (422.10)

$$P(\text{Fehler 2. Art}) = 0,38 ,$$

so daß sich die Trennschärfe des Tests, also die Wahrscheinlichkeit des Ablehnens einer falschen Nullhypothese bei richtiger Alternativhypothese H_1 aus (422.11) zu $\beta = 0,62$ ergibt. Berechnet man mit $T = 36,1$ den Wert für α_T aus (422.12), folgt $\alpha_T = 0,009$. Für $\alpha = 0,01$ ist also die Hypothese wegen (422.13) abzulehnen, wie bereits mit Hilfe von $F_{0,99;1,3}$ festgestellt wurde.

e) Test des empirischen multiplen Korrelationskoeffizienten.

Wendet man die Hypothese (423.10) auf sämtliche u Parameter β_i eines Modells mit vollem Rang an, ergibt sich die

$$\text{Hypothese } H_0 : \boldsymbol{\beta} = \mathbf{0} \quad \text{gegen} \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0} \quad (423.13)$$

und die Testgröße T mit ihrer Verteilung aus (423.11)

$$T = \frac{1}{u\hat{\sigma}^2} \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}} \quad \text{mit} \quad T \sim F(u, n - u). \quad (423.14)$$

In Analogie zu (232.9) und zur Definition des multiplen Korrelationskoeffizienten bei (351.11) bezeichnet man

$$R^2 = \frac{(\mathbf{y}' \hat{\mathbf{y}})^2}{(\mathbf{y}' \mathbf{y})(\hat{\mathbf{y}}' \hat{\mathbf{y}})} \quad (423.15)$$

als *empirischen multiplen Korrelationskoeffizienten* der Beobachtungen \mathbf{y} und der geschätzten Erwartungswerte $\hat{\mathbf{y}}$ der Beobachtungen. In einem Modell mit einer Konstanten als unbekannter Parameter, der aber nicht in der Hypothese (423.13) enthalten ist, so daß u in (423.14) durch $u - 1$ zu ersetzen ist, beziehen sich \mathbf{y} und $\hat{\mathbf{y}}$ auf den Mittelwert der Beobachtungen (FAHRMEIR und HAMERLE 1984, S.99). Mit (322.9) und (323.4) folgt $\mathbf{y}' \hat{\mathbf{y}} = \hat{\mathbf{y}}' \hat{\mathbf{y}} = \hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}$ und damit

$$R^2 = \frac{\hat{\boldsymbol{\beta}}' \mathbf{X}' \mathbf{X} \hat{\boldsymbol{\beta}}}{\mathbf{y}' \mathbf{y}} \quad \text{sowie mit (325.2)} \quad 1 - R^2 = \frac{\Omega}{\mathbf{y}' \mathbf{y}}.$$

Folglich gilt für die Testgröße T in (423.14)

$$T = \frac{(n - u)R^2}{u(1 - R^2)}, \quad (423.16)$$

so daß man die Hypothese (423.13) als Test des empirischen multiplen Korrelationskoeffizienten R^2 interpretieren kann, der auf $R^2 = 0$ gegen $R^2 \neq 0$ getestet wird.

424 Hypothesentest für die Varianz der Gewichtseinheit

Nach (411.3) gilt $(n - q)\hat{\sigma}^2/\sigma^2 \sim \chi^2(n - q)$. Soll daher die

$$\text{Hypothese } H_0 : \sigma^2 = \sigma_0^2 \quad \text{gegen} \quad H_1 : \sigma^2 > \sigma_0^2 \quad (424.1)$$

getestet werden, läßt sich die Teststatistik

$$T = (n - q)\hat{\sigma}^2/\sigma_0^2 \quad (424.2)$$

bilden, und man lehnt die Hypothese mit einem Signifikanzniveau von α ab, falls

$$T > \chi_{1-\alpha; n-q}^2 \quad (424.3)$$

gilt, wobei $\chi_{1-\alpha;n-q}^2$ nach (261.10) das $(1 - \alpha)$ -Fraktile der χ^2 -Verteilung bezeichnet. Die

$$\text{Hypothese } H_0 : \sigma^2 = \sigma_0^2 \text{ gegen } H_1 : \sigma^2 < \sigma_0^2 \quad (424.4)$$

wird abgelehnt, falls

$$T < \chi_{\alpha;n-q}^2 \quad (424.5)$$

gilt, und soll schließlich die

$$\text{Hypothese } H_0 : \sigma^2 = \sigma_0^2 \text{ gegen } H_1 : \sigma^2 \neq \sigma_0^2 \quad (424.6)$$

getestet werden, so ist sie mit einem Signifikanzniveau von α abzulehnen, falls

$$T > \chi_{1-\alpha/2;n-q}^2 \text{ oder } T < \chi_{\alpha/2;n-q}^2 \quad (424.7)$$

gilt. Diesen Test bezeichnet man im Gegensatz zu den bisherigen *einseitigen* Hypothesentests als *zweiseitigen* Test, da für ihn die Wahrscheinlichkeit an den beiden Seiten des Graphen der Verteilung berechnet wird.

Soll die Wahrscheinlichkeit des Fehlers 2. Art der Hypothese (424.1) angegeben werden, falls die Alternativhypothese $H_1 : \sigma^2 = \sigma_1^2 > \sigma_0^2$ richtig ist, so ist $P(T < \chi_{1-\alpha;n-q}^2)$ unter der Bedingung $\sigma_0^2 T / \sigma_1^2 \sim \chi^2(n - q)$ zu ermitteln, da σ_0^2 in der Teststatistik T aus (424.2) durch σ_1^2 zu ersetzen ist. Somit folgt

$$P\left(\frac{\sigma_0^2}{\sigma_1^2} T < \frac{\sigma_0^2}{\sigma_1^2} \chi_{1-\alpha;n-q}^2\right) = 1 - \beta \quad (424.8)$$

und mit (261.5)

$$1 - \beta = F(\chi^2; n - q) \text{ mit } \chi^2 = \frac{\sigma_0^2}{\sigma_1^2} \chi_{1-\alpha;n-q}^2. \quad (424.9)$$

Entsprechende Betrachtungen gelten auch für die Alternativhypothese $H_1 : \sigma^2 = \sigma_1^2 < \sigma_0^2$ in (424.4) und für $H_1 : \sigma^2 = \sigma_1^2 \neq \sigma_0^2$ in (424.6). Im letzten Fall ergibt sich

$$1 - \beta = F(\chi_1^2; n - q) - F(\chi_2^2; n - q) \\ \text{mit } \chi_1^2 = \frac{\sigma_0^2}{\sigma_1^2} \chi_{1-\alpha/2;n-q}^2 \text{ und } \chi_2^2 = \frac{\sigma_0^2}{\sigma_1^2} \chi_{\alpha/2;n-q}^2. \quad (424.10)$$

425 Test einer allgemeinen Hypothese im multivariaten Gauß–Markoff–Modell

Die Parameter β_i des multivariaten Gauß–Markoff–Modells (371.1) lassen sich getrennt von den Parametern β_j mit $i \neq j$ wie bei der Hypothesenprüfung (422.1) im univariaten Modell durch die allgemeine

$$\text{Hypothese } H_0 : \mathbf{H}\beta_i = \mathbf{w}_i \quad \text{gegen} \quad H_1 : \mathbf{H}\beta_i = \bar{\mathbf{w}}_i \neq \mathbf{w}_i$$

testen. Darüber hinaus kann man aber sämtliche Parameter β_i für $i \in \{1, \dots, p\}$ in der Parametermatrix \mathbf{B} in (371.3) einer gemeinsamen Hypothese unterwerfen, indem die p Vektoren \mathbf{w}_i und $\bar{\mathbf{w}}_i$ in den $r \times p$ Matrizen \mathbf{W} und $\bar{\mathbf{W}}$ zusammengefaßt werden, so daß die

$$\text{Hypothese } H_0 : \mathbf{H}\mathbf{B} = \mathbf{W} \quad \text{gegen} \quad H_1 : \mathbf{H}\mathbf{B} = \bar{\mathbf{W}} \neq \mathbf{W} \quad (425.1)$$

eingeführt wird, die den Restriktionen $\mathbf{H}\mathbf{B} = \mathbf{W}$ im Modell (376.2) entspricht.

Für diese Hypothese führt der Likelihood–Quotiententest (421.3) auf die Testgröße

$$\lambda(\text{vec}\mathbf{Y}) = \frac{\sup_{\mathbf{B}, \Sigma \in b} L(\text{vec}\mathbf{Y}; \mathbf{B}, \Sigma)}{\sup_{\mathbf{B}, \Sigma \in B} L(\text{vec}\mathbf{Y}; \mathbf{B}, \Sigma)}. \quad (425.2)$$

Der Parameterraum B enthält als Parameter neben \mathbf{B} auch Σ und umfaßt die Menge der Parameter ohne Restriktionen, während b wieder die Teilmenge der Parameter bedeutet, die den Restriktionen $\mathbf{H}\mathbf{B} = \mathbf{W}$ unterworfen sind. Die Maximum–Likelihood–Schätzung von \mathbf{B} und Σ für normalverteilte Beobachtungen im Modell mit nicht vollem Rang ergibt sich bei Restriktionen aus (376.5) und (376.9) zu $\tilde{\mathbf{B}}$ und $\tilde{\Sigma}$ und ohne Restriktionen aus (372.5) und (373.6) zu $\bar{\mathbf{B}}$ und $\bar{\Sigma}$. Mit (372.7) erhält man daher für (425.2)

$$\lambda(\text{vec}\mathbf{Y}) = \frac{(2\pi)^{np/2} (\det \tilde{\Sigma})^{n/2} e^{np/2}}{(2\pi)^{np/2} (\det \bar{\Sigma})^{n/2} e^{np/2}} = \frac{(\det \tilde{\Sigma})^{n/2}}{(\det \bar{\Sigma})^{n/2}}.$$

Anstelle von $\lambda(\text{vec}\mathbf{Y})$ wird die mit $\lambda(\text{vec}\mathbf{Y})$ monoton wachsende Testgröße $\Lambda_{p,r,n-q} = [\lambda(\text{vec}\mathbf{Y})]^{2/n}$ eingeführt, die sich mit (373.6), (376.7) und (376.9) ergibt zu

$$\Lambda_{p,r,n-q} = \frac{\det \mathbf{\Omega}}{\det(\mathbf{\Omega} + \mathbf{R})}. \quad (425.3)$$

Man bezeichnet sie als *Likelihood–Quotientenkriterium* von WILKS (1932).

Nach (412.5) besitzt die Matrix $\mathbf{\Omega}$ die zentrale und nach (412.7) die Matrix \mathbf{R} die nichtzentrale Wishart-Verteilung. Ist die Nullhypothese $\mathbf{HB} = \mathbf{W}$ in (425.1) gültig, ist $\mathbf{\Lambda} = \mathbf{0}$ in (412.7) und für \mathbf{R} folgt die zentrale Wishart-Verteilung $\mathbf{R} \sim W(r, \mathbf{\Sigma})$. Die Verteilung der Teststatistik $\Lambda_{p,r,n-q}$ ergibt sich dann aus (286.3) bis (286.5), so daß die Hypothese (425.1) mit einem Signifikanzniveau von α abzulehnen ist, falls gilt

$$\Lambda_{p,r,n-q} < \Lambda_{\alpha;p,r,n-q} , \quad (425.4)$$

wobei $\Lambda_{\alpha;p,r,n-q}$ das α -Fraktile (286.9) der Verteilung für $\Lambda_{p,r,n-q}$ im Falle der gültigen Nullhypothese bedeutet.

Ist nicht die Nullhypothese in (425.1), sondern die Alternativhypothese $\mathbf{HB} = \bar{\mathbf{W}}$ richtig, gilt für die Matrix \mathbf{R} die nichtzentrale Wishart-Verteilung $\mathbf{R} \sim W'(r, \mathbf{\Sigma}, \mathbf{\Lambda})$ mit $\mathbf{HB} = \bar{\mathbf{W}}$ in (412.7). Die Verteilungsfunktion der Teststatistik $\Lambda_{p,r,n-q}$ folgt dann aus (286.10) und damit die Wahrscheinlichkeit $1 - \beta$ des Fehlers 2. Art wegen (421.5) aus

$$P(\Lambda_{p,r,n-q} > \Lambda_{\alpha;p,r,n-q}) = 1 - \beta .$$

Mit (286.8) läßt sich wie in (422.12) die Wahrscheinlichkeit α_{Λ} berechnen

$$\alpha_{\Lambda} = \int_0^{\Lambda_{p,r,n-q}} f(\Lambda) d\Lambda , \quad (425.5)$$

falls $f(\Lambda)$ die Dichte der Testgröße bedeutet. Gilt für ein vorgegebenes Signifikanzniveau α

$$\alpha_{\Lambda} < \alpha , \quad (425.6)$$

ist die Nullhypothese abzulehnen.

Mit der Wahrscheinlichkeit $\alpha_{\Lambda U}$ anstelle von α_{Λ} kann man wie in (422.14) einen weniger empfindlichen Test der Hypothese (425.1) angeben (KOCH und RIESMEIER 1985)

$$\alpha_{\Lambda U} = 1 - (1 - c)^{-1} \left(\int_{\Lambda_{p,r,n-q}}^1 f(\Lambda) d\Lambda - c \right) \quad (425.7)$$

mit

$$\alpha_{\Lambda U} = 1 \quad \text{für} \quad \Lambda_{p,r,n-q}(\mathbf{W}_R) \leq \Lambda_{p,r,n-q}$$

und

$$c = \int_{\Lambda_{p,r,n-q}(\mathbf{W}_R)}^1 f(\Lambda) d\Lambda$$

sowie

$$\Lambda_{p,r,n-q}(\mathbf{W}_R) = \frac{\det \mathbf{\Omega}}{\det(\mathbf{\Omega} + \mathbf{R})}$$

mit $\mathbf{\Omega}$ aus (425.3) und

$$\mathbf{R} = (\mathbf{H}\bar{\mathbf{B}} - \mathbf{W}_R)'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\bar{\mathbf{B}} - \mathbf{W}_R).$$

Die Matrix $\mathbf{H}\bar{\mathbf{B}} - \mathbf{W}_R$ enthält wieder die vorzugebenden Werte, die den Bereich bestimmen, der von der statistischen Beurteilung ausgeschlossen wird. Die Nullhypothese wird abgelehnt, falls

$$\alpha_{\Lambda U} < \alpha \quad (425.8)$$

oder falls

$$\Lambda_{p,r,n-q} < \Lambda_{\alpha-c\alpha;p,r,n-q} \quad (425.9)$$

gilt, wobei $\Lambda_{\alpha-c\alpha;p,r,n-q}$ das $(\alpha - c\alpha)$ -Fraktile der Verteilung für $\Lambda_{p,r,n-q}$ bedeutet.

Die Hypothese (425.1) erlaubt die Prüfung der Spalten der Parametermatrix \mathbf{B} , also der Parameter, die durch ein Merkmal geschätzt werden. Will man gleichzeitig auch die Zeilen von \mathbf{B} testen, also die Parameter aus verschiedenen Merkmalen, so ist die

$$\text{Hypothese } H_0 : \mathbf{HBU} = \mathbf{W} \quad \text{gegen } H_1 : \mathbf{HBU} = \bar{\mathbf{W}} \neq \mathbf{W} \quad (425.10)$$

einzuführen, in der \mathbf{U} eine $p \times s$ Matrix mit $\text{rg}\mathbf{U} = s < p$ und \mathbf{W} eine $r \times s$ Matrix bedeuten. Ersetzt man in den Modellen (371.3) und (376.2) \mathbf{B} durch \mathbf{BU} und \mathbf{Y} durch \mathbf{YU} , ergibt sich mit (137.6) $\text{vec}(\mathbf{YU}) = (\mathbf{U}' \otimes \mathbf{I})\text{vec}\mathbf{Y}$ und aus (233.2) wegen (131.23) und (131.26) die Kovarianzmatrix

$$D(\text{vec}(\mathbf{YU})) = (\mathbf{U}' \otimes \mathbf{I})(\mathbf{\Sigma} \otimes \mathbf{I})(\mathbf{U}' \otimes \mathbf{I})' = (\mathbf{U}'\mathbf{\Sigma}\mathbf{U}) \otimes \mathbf{I}.$$

Durch die Substitution wird also ein multivariates Modell in ein anderes überführt, so daß \mathbf{BU} anstelle von \mathbf{B} in den Modellen (371.3) und (376.2) geschätzt wird. Daher ist aus (373.3) und (376.7) ersichtlich, daß für die Hypothese (425.10) die $s \times s$ Matrizen $\mathbf{\Omega}$ und \mathbf{R} in (425.3) oder (425.7) zu berechnen sind aus

$$\mathbf{\Omega} = \mathbf{U}'(\mathbf{X}\bar{\mathbf{B}} - \mathbf{Y})'(\mathbf{X}\bar{\mathbf{B}} - \mathbf{Y})\mathbf{U} \quad (425.11)$$

und

$$\mathbf{R} = (\mathbf{H}\bar{\mathbf{B}}\mathbf{U} - \mathbf{W})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\bar{\mathbf{B}}\mathbf{U} - \mathbf{W}). \quad (425.12)$$

Weiter folgen wegen (412.5) und (412.7) für $\mathbf{\Omega}$ und \mathbf{R} Wishart-Verteilungen mit $n - q$ und r Freiheitsgraden, so daß in (425.3) oder (425.7) lediglich noch p durch s zu ersetzen ist.

1. Beispiel: Bei einer Auswertung von Wiederholungsmessungen zur Überwachung von Bauwerken enthalte die Matrix \mathbf{B} der unbekannt Parameter die ebenen Koordinaten x_{ki}, y_{ki} , bei denen der erste Index k die Punktnummer und der zweite Index i die Wiederholungsmessung bezeichne. Liegen drei Wiederholungsmessungen vor und soll geprüft werden, ob die Koordinatendifferenzen des zweiten und dritten Punktes zwischen den Wiederholungsmessungen mit Null identisch sind, folgt als Nullhypothese H_0

$$\begin{aligned} x_{21} - x_{22} = 0, & \quad x_{22} - x_{23} = 0 \\ y_{21} - y_{22} = 0, & \quad y_{22} - y_{23} = 0 \\ x_{31} - x_{32} = 0, & \quad x_{32} - x_{33} = 0 \\ y_{31} - y_{32} = 0, & \quad y_{32} - y_{33} = 0. \end{aligned}$$

In der Form der allgemeinen Hypothese (425.10) erhält man

$$\mathbf{HBU} = \mathbf{W}$$

mit

$$\mathbf{H} = \begin{vmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots \end{vmatrix}$$

$$\mathbf{B} = \begin{vmatrix} x_{11} & x_{12} & x_{13} \\ y_{11} & y_{12} & y_{13} \\ x_{21} & x_{22} & x_{23} \\ y_{21} & y_{22} & y_{23} \\ x_{31} & x_{32} & x_{33} \\ y_{31} & y_{32} & y_{33} \\ \dots & \dots & \dots \end{vmatrix}$$

$$\mathbf{U} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{vmatrix}, \quad \mathbf{W} = \begin{vmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{vmatrix}.$$

Im allgemeinen stellt der Likelihood-Quotiententest nach (425.3) keinen gleichmäßig besten Test dar (HUMAK 1977, S.191). Mit den Eigenwerten der Matrix $\mathbf{R}\mathbf{\Omega}^{-1}$, denn $\mathbf{\Omega}$ kann wegen (412.5) als positiv definit angenommen

werden, gibt man daher zum Test der Hypothese (425.1) oder (425.10) weitere Teststatistiken an, denn nach (142.7) sind die Eigenwerte invariant gegenüber orthogonalen Transformationen und zum anderen gilt nach (142.6)

$$\mathbf{C}'\mathbf{R}\boldsymbol{\Omega}^{-1}\mathbf{C} = \boldsymbol{\Lambda} \quad \text{mit} \quad \mathbf{C}'\mathbf{C} = \mathbf{I} \quad \text{mit} \quad \boldsymbol{\Lambda} = \text{diag}(\lambda_i), \quad (425.13)$$

worin λ_i die Eigenwerte der Matrix $\mathbf{R}\boldsymbol{\Omega}^{-1}$ bezeichnen. Für das Likelihood-Quotientenkriterium (425.3) folgt dann mit (136.10) und (136.13)

$$\begin{aligned} \Lambda_{p,r,n-q} &= \frac{\det \boldsymbol{\Omega}}{\det(\boldsymbol{\Omega} + \mathbf{R})} = \frac{\det \mathbf{I}}{\det(\mathbf{I} + \mathbf{R}\boldsymbol{\Omega}^{-1})} \\ &= \frac{\det \mathbf{I}}{\det(\mathbf{C}'\mathbf{C} + \boldsymbol{\Lambda})} = \frac{\det \mathbf{I}}{\det(\mathbf{I} + \boldsymbol{\Lambda})} \end{aligned}$$

und mit (136.9)

$$\Lambda_{p,r,n-q} = \prod_{i=1}^p (1 + \lambda_i)^{-1}. \quad (425.14)$$

Mit der Summe der Eigenwerte $\sum_{i=1}^p \lambda_i = \text{sp}\boldsymbol{\Lambda} = \text{sp}(\mathbf{C}'\mathbf{R}\boldsymbol{\Omega}^{-1}\mathbf{C}) = \text{sp}(\mathbf{R}\boldsymbol{\Omega}^{-1})$ wegen (137.3) und (425.13) ergibt sich als Teststatistik das *Spurkriterium* von LAWLEY (1938) und HOTELLING (1951)

$$T_{p,r,n-q}^2 = \text{sp}(\mathbf{R}\boldsymbol{\Omega}^{-1}). \quad (425.15)$$

Die Verteilung dieser Testgröße erhält man aus (287.2), falls die Nullhypothese in (425.1) oder (425.10) richtig ist. Da neben $\boldsymbol{\Omega}$ auch \mathbf{R} wegen (412.7) als positiv definit anzunehmen ist, folgt mit $\mathbf{R} = \mathbf{G}\mathbf{G}'$ aus (143.5) $\text{sp}(\mathbf{R}\boldsymbol{\Omega}^{-1}) = \text{sp}(\mathbf{G}'\boldsymbol{\Omega}^{-1}\mathbf{G}) > 0$ wegen (143.6), so daß $T_{p,r,n-q}^2$ mit dem Anwachsen der Elemente von \mathbf{R} ebenfalls wächst. Die Nullhypothese in (425.1) oder (425.10) ist also abzulehnen, falls gilt

$$T_{p,r,n-q}^2 > T_{1-\alpha;p,r,n-q}^2, \quad (425.16)$$

wobei $T_{1-\alpha;p,r,n-q}^2$ das mit (287.4) zu bestimmende $(1-\alpha)$ -Fraktile der Verteilung für $T_{p,r,n-q}^2$ bezeichnet. Wie in (422.12) läßt sich auch die Wahrscheinlichkeit α_{T^2} berechnen

$$\alpha_{T^2} = \int_{T_{p,r,n-q}^2}^{\infty} f(T^2) dT^2, \quad (425.17)$$

falls $f(T^2)$ die in (287.2) angegebene Dichte bedeutet. Gilt für ein vorgegebenes Signifikanzniveau α

$$\alpha_{T^2} < \alpha, \quad (425.18)$$

ist die Nullhypothese abzulehnen. Die Wahrscheinlichkeit des Fehlers 2. Art läßt sich mit der in (MUIRHEAD 1972) angegebenen Verteilungsfunktion berechnen.

Als weitere Testgröße ist das *Maximalwurzel-Kriterium* von ROY (1957)

$$V = \lambda_{\max} \quad (425.19)$$

zu nennen. Diese Teststatistik und noch weitere (KRES 1983, S.5; ROY et al. 1971, S.73) werden hier nicht behandelt, denn numerische Vergleiche (ITO 1962; PILLAI und JAYACHANDRAN 1967; ROY et al. 1971, S.75), die allerdings auf kleine Werte für p , eine geringe Anzahl von Variationen für r und $n - q$ und kleine Abweichungen von der Nullhypothese beschränkt blieben, haben gezeigt, daß die Trennschärfen der Tests sich nicht wesentlich unterscheiden. Daher ist kein Testverfahren besonders zu bevorzugen, so daß bei der Auswahl die Möglichkeiten einer einfachen Berechnung der Testgrößen zu beachten sind, die für (425.3) und (425.15) gegeben sind.

2. Beispiel: Im Beispiel des Kapitels 374 soll jetzt geprüft werden, ob der Faktor A , also die ungünstige und günstige Geschäftslage überhaupt einen Einfluß auf die Preisgestaltung besitzt. Hierzu werden sowohl die beiden Hypothesen

$$H_0 : \alpha_{1i} = \alpha_{2i} = 0 \quad \text{gegen} \quad H_1 : \alpha_{1i} \neq 0, \alpha_{2i} \neq 0 \\ \text{für jedes } i \in \{1, 2\}$$

getrennt in den beiden univariaten Modellen, die durch $\mathbf{X}\boldsymbol{\beta}_i = E(\mathbf{y}_i)$ gegeben sind, als auch die multivariate Hypothese

$$H_0 : \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} \quad \text{gegen} \quad H_1 : \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} \neq \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix}$$

geprüft. Durch die beiden Hypothesen wird nach (334.17) jeweils nur eine linear unabhängige Restriktion eingeführt, da aus der Restriktion $\alpha_{1i} + \alpha_{2i} = 0$ zur Berechnung einer symmetrischen reflexiven generalisierten Inversen mit $\alpha_{1i} = 0$ bereits $\alpha_{2i} = 0$ folgt, so daß $r = 1$ gilt und die Hypothesen in den beiden univariaten Modellen auch durch

$$H_0 : \mathbf{H}\boldsymbol{\beta}_i = \mathbf{0} \quad \text{gegen} \quad H_1 : \mathbf{H}\boldsymbol{\beta}_i \neq \mathbf{0}$$

und im multivariaten Modell durch

$$H_0 : \mathbf{H}\mathbf{B} = \mathbf{0} \quad \text{gegen} \quad H_1 : \mathbf{H}\mathbf{B} \neq \mathbf{0}$$

mit $\mathbf{H} = |1, 0, 0|$ formuliert werden können.

Mit $\alpha_{1i} = \alpha_{2i} = 0$ entfallen in dem angegebenen Normalgleichungssystem des Beispiels die Spalten und Zeilen α_{1i}, α_{2i} und k_i , so daß durch einmalige Elimination nach (374.2) sich ergibt

$$\mathbf{\Omega}_H = \begin{vmatrix} 63,683 & 57,517 \\ & 54,558 \end{vmatrix}.$$

Diese Matrix läßt sich mit $\mathbf{H} = |1, 0, 0|$ auch nach (376.7) berechnen. Aus (422.4) folgen dann die Testgrößen T und aus (422.12) die Wahrscheinlichkeiten α_T zu

$$T = \frac{(63,683 - 1,667)/1}{1,667/4} = 148,8 \quad \text{mit} \quad \alpha_T = 0,000$$

für $H_0 : \alpha_{11} = \alpha_{21} = 0$

und

$$T = \frac{(54,558 - 2,291)/1}{2,291/4} = 91,26 \quad \text{mit} \quad \alpha_T = 0,001$$

für $H_0 : \alpha_{12} = \alpha_{22} = 0$.

Bei einem Signifikanzniveau von $\alpha = 0,05$ sind beide univariaten Hypothesen wegen (422.13) abzulehnen. Die gleiche Aussage ergibt sich auch aus (422.7), da $F_{0,95;1,4} = 7,71$ gilt.

Für die multivariate Hypothese $\mathbf{HB} = \mathbf{0}$ berechnen sich das Likelihood-Quotienten-Kriterium (425.3) und die Wahrscheinlichkeit α_Λ aus (425.5) zu

$$\Lambda_{2,1,4} = \frac{\det \mathbf{\Omega}}{\det \mathbf{\Omega}_H} = \frac{3,479}{166,21} = 0,0209 \quad \text{und} \quad \alpha_\Lambda = 0,003.$$

Für ein Signifikanzniveau von $\alpha = 0,05$ erhält man $\Lambda_{0,05;2,1,4} = 0,136$, so daß nach (425.4) und (425.6) die Nullhypothese abzulehnen ist. Mit dem Spurkriterium folgt aus (425.15) und (425.17)

$$T_{2,1,4}^2 = \text{sp}((\mathbf{\Omega}_H - \mathbf{\Omega})\mathbf{\Omega}^{-1}) = 46,80 \quad \text{und} \quad \alpha_{T^2} = 0,003$$

sowie $T_{0,95;2,1,4}^2 = 6,37$ aus (287.4), so daß wegen (425.16) und (425.18) die Nullhypothese ebenfalls abzulehnen ist. Das Datenmaterial läßt also Einflüsse der Geschäftslage auf die Preisgestaltung erkennen.

Die Schätzwerte $\hat{\alpha}_{11}$ und $\hat{\alpha}_{21}$ unterscheiden sich von $\hat{\alpha}_{12}$ und $\hat{\alpha}_{22}$ jeweils nur um DM 0,04, so daß zu prüfen ist, ob die Preise der Waren 1 und 2 überhaupt unterschiedliche Ergebnisse für die Einflüsse einer ungünstigen oder günstigen Geschäftslage rechtfertigen. Die entsprechende multivariate Hypothese lautet

$$H_0 : \alpha_{i1} - \alpha_{i2} = 0 \quad \text{für jedes} \quad i \in \{1, 2\}$$

gegen $H_1 : \alpha_{i1} - \alpha_{i2} \neq 0 \quad \text{für ein} \quad i \in \{1, 2\}.$

Sie folgt mit $\mathbf{H} = |1, 0, 0|$, $\mathbf{U}' = |1, -1|$ und $\mathbf{W} = 0$ aus (425.10), denn mit $\alpha_{11} = \alpha_{12}$ und $\alpha_{1i} + \alpha_{2i} = 0$ ergibt sich $\alpha_{21} = \alpha_{22}$. Aus (425.11) und (425.12) erhält man $\mathbf{\Omega} = 2,792$ sowie $\mathbf{R} = 0,415$ und daher

$$\Lambda_{1,1,4} = 0,871 \quad \text{und} \quad \alpha_{\Lambda} = 0,48.$$

Für ein Signifikanzniveau von $\alpha = 0,05$ gilt $\Lambda_{0,05;1,1,4} = 0,342$, so daß nach (425.4) und (425.6) die Nullhypothese zu akzeptieren ist. Mit dem Spurkriterium folgt

$$T_{1,1,4}^2 = 0,149 \quad \text{und} \quad \alpha_{T^2} = 0,48$$

sowie $T_{0,95;1,1,4}^2 = 1,93$ aus (287.4), so daß wegen (425.16) und (425.18) die Nullhypothese ebenfalls anzunehmen ist. Die Preise der Waren 1 und 2 rechtfertigen also keine unterschiedlichen Ergebnisse für die Effekte der ungünstigen und günstigen Geschäftslage.

426 Hypothese der Identität einer Kovarianzmatrix mit einer gegebenen Matrix

Sollen die Schätzwerte einer Kovarianzmatrix $\mathbf{\Sigma}$ durch eine vorgegebene, positiv definite Kovarianzmatrix $\mathbf{\Sigma}_0$ ersetzt werden, läßt sich die

$$\text{Hypothese } H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0 \quad \text{gegen} \quad H_1 : \mathbf{\Sigma} \neq \mathbf{\Sigma}_0 \quad (426.1)$$

testen. Ein Beispiel ist mit (375.4) gegeben, wo eine geschätzte Kovarianzmatrix mit Hilfe einer Kovarianzfunktion angenähert wird. Weiter läßt sich dieser Test anwenden, wenn geschätzte Kovarianzmatrizen durch Matrizen einfacherer Struktur, eventuell durch Diagonalmatrizen ersetzt werden sollen.

Zur Prüfung der Hypothese (426.1) wird eine Teststatistik in Analogie zum Likelihood-Quotiententest (421.3) abgeleitet. Wie für (425.2) ergibt sich bei normalverteilten Beobachtungen die Maximum-Likelihood-Schätzung für \mathbf{B} zu $\bar{\mathbf{B}}$ und für $\mathbf{\Sigma}$ zu $\bar{\mathbf{\Sigma}}$ und im Falle der Restriktionen $\mathbf{\Sigma} = \mathbf{\Sigma}_0$ die Schätzung für \mathbf{B} zu $\bar{\mathbf{B}}$ und für $\mathbf{\Sigma}$ zu $\mathbf{\Sigma}_0$. Daher folgt mit (372.7) die Testgröße $\lambda(\text{vec}\mathbf{Y})$ aus (421.3) zu

$$\lambda(\text{vec}\mathbf{Y}) = \frac{(2\pi)^{np/2} (\det \bar{\mathbf{\Sigma}})^{n/2} e^{np/2}}{(2\pi)^{np/2} (\det \mathbf{\Sigma}_0)^{n/2} e^{n \text{sp}(\mathbf{\Sigma}_0^{-1} \bar{\mathbf{\Sigma}})/2}},$$

für die wegen (421.3) $0 \leq \lambda(\text{vec}\mathbf{Y}) \leq 1$ mit $\lambda(\text{vec}\mathbf{Y}) = 1$ für $\bar{\mathbf{\Sigma}} = \mathbf{\Sigma}_0$ gilt. Anstelle der nicht erwartungstreuen Schätzung $\bar{\mathbf{\Sigma}}$ von $\mathbf{\Sigma}$ sei die erwartungstreue Schätzung $\hat{\mathbf{\Sigma}}$ eingeführt, für die nach (412.9) die Wishart-Verteilung

$(n - q)\hat{\Sigma} \sim W(n - q, \Sigma)$ gilt. Ersetzt man weiter n durch $n - q$, ergibt sich die Testgröße λ_1

$$\lambda_1 = (\det \hat{\Sigma})^{(n-q)/2} e^{(n-q)p/2} (\det \Sigma_0)^{-(n-q)/2} e^{-(n-q) \text{sp}(\hat{\Sigma}\Sigma_0^{-1})/2},$$

für die $0 \leq \lambda_1 \leq 1$ mit $\lambda_1 = 1$ für $\hat{\Sigma} = \Sigma_0$ gilt. Da Σ_0 nach Voraussetzung und $\hat{\Sigma}$ nach (412.5) und (412.9) positiv definit sind, folgt nämlich $\lambda_1 \geq 0$ wegen (143.2). Weiter ist $\lambda_1 \leq 1$ wegen $\det(\hat{\Sigma}\Sigma_0^{-1}) \exp(p - \text{sp}(\hat{\Sigma}\Sigma_0^{-1})) \leq 1$ oder mit $\Sigma_0^{-1} = \mathbf{G}\mathbf{G}'$ nach (143.5) $\det(\mathbf{G}'\hat{\Sigma}\mathbf{G}) \leq \exp(\text{sp}(\mathbf{G}'\hat{\Sigma}\mathbf{G}) - p)$, denn bezeichnet man mit l_1 bis l_p die Eigenwerte der wegen (143.7) positiv definiten Matrix $\mathbf{G}'\hat{\Sigma}\mathbf{G}$, ergibt sich mit (142.6) $\det(\mathbf{G}'\hat{\Sigma}\mathbf{G}) = l_1 \dots l_p \leq \exp((l_1 - 1) + \dots + (l_p - 1)) = \exp(l_1 - 1) \dots \exp(l_p - 1)$, da $l_i \leq \exp(l_i - 1)$ gilt, was aus (224.3) folgt. Mit $\lambda_{p,n-q} = -2 \ln \lambda_1$ erhält man dann die Teststatistik

$$\lambda_{p,n-q} = (n - q) [\ln(\det \Sigma_0 / \det \hat{\Sigma}) - p + \text{sp}(\hat{\Sigma}\Sigma_0^{-1})], \quad (426.2)$$

deren Verteilung wegen (412.9) im Falle der gültigen Nullhypothese $\Sigma = \Sigma_0$ mit (287.6) bestimmt ist. Für $\hat{\Sigma} = \Sigma_0$ ergibt sich $\lambda_{p,n-q} = 0$, und für $\hat{\Sigma} \neq \Sigma_0$ folgt $\lambda_{p,n-q} > 0$. Die Nullhypothese in (426.1) ist also abzulehnen, falls

$$\lambda_{p,n-q} > \lambda_{1-\alpha;p,n-q} \quad (426.3)$$

gilt, wobei $\lambda_{1-\alpha;p,n-q}$ das $(1 - \alpha)$ -Fraktile der Verteilung (287.6) bedeutet.

Die Hypothese (426.1) läßt sich auch für eine Kovarianzmatrix $\hat{\Sigma}$ prüfen, die nach ihrer Schätzung einer Transformation, beispielsweise $\hat{\Sigma}_A = \mathbf{A}\hat{\Sigma}\mathbf{A}'$, unterworfen wird. Aus $(n - q)\hat{\Sigma} \sim W(n - q, \Sigma)$ folgt nämlich mit (284.1) für $\hat{\Sigma}_A$ die Verteilung $(n - q)\hat{\Sigma}_A \sim W(n - q, \mathbf{A}\Sigma\mathbf{A}')$, sofern \mathbf{A} vollen Zeilenrang besitzt.

43 Bereichsschätzung

431 Konfidenzintervalle

Während bei der Punktschätzung Schätzwerte für die linearen Funktionen der Parameter und für die Parameter selbst bestimmt werden, sollen jetzt die Bereiche gesucht werden, in denen die Funktionen der unbekannt Parameter oder die Parameter selbst bei einer vorgegebenen Wahrscheinlichkeit liegen. Die Grenzen der Bereiche, ausgedrückt durch Intervalle, sind als Funktionen der Beobachtungen zu bestimmen.

Definition: Es seien $u(\mathbf{y})$ und $o(\mathbf{y})$ mit $u(\mathbf{y}) < o(\mathbf{y})$ lineare skalarwertige Funktionen der Beobachtungen \mathbf{y} und $g(\boldsymbol{\beta})$ eine lineare skalarwertige Funktion

der unbekanntem Parameter β , dann bezeichnet man das Intervall $(u(\mathbf{y}), o(\mathbf{y}))$, für das

$$P(u(\mathbf{y}) < g(\beta) < o(\mathbf{y})) = 1 - \alpha \quad \text{mit} \quad 0 < \alpha < 1$$

gilt, als *Konfidenzintervall* zum *Konfidenzniveau* $1 - \alpha$ und $u(\mathbf{y})$ sowie $o(\mathbf{y})$ als *Konfidenzgrenzen*. (431.1)

Statt Konfidenzintervall sagt man auch *Vertrauensintervall* und wählt wie beim Hypothesentest im allgemeinen $\alpha = 0,1$; $\alpha = 0,05$ oder $\alpha = 0,01$. Ein *gleichmäßig bestes Konfidenzintervall* besitzt unter allen Konfidenzintervallen die Eigenschaft, daß die Wahrscheinlichkeit, daß das Intervall falsche Werte für die Parameter β enthält, minimal ist. Gleichmäßig beste Konfidenzintervalle ergeben sich aus gleichmäßig besten Hypothesentests (HUMAK 1977, S.335; WITTING und NÖLLE 1970, S.37), die, wie bereits im Kapitel 421 erwähnt, in vielen Fällen mit dem Likelihood-Quotiententest (421.3) erhalten werden. Aufgrund dieser Hypothesentests werden daher im folgenden Kapitel die Konfidenzintervalle angegeben. Die enge Beziehung zwischen Hypothesentests und Konfidenzintervallen liegt darin begründet, daß durch einen Hypothesentest für die Parameter β zum Signifikanzniveau α , in den man als Nullhypothese sämtliche Werte von β im Parameterraum B einführt, ein Konfidenzbereich für β zum Konfidenzniveau $1 - \alpha$ erhalten wird.

432 Konfidenzintervalle für Parameter, für lineare Funktionen der Parameter und Konfidenzhyperellipsoide

Die Konfidenzintervalle für Parameter beziehen sich wieder wie im Kapitel 423 auf die erwartungstreu schätzbaren projizierten Parameter, sofern sie in einem Gauß-Markoff-Modell mit nicht vollem Rang definiert sind.

a) Konfidenzintervall für einen Parameter

Setzt man mit $\beta_b = (\mathbf{X}'\mathbf{X})_{rs}^{-1}\mathbf{X}'\mathbf{X}\beta = (\beta_i)$ in (423.2) $\beta_{oi} = \beta_i$, so folgt, sofern $\hat{\sigma}_i^2 \neq 0$ gilt, mit $\hat{\sigma}_i = (\hat{\sigma}_i^2)^{1/2}$ nach (423.4) $(\hat{\beta}_i - \beta_i)/\hat{\sigma}_i \sim t(n - q)$ und mit $t_{1-\alpha;n-q}$ aus (265.4) das Konfidenzintervall, das der Hypothese (423.1) entspricht,

$$P(-t_{1-\alpha;n-q} < (\hat{\beta}_i - \beta_i)/\hat{\sigma}_i < t_{1-\alpha;n-q}) = 1 - \alpha$$

oder

$$P(\hat{\sigma}_i t_{1-\alpha;n-q} > \beta_i - \hat{\beta}_i > -\hat{\sigma}_i t_{1-\alpha;n-q}) = 1 - \alpha .$$

Hiermit ergibt sich das Konfidenzintervall für den Parameter β_i zum Konfidenzniveau $1 - \alpha$ zu

$$P(\hat{\beta}_i - \hat{\sigma}_i t_{1-\alpha;n-q} < \beta_i < \hat{\beta}_i + \hat{\sigma}_i t_{1-\alpha;n-q}) = 1 - \alpha . \quad (432.1)$$

b) Konfidenzintervall für eine lineare Funktion der Parameter

Es sei $\mathbf{b}'\boldsymbol{\beta}$ eine schätzbare lineare Funktion der Parameter $\boldsymbol{\beta}$, dann erhält man mit $w_0 = \mathbf{b}'\boldsymbol{\beta}$ aus (423.7) $(\widehat{\mathbf{b}'\boldsymbol{\beta}} - \mathbf{b}'\boldsymbol{\beta})/[\widehat{V}(\widehat{\mathbf{b}'\boldsymbol{\beta}})]^{1/2} \sim t(n-q)$, so daß sich (432.1) entsprechend das Konfidenzintervall für $\mathbf{b}'\boldsymbol{\beta}$ zum Konfidenzniveau $1-\alpha$ ergibt, das der Hypothese (423.5) entspricht,

$$P(\widehat{\mathbf{b}'\boldsymbol{\beta}} - [\widehat{V}(\widehat{\mathbf{b}'\boldsymbol{\beta}})]^{1/2}t_{1-\alpha;n-q} < \mathbf{b}'\boldsymbol{\beta} < \widehat{\mathbf{b}'\boldsymbol{\beta}} + [\widehat{V}(\widehat{\mathbf{b}'\boldsymbol{\beta}})]^{1/2}t_{1-\alpha;n-q}) = 1 - \alpha . \quad (432.2)$$

Beispiel: Für das Beispiel des Kapitels 374 soll das Konfidenzintervall der linearen Funktion $S = \alpha_{11} + \mu_1$ der Parameter α_{11} und μ_1 bestimmt werden, so daß S den Preis der Waren 1 bei ungünstiger Geschäftslage angibt. Mit $\hat{S} = |1, 0, 1|\hat{\boldsymbol{\beta}}_1$ erhält man mit $(\mathbf{X}'\mathbf{X})_{rs}^-$ und $\hat{\sigma}_1^2$

$$\widehat{V}(\hat{S}) = 0,417 \cdot 0,00334 = 0,00139$$

und mit $t_{0,95;4} = 2,78$ sowie $\hat{S} = 1,48$ das Konfidenzintervall von S zum Niveau $1 - \alpha = 0,95$

$$P(1,38 < \alpha_{11} + \mu_1 < 1,58) = 0,95 .$$

Mit einer Wahrscheinlichkeit von 95 % liegt somit der Preis der Waren 1 bei ungünstiger Geschäftslage in dem angegebenen Intervall von DM 1,38 und DM 1,58.

c) Konfidenzbereich für mehrere Parameter

Substituiert man in (423.9) $\beta_{oi} = \beta_i$ mit $\boldsymbol{\beta}_{j\dots k} = \mathbf{H}\boldsymbol{\beta}_b = (\beta_i)$ für $i \in \{j, j+1, \dots, k\}$, erhält man

$$(\hat{\boldsymbol{\beta}}_{j\dots k} - \boldsymbol{\beta}_{j\dots k})'(\boldsymbol{\Sigma}_{\boldsymbol{\beta}})_{j\dots k}^{-1}(\hat{\boldsymbol{\beta}}_{j\dots k} - \boldsymbol{\beta}_{j\dots k})/[(k-j+1)\hat{\sigma}^2] \sim F(k-j+1, n-q) .$$

Setzt man nun mit (263.5)

$$(\hat{\boldsymbol{\beta}}_{j\dots k} - \boldsymbol{\beta}_{j\dots k})'(\boldsymbol{\Sigma}_{\boldsymbol{\beta}})_{j\dots k}^{-1}(\hat{\boldsymbol{\beta}}_{j\dots k} - \boldsymbol{\beta}_{j\dots k})/[(k-j+1)\hat{\sigma}^2] = F_{1-\alpha;k-j+1,n-q} , \quad (432.3)$$

stellt diese Gleichung den Konfidenzbereich für die Parameter β_j bis β_k zum Konfidenzniveau $1 - \alpha$ dar. Da die Matrix $(\boldsymbol{\Sigma}_{\boldsymbol{\beta}})_{j\dots k}^{-1}$ wegen (233.5) positiv definit ist, denn ihre Inverse wurde für (423.9) als regulär vorausgesetzt, besitzt

wegen der quadratischen Form auf der linken Seite von (432.3) der Konfidenzbereich die Gestalt eines Hyperellipsoides mit dem Mittelpunkt $\hat{\beta}_{j\dots k}$, so daß ein *Konfidenzhyperellipsoid* erhalten wird. Dies ergibt sich aus den folgenden Überlegungen.

Bezeichnet man mit x_i für $i \in \{j, j+1, \dots, k\}$ die Koordinaten eines im E^{k-j+1} orthogonalen Koordinatensystems, dessen Achsen in Richtung der $k-j+1$ Halbachsen c_i eines Hyperellipsoides zeigen und dessen Ursprung im Mittelpunkt des Hyperellipsoides liegt, so ist die Gleichung des Hyperellipsoides gegeben durch

$$\sum_{i=j}^k x_i^2 / c_i^2 = 1, \quad (432.4)$$

denn eine x_l, x_m -Koordinatenebene mit $l \neq m$ schneidet das Hyperellipsoid in der Ellipse $x_l^2 / c_l^2 + x_m^2 / c_m^2 = 1$. Die Gleichung (432.3) läßt sich auf die Form (432.4) bringen, indem die Matrix $(\Sigma_\beta)_{j\dots k}^{-1}$ mit der orthogonalen Matrix \mathbf{C} ihrer Eigenvektoren mit $\mathbf{C}'\mathbf{C} = \mathbf{I}$ nach (142.6) durch $\mathbf{C}'(\Sigma_\beta)_{j\dots k}^{-1}\mathbf{C} = \mathbf{\Lambda}^{-1}$ in die Diagonalmatrix $\mathbf{\Lambda}^{-1}$ ihrer Eigenwerte überführt wird. Setzt man weiter

$$\mathbf{x} = \mathbf{C}'(\hat{\beta}_{j\dots k} - \beta_{j\dots k}), \quad (432.5)$$

gilt $(\hat{\beta}_{j\dots k} - \beta_{j\dots k})'(\Sigma_\beta)_{j\dots k}^{-1}(\hat{\beta}_{j\dots k} - \beta_{j\dots k}) = \mathbf{x}'\mathbf{\Lambda}^{-1}\mathbf{x}$, so daß eine (432.4) entsprechende Darstellung erhalten wird. Die Matrix \mathbf{C} der Eigenvektoren enthält nach (141.4) die Richtungskosinus zwischen dem orthogonalen Koordinatensystem, in dem die Vektoren $\hat{\beta}_{j\dots k}$ und $\beta_{j\dots k}$ dargestellt sind, und dem orthogonalen (x_j, \dots, x_k) -Koordinatensystem. Die Inverse der Matrix $\mathbf{\Lambda}^{-1}$ führt zu den Halbachsen c_i , so daß man mit $(\mathbf{\Lambda}^{-1})^{-1} = \mathbf{C}'(\Sigma_\beta)_{j\dots k}\mathbf{C}$ wegen (131.14) und (141.2) zweckmäßig $(\Sigma_\beta)_{j\dots k}$ diagonalisiert,

$$\mathbf{C}'(\Sigma_\beta)_{j\dots k}\mathbf{C} = \mathbf{\Lambda} \quad \text{mit} \quad \mathbf{\Lambda} = \text{diag}(\lambda_j, \lambda_{j+1}, \dots, \lambda_k). \quad (432.6)$$

Somit ergeben sich die Halbachsen c_i des Konfidenzhyperellipsoides (432.3) zu

$$c_i = (\hat{\sigma}^2(k-j+1)\lambda_i F_{1-\alpha; k-j+1, n-q})^{1/2}. \quad (432.7)$$

Mit $(k-j+1)F_{1-\alpha; k-j+1, n-q} = 1$ erhält man das sogenannte *Fehlerhyperellipsoid*, das in der Ausgleichsrechnung häufig berechnet wird. Bezeichnet man nämlich mit $\bar{\beta}_{j\dots k}$ einen $(k-j+1) \times 1$ Vektor von Konstanten und fragt nach der mit $\hat{\sigma}^2$ berechneten Varianz $\hat{\sigma}_s^2$ der Länge s des Differenzvektors $\hat{\beta}_{j\dots k} - \bar{\beta}_{j\dots k}$, ergibt sich mit (123.3)

$$s = ((\hat{\beta}_j - \bar{\beta}_j)^2 + (\hat{\beta}_{j+1} - \bar{\beta}_{j+1})^2 + \dots + (\hat{\beta}_k - \bar{\beta}_k)^2)^{1/2}$$

und mit (233.2) und (233.4)

$$\hat{\sigma}_s^2 = \hat{\sigma}^2 \mathbf{z}' (\boldsymbol{\Sigma}_\beta)_{j\dots k} \mathbf{z} \quad (432.8)$$

mit $\mathbf{z}' = \left| (\hat{\beta}_j - \bar{\beta}_j)/s|_0, \dots, (\hat{\beta}_k - \bar{\beta}_k)/s|_0 \right|$ und

$$\mathbf{z}' \mathbf{z} = 1. \quad (432.9)$$

Sucht man nach den Extremwerten von $\hat{\sigma}_s^2$, so sind die Extrema der quadratischen Form (432.8) unter der Bedingung (432.9) zu finden. Nach (142.2) ergeben sie sich aus den Eigenwerten und Eigenvektoren von $(\boldsymbol{\Sigma}_\beta)_{j\dots k}$, also aus (432.6). Folglich berechnen sich aus $\hat{\sigma}^2 \lambda_i$ die Maximalwerte der Varianz $\hat{\sigma}_s^2$ der Länge s und aus \mathbf{C} ihre Richtungskosinus. Die Verbindungshyperfläche der Maximalwerte bildet das Fehlerhyperellipsoid.

Beispiel: Es gelte

$$\boldsymbol{\beta}_{j\dots k} = \begin{vmatrix} \beta_j \\ \beta_k \end{vmatrix}, \quad \hat{\boldsymbol{\beta}}_{j\dots k} = \begin{vmatrix} \hat{\beta}_j \\ \hat{\beta}_k \end{vmatrix} \quad \text{und} \quad (\boldsymbol{\Sigma}_\beta)_{j\dots k} = \begin{vmatrix} \sigma_{jj} & \sigma_{jk} \\ \sigma_{jk} & \sigma_{kk} \end{vmatrix},$$

und die Koordinaten der Vektoren $\boldsymbol{\beta}_{j\dots k}$ und $\hat{\boldsymbol{\beta}}_{j\dots k}$ seien in dem zweidimensionalen, orthogonalen (z_1, z_2) -Koordinatensystem definiert. Gesucht ist die durch (432.3) definierte Konfidenzellipse zum Konfidenzniveau $1 - \alpha$, bestimmt durch ihre Achsen, die in Richtung des orthogonalen (x_1, x_2) -Koordinatensystems zeigen sollen, und durch den Winkel zwischen der x_1 - und z_1 -Achse, der sich zwischen der x_2 - und z_2 -Achse wiederholt.

Die Eigenwerte der Matrix $(\boldsymbol{\Sigma}_\beta)_{j\dots k}$ ergeben sich mit (142.3) aus

$$\det \begin{vmatrix} \sigma_{jj} - \lambda & \sigma_{jk} \\ \sigma_{jk} & \sigma_{kk} - \lambda \end{vmatrix} = 0,$$

woraus mit (136.5) $\lambda^2 - \lambda(\sigma_{jj} + \sigma_{kk}) + \sigma_{jj}\sigma_{kk} - \sigma_{jk}^2 = 0$ oder

$$\lambda_{j,k} = \frac{1}{2}(\sigma_{jj} + \sigma_{kk}) \pm \left[\frac{1}{4}(\sigma_{jj} + \sigma_{kk})^2 - \sigma_{jj}\sigma_{kk} + \sigma_{jk}^2 \right]^{1/2}$$

folgt, so daß sich die Eigenwerte λ_j und λ_k ergeben mit

$$\lambda_{j,k} = \frac{1}{2} \left(\sigma_{jj} + \sigma_{kk} \pm \left[(\sigma_{jj} - \sigma_{kk})^2 + 4\sigma_{jk}^2 \right]^{1/2} \right). \quad (432.10)$$

Der zu λ_j gehörende Eigenvektor \mathbf{c}_j mit $\mathbf{c}_j = |c_{1j}, c_{2j}|'$ ist nach (142.2) definiert durch $((\boldsymbol{\Sigma}_\beta)_{j\dots k} - \lambda_j \mathbf{I}) \mathbf{c}_j = \mathbf{0}$ oder

$$\begin{aligned} c_{1j}(\sigma_{jj} - \lambda_j) + c_{2j}\sigma_{jk} &= 0 \\ c_{1j}\sigma_{jk} + c_{2j}(\sigma_{kk} - \lambda_j) &= 0. \end{aligned} \quad (432.11)$$

Mit dem zu λ_k gehörenden Eigenvektor \mathbf{c}_k gilt $\mathbf{C} = |\mathbf{c}_j, \mathbf{c}_k|$ für die Matrix \mathbf{C} der Eigenvektoren von $(\boldsymbol{\Sigma}_\beta)_{j\dots k}$. Die erste Zeile von \mathbf{C}' , die nach (432.5) für die Koordinatentransformation vom (z_1, z_2) -System ins (x_1, x_2) -System benötigt wird, enthält wegen (141.4) und (141.5) die Elemente $\cos \theta$ und $\sin \theta$, falls θ den Winkel im Gegenuhrzeigersinn zwischen der z_1 - und x_1 -Achse bezeichnet. Aus (432.11) folgt dann

$$\tan \theta = \frac{c_{2j}}{c_{1j}} = \frac{\lambda_j - \sigma_{jj}}{\sigma_{jk}} = \frac{\sigma_{jk}}{\lambda_j - \sigma_{kk}}.$$

Mit $\tan 2\theta = 2 \sin \theta \cos \theta / (\cos^2 \theta - \sin^2 \theta) = 2 \tan \theta / (1 - \tan^2 \theta)$ folgt weiter

$$\begin{aligned} \tan 2\theta &= \frac{2\sigma_{jk}(\lambda_j - \sigma_{kk})}{(\lambda_j - \sigma_{kk})^2 - \sigma_{jk}^2} \\ &= \frac{2\sigma_{jk}(\sigma_{jj} - \sigma_{kk} + [(\sigma_{jj} - \sigma_{kk})^2 + 4\sigma_{jk}^2]^{1/2})}{(\sigma_{jj} - \sigma_{kk})^2 + (\sigma_{jj} - \sigma_{kk})[(\sigma_{jj} - \sigma_{kk})^2 + 4\sigma_{jk}^2]^{1/2}}, \end{aligned}$$

so daß sich schließlich ergibt

$$\tan 2\theta = 2\sigma_{jk} / (\sigma_{jj} - \sigma_{kk}). \quad (432.12)$$

Hiermit wird der Winkel θ zwischen der z_1 -Achse und der großen Halbachse c_j der Konfidenzellipse berechnet, wobei für c_j wegen (432.7) gilt

$$c_j = (2\hat{\sigma}^2 \lambda_j F_{1-\alpha; 2, n-q})^{1/2} \quad (432.13)$$

mit $F_{1-\alpha; 2, n-q}$ aus (263.7) und λ_j aus (432.10). Der Winkel zwischen der z_1 -Achse und der aus λ_k folgenden kleinen Halbachse beträgt $\theta + 100^\circ$.

d) Konfidenzbereich für mehrere lineare Funktionen der Parameter

Der Konfidenzbereich der linearen Funktionen $\mathbf{H}\boldsymbol{\beta} - \mathbf{w}$ der Parameter $\boldsymbol{\beta}$, der der allgemeinen Hypothese (422.1) entspricht, ergibt sich mit (422.4) und (422.5) zu

$$(\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w})'(\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}')^{-1}(\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}) / (r\hat{\sigma}^2) = F_{1-\alpha; r, n-q}. \quad (432.14)$$

Wie man den Erläuterungen zu (432.3) entnehmen kann, besitzt (432.14) die Gestalt eines Hyperellipsoides mit dem Mittelpunkt $\mathbf{H}\bar{\boldsymbol{\beta}}$, falls der Vektor \mathbf{w} als variabel angesehen wird. Entsprechend (432.7) ergeben sich die Halbachsen des Hyperellipsoides (432.14) mit den Eigenwerten der Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'$ und die Richtungskosinus der Halbachsen aus der Matrix ihrer Eigenvektoren.

Das Hyperellipsoid (432.14) läßt sich in der Weise deuten, daß es die Grenze

zwischen den Werten für $\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}$ angibt, die zur Annahme oder zur Ablehnung der Nullhypothese in (422.1) führen. Innerhalb des Hyperellipsoides liegen die Differenzen $\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}$, für die die Hypothese akzeptiert, und außerhalb die Werte, für die die Hypothese abgelehnt wird. Das Konfidenzhyperellipsoid (432.14) kann man also dafür verwenden, Hypothesentests graphisch vorzunehmen. Da die Halbachsen des Hyperellipsoides, abgesehen von $\hat{\sigma}^2$, nicht von den Beobachtungen, sondern lediglich von den Eigenwerten der Matrix $\mathbf{H}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{H}'$ abhängen, kann man sich bereits bei der Planung eines Experimentes darüber Gewißheit verschaffen, welche Größenordnungen die Differenzen $\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}$ erreichen müssen, um zu einer Ablehnung der Nullhypothese zu führen. Als Beispiel seien die durch eine Deformationsanalyse aufzudeckenden Punktverschiebungen genannt (KOCH 1985A).

Es soll nun angenommen werden, daß die Alternativhypothese $\mathbf{H}\boldsymbol{\beta} = \bar{\mathbf{w}}$ in (422.1) gültig ist und daß der Differenzvektor $\bar{\mathbf{w}} - \mathbf{w}$ durch den mit (432.14) definierten Vektor $\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}$ bestimmt ist. Aus dem Vergleich des Nichtzentralitätsparameters λ in (422.6) mit (432.14) folgt dann, falls $\sigma^2 = \hat{\sigma}^2$ gilt,

$$\lambda = rF_{1-\alpha; r, n-q} . \quad (432.15)$$

Mit λ läßt sich aus (422.11) die Trennschärfe β berechnen, das heißt die Wahrscheinlichkeit, mit der die durch (432.14) definierten Differenzen $\mathbf{H}\bar{\boldsymbol{\beta}} - \mathbf{w}$ bestimmt werden.

433 Konfidenzintervall für die Varianz der Gewichtseinheit

Mit dem Hypothesentest (424.6) erhält man

$$P(\chi_{\alpha/2; n-q}^2 < (n-q)\hat{\sigma}^2/\sigma_0^2 < \chi_{1-\alpha/2; n-q}^2) = 1 - \alpha ,$$

woraus mit $\sigma_0^2 = \sigma^2$ das Konfidenzintervall für die Varianz σ^2 der Gewichtseinheit zum Konfidenzniveau $1 - \alpha$ folgt

$$P\left(\frac{(n-q)\hat{\sigma}^2}{\chi_{1-\alpha/2; n-q}^2} < \sigma^2 < \frac{(n-q)\hat{\sigma}^2}{\chi_{\alpha/2; n-q}^2}\right) = 1 - \alpha \quad (433.1)$$

oder aufgrund der gleichen Überlegungen, die nach (265.3) führen, das Konfidenzintervall für die Standardabweichung σ

$$P\left(\left[\frac{(n-q)\hat{\sigma}^2}{\chi_{1-\alpha/2; n-q}^2}\right]^{1/2} < \sigma < \left[\frac{(n-q)\hat{\sigma}^2}{\chi_{\alpha/2; n-q}^2}\right]^{1/2}\right) = 1 - \alpha . \quad (433.2)$$

Da die χ^2 -Verteilung nicht symmetrisch ist, besitzen die durch (433.1) und (433.2) definierten Konfidenzintervalle keine minimale Länge, was aber für

praktische Anwendungen unerheblich ist. Die Bedingungen für ein minimales Konfidenzintervall sind bei (MOOD et al. 1974, S.383) angegeben.

Beispiel: Im Beispiel des Kapitels 374 soll das Konfidenzintervall für die Standardabweichung σ_1 zum Konfidenzniveau $1-\alpha = 0,95$ angegeben werden. Mit $(\hat{\sigma}_1^2)^{1/2} = 0,646$, $\chi_{0,975;4}^2 = 11,14$ und $\chi_{0,025;4}^2 = 0,484$ erhält man

$$P(0,39 < \sigma_1 < 1,86) = 0,95.$$

44 Ausreißertest

441 Hypothesentest für Ausreißer

Wie bereits in den Kapiteln 345 und 381 erwähnt, können aus irgendwelchen Gründen Beobachtungen durch Ausreißer grob verfälscht sein. Diese Beobachtungen müssen entdeckt und ausgeschieden werden, damit sie die Parameterschätzung nicht beeinflussen. Mit Hilfe eines Hypothesentests, eines *Ausreißertests*, soll dies geschehen, der hier als Spezialfall des Tests einer allgemeinen Hypothese im univariaten Modell abgeleitet wird (KOCH 1983A). Ausreißertests im univariaten Modell sind bei (BARNETT und LEWIS 1978, S.208; HAWKINS 1980, S.104) angegeben. Die Suche nach Ausreißern, die auf robusten Schätzungen beruhen, wird im Kapitel 38 behandelt.

Das Modell (345.7) sieht Ausreißer ϵ in den Beobachtungen \mathbf{y} vor. Einen Ausreißertest erhält man daher, wenn in diesem Modell die Nullhypothese H_0 , daß alle Ausreißer gleich Null sind, gegen die Alternativhypothese H_1 , daß Ausreißer der Größe $\bar{\epsilon}$ vorliegen, getestet wird, also

$$H_0 : \epsilon = \mathbf{0} \quad \text{gegen} \quad H_1 : \epsilon = \bar{\epsilon} \neq \mathbf{0}$$

oder (422.1) entsprechend als lineare Hypothese für die projizierten Parameter in (345.2) formuliert

$$H_0 : \left| \begin{array}{c} \beta_b \\ \epsilon \end{array} \right| = \mathbf{0} \quad \text{gegen} \quad H_1 : \left| \begin{array}{c} \beta_b \\ \epsilon \end{array} \right| = \left| \begin{array}{c} \beta_b \\ \bar{\epsilon} \end{array} \right| \neq \mathbf{0}. \quad (441.1)$$

Für die Testgröße T dieser Hypothese folgt aus (422.4)

$$T = \frac{R/r}{\Omega_\epsilon/(n-q-r)} \quad (441.2)$$

mit

$$R = \left(\begin{array}{c|c} \mathbf{0}, \mathbf{I} & \begin{array}{c} \hat{\beta}_b \\ \hat{\epsilon} \end{array} \end{array} \right)' \left(\begin{array}{c|c} \mathbf{0}, \mathbf{I} & \begin{array}{c} \mathbf{X}'\mathbf{P}\mathbf{X} & \mathbf{X}'\mathbf{P}\mathbf{Z} \\ \mathbf{Z}'\mathbf{P}\mathbf{X} & \mathbf{Z}'\mathbf{P}\mathbf{Z} \end{array} \left[\begin{array}{c} \mathbf{0} \\ \mathbf{I} \end{array} \right]_{rs} \end{array} \right)^{-1} \\ \left(\begin{array}{c|c} \mathbf{0}, \mathbf{I} & \begin{array}{c} \hat{\beta}_b \\ \hat{\epsilon} \end{array} \end{array} \right)$$

oder wegen (345.11) und (345.12)

$$R = \hat{\epsilon}' \mathbf{P}\mathbf{Z} (\mathbf{Z}' \mathbf{P} \mathbf{Q}_e \mathbf{P}\mathbf{Z})^{-1} \mathbf{Z}' \mathbf{P} \hat{\epsilon}. \quad (441.3)$$

Die quadratische Form Ω_ϵ der Residuen im Modell (345.7) ergibt sich mit Ω aus (345.6) zu

$$\Omega_\epsilon = \Omega - R,$$

denn wird die Nullhypothese H_0 in (441.1) als Restriktion in das Modell (345.7) eingeführt, erhält man das Modell (345.1) und daher $\Omega = \Omega_\epsilon + R$. Ist die Alternativhypothese H_1 in (441.1) richtig, besitzt T nach (422.6) die nichtzentrale F -Verteilung

$$T \sim F'(r, n - q - r, \lambda) \quad (441.4)$$

mit dem Nichtzentralitätsparameter

$$\lambda = \frac{1}{\sigma_\epsilon^2} \left(\begin{array}{c|c} \mathbf{0}, \mathbf{I} & \begin{array}{c} \beta_b \\ \bar{\epsilon} \end{array} \end{array} \right)' \left(\begin{array}{c|c} \mathbf{0}, \mathbf{I} & \begin{array}{c} \mathbf{X}'\mathbf{P}\mathbf{X} & \mathbf{X}'\mathbf{P}\mathbf{Z} \\ \mathbf{Z}'\mathbf{P}\mathbf{X} & \mathbf{Z}'\mathbf{P}\mathbf{Z} \end{array} \left[\begin{array}{c} \mathbf{0} \\ \mathbf{I} \end{array} \right]_{rs} \end{array} \right)^{-1} \\ \left(\begin{array}{c|c} \mathbf{0}, \mathbf{I} & \begin{array}{c} \beta_b \\ \bar{\epsilon} \end{array} \end{array} \right)$$

oder mit (345.11)

$$\lambda = \bar{\epsilon}' \mathbf{Z}' \mathbf{P} \mathbf{Q}_e \mathbf{P}\mathbf{Z} \bar{\epsilon} / \sigma_\epsilon^2.$$

Die Testgröße T soll jetzt in die Testgröße τ^2 transformiert werden

$$\tau^2 = \frac{(n - q)T}{n - q - r + rT} \quad (441.5)$$

oder mit (441.2) und der positiven Quadratwurzel

$$\tau = \left(\frac{R/r}{\Omega/(n - q)} \right)^{1/2}. \quad (441.6)$$

Die Transformation (441.5) bewirkt also, daß in (441.6) Ω und nicht mehr Ω_ϵ wie in (441.2) eingeht. Als Verteilung für die Testgröße τ gilt analog zu (441.4)

$$\tau \sim \tau'(r, n - q - r, \lambda), \quad (441.7)$$

worin $\tau'(r, n - q - r, \lambda)$ die nichtzentrale Form der τ -Verteilung (POPE 1976) bezeichnet. Um Fraktile für die τ -Verteilung zu erhalten, wird (441.5) nach T aufgelöst, und man erhält mit (441.4)

$$T = \frac{(n - q - r)\tau^2}{n - q - r\tau^2} \sim F'(r, n - q - r, \lambda). \quad (441.8)$$

Den Überlegungen entsprechend, die zu (286.9) führen, ergibt sich das α -Fraktile $\tau'_{\alpha; r, n - q - r, \lambda}$ der nichtzentralen τ -Verteilung aus dem α -Fraktile $F'_{\alpha; r, n - q - r, \lambda}$ der nichtzentralen F -Verteilung zu

$$\tau'_{\alpha; r, n - q - r, \lambda} = \left(\frac{(n - q)F'_{\alpha; r, n - q - r, \lambda}}{n - q - r + rF'_{\alpha; r, n - q - r, \lambda}} \right)^{1/2}. \quad (441.9)$$

Mit der Testgröße (441.6) und ihrer Verteilung (441.7) lassen sich nun die Tests für Ausreißer in beliebig vielen Beobachtungen ableiten. Ein Testverfahren für Ausreißer in Beobachtungspaaren befindet sich bei (KOCH 1985B). Im folgenden soll der wichtigste Fall der Ausreißertests behandelt werden, daß nach einem Ausreißer in einer Beobachtung gesucht wird.

442 Test für einen Ausreißer

Die allgemeine Hypothese (441.1) wird jetzt derart spezialisiert, daß sie nur einen Ausreißer Δy_k in der Beobachtung y_k enthält, also

$$H_0 : \Delta y_k = 0 \quad \text{gegen} \quad H_1 : \Delta y_k = \Delta \bar{y}_k \neq 0. \quad (442.1)$$

Dann gilt $r = 1$ und (345.8), womit für (441.3) folgt

$$R = \hat{e}' P e_k (e_k' P Q_e P e_k)^{-1} e_k' P \hat{e}.$$

Es wird nun das transformierte Residuum e_k definiert und die Größe q_k^2 , die zur Berechnung der Varianz von e_k benötigt wird, mit

$$e_k = e_k' P \hat{e} \quad \text{und} \quad q_k^2 = e_k' P Q_e P e_k, \quad (442.2)$$

denn aufgrund von (233.2) erhält man mit (345.5)

$$V(e_k) = \sigma^2 q_k^2. \quad (442.3)$$

Mit der Schätzung $\hat{\sigma}^2$ von σ^2 im Modell (345.1) nach (331.12)

$$\hat{\sigma}^2 = \Omega / (n - q)$$

ergibt sich dann anstelle von (441.6) und (441.7) die Testgröße τ_k

$$\tau_k = e_k / (\hat{\sigma} q_k) \quad (442.4)$$

und ihre Verteilung im Falle der gültigen Alternativhypothese in (442.1)

$$\tau_k \sim \tau'(1, n - q - 1, \lambda) \quad \text{mit} \quad \lambda = \Delta \bar{y}_k^2 q_k^2 / \sigma_\epsilon^2. \quad (442.5)$$

Die Testgröße τ_k für die Hypothese (442.1), daß der Ausreißer Δy_k der Beobachtung y_k gleich Null ist, berechnet sich also aus dem *studentisierten* transformierten Residuum, das heißt aus e_k , das durch seine Standardabweichung dividiert wird, die mit Hilfe der geschätzten Varianz $\hat{\sigma}^2$ der Gewichtseinheit berechnet wird.

Die Verteilung (442.5) gilt, wie bereits erwähnt, für den Fall der richtigen Alternativhypothese. Ist die Nullhypothese $\Delta y_k = 0$ in (442.1) korrekt, folgt $\lambda = 0$ und damit für τ_k die zentrale τ -Verteilung

$$\tau_k \sim \tau(1, n - q - 1). \quad (442.6)$$

Die Berechnung der Testgröße τ_k vereinfacht sich, falls die Gewichtsmatrix \mathbf{P} im Modell (345.7) Diagonalform besitzt, also

$$\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_n). \quad (442.7)$$

Dann erhält man mit $\hat{e} = (\hat{e}_i)$ für e_k und mit $\mathbf{Q}_e = (q_{eij})$ für q_k^2

$$e_k = p_k \hat{e}_k \quad \text{und} \quad q_k^2 = p_k^2 q_{ekkk}. \quad (442.8)$$

Für das k -te Diagonalelement r_k der Matrix $\mathbf{Q}_e \mathbf{P}$ gilt wegen (442.7)

$$r_k = p_k q_{ekkk}. \quad (442.9)$$

Wie bereits im Zusammenhang mit (385.11) erwähnt, gibt r_k die Teilredundanz an, also den Anteil der k -ten Beobachtung an der Gesamtredundanz $n - q$ der Beobachtungen, denn mit (345.5) erhält man

$$\begin{aligned} \text{sp}(\mathbf{Q}_e \mathbf{P}) &= \text{sp}(\mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{P} \mathbf{X})_{rs}^- \mathbf{X}' \mathbf{P}) \\ &= n - \text{rg}((\mathbf{X}' \mathbf{P} \mathbf{X})_{rs}^- \mathbf{X}' \mathbf{P} \mathbf{X}) = n - q \end{aligned}$$

wegen (152.3) und (153.4). Mit (442.8) und (442.9) folgt dann anstelle von (442.4)

$$\tau_k = \hat{e}_k / (\hat{\sigma} \sqrt{q_{ekkk}}) = \hat{e}_k / [(\hat{\sigma} / \sqrt{p_k}) \sqrt{r_k}]. \quad (442.10)$$

Die Testgröße τ_k ergibt sich nun aus dem studentisierten Residuum von y_k . Im zweiten Ausdruck gibt $\hat{\sigma}/\sqrt{p_k}$ wegen $D(\mathbf{y}) = \sigma^2 \mathbf{P}^{-1}$ die mit $\hat{\sigma}$ berechnete Standardabweichung der Beobachtung y_k an.

Wird das Modell (345.7) im Falle einer diagonalen Gewichtsmatrix durch die Transformation (321.4) in ein Modell mit $\sigma_e^2 \mathbf{I}$ als Kovarianzmatrix der Beobachtungen überführt, folgt der Vektor \hat{e}_h der homogenisierten Residuen mit

$$\hat{e}_h = \text{diag}(\sqrt{p_1}, \sqrt{p_2}, \dots, \sqrt{p_n}) \hat{e}$$

und mit $\hat{e}_{hk} = (\hat{e}_{hi})$ das k -te homogenisierte Residuum \hat{e}_{hk}

$$\hat{e}_{hk} = \sqrt{p_k} \hat{e}_k .$$

Bezeichnet man seine Varianz mit

$$V(\hat{e}_{hk}) = \sigma^2 q_{hkk} ,$$

ergibt sich mit (233.2) und (345.5)

$$q_{hkk} = p_k q_{ekk} = r_k \quad (442.11)$$

wegen (442.9). Anstelle von (442.10) ergibt sich dann als Testgröße das studentisierte homogenisierte Residuum

$$\tau_k = \hat{e}_{hk} / (\hat{\sigma} \sqrt{q_{hkk}}) = \hat{e}_{hk} / (\hat{\sigma} \sqrt{r_k}) . \quad (442.12)$$

443 Fraktilwerte für den Ausreißertest

Mit der Testgröße τ_k aus (442.4), (442.10) oder (442.12) und ihrer Verteilung läßt sich nun die Hypothese testen, daß ein Ausreißer in der Beobachtung y_k enthalten ist. Für einen Ausreißertest möchte man aber nicht nur die Beobachtung y_k , sondern alle Beobachtungen von y_1 bis y_n prüfen. Bezeichnet man mit α das Signifikanzniveau des Tests der Nullhypothese, daß kein Ausreißer in den Beobachtungen y_1 bis y_n vorliegt, und mit $c_{1-\alpha;n,n-q}$ das $(1-\alpha)$ -Fraktil der Verteilung der Testgröße, dann gilt für die Wahrscheinlichkeit des Fehlers 1. Art des Tests

$$P(\tau_1 > c_{1-\alpha;n,n-q} \cup \tau_2 > c_{1-\alpha;n,n-q} \cup \dots \cup \tau_n > c_{1-\alpha;n,n-q}) = \alpha . \quad (443.1)$$

Für beliebiges l und m mit $l \neq m$ ergibt sich mit (213.7)

$$\begin{aligned} P(\tau_l > c_{1-\alpha;n,n-q} \cup \tau_m > c_{1-\alpha;n,n-q}) &= P(\tau_l > c_{1-\alpha;n,n-q}) \\ &+ P(\tau_m > c_{1-\alpha;n,n-q}) - P(\tau_l > c_{1-\alpha;n,n-q} \cap \tau_m > c_{1-\alpha;n,n-q}) . \end{aligned}$$

Wendet man diese Gleichung rekursiv auf (443.1) an und vernachlässigt die Wahrscheinlichkeiten der Schnittmengen, die klein sind, erhält man

$$P(\tau_1 > c_{1-\alpha;n,n-q}) + P(\tau_2 > c_{1-\alpha;n,n-q}) \\ + \dots + P(\tau_n > c_{1-\alpha;n,n-q}) \geq \alpha$$

oder

$$nP(\tau_i > c_{1-\alpha;n,n-q}) \geq \alpha \quad \text{und} \quad P(\tau_i > c_{1-\alpha;n,n-q}) \geq \alpha/n$$

und näherungsweise

$$P(\tau_i > c_{1-\alpha;n,n-q}) = \alpha/n$$

oder

$$P(\tau_i < c_{1-\alpha;n,n-q}) = 1 - \alpha/n \quad (443.2)$$

mit

$$c_{1-\alpha;n,n-q} = \tau_{1-\alpha/n;1,n-q-1} \cdot \quad (443.3)$$

Die einzelne Beobachtung y_k muß also mit einem Signifikanzniveau von α/n getestet werden, damit α als Signifikanzniveau für den Test aller Beobachtungen erhalten wird. Mit diesem Test werden weniger Ausreißer entdeckt, als wenn die Ungleichung angewendet würde, denn der Fraktilwert (443.3) ist größer als der Fraktilwert, der aus der Ungleichung resultiert.

Bei einem Ausreißertest ist somit für jede Beobachtung y_k mit $k \in \{1, \dots, n\}$ die Testgröße τ_k nach (442.4), (442.10) oder (442.12) zu bilden. Die Beobachtung y_k enthält einen Ausreißer, falls

$$\tau_k > c_{1-\alpha;n,n-q} \quad (443.4)$$

gilt. Die Fraktilwerte ergeben sich mit (443.3) aus (441.9). Tafeln für $c_{1-\alpha;n,n-q}$ findet man bei (LUND 1975; BARNETT und LEWIS 1978, S.335). Sehr umfangreiche Tafeln, die auf dem Signifikanzniveau $(1 - \alpha)^{1/n} \approx 1 - \alpha/n$ basieren, sind bei (POPE 1976) abgedruckt.

Ein ähnlicher Ausreißertest, der nicht auf studentisierten, sondern standardisierten Residuen beruht, Residuen also, die durch ihre Standardabweichungen dividiert werden, wurde von BAARDA (1968) entwickelt, siehe auch (BENNING 1982; FÖRSTNER 1979A; GRÜN 1978; KOCH 1983A; KOK 1984).

Beispiel: Im Beispiel des Kapitels 374 erhält man die homogenisierten Residuenvektoren \hat{e}_1 und \hat{e}_2 aus $\hat{e}_i = \mathbf{X}\hat{\beta}_i - \mathbf{y}_i$ und ihre mit $\hat{\sigma}_i^2$ berechneten Standardabweichungen aus den Diagonalelementen der Matrix $\hat{\sigma}_i^2(\mathbf{I} -$

$\mathbf{X}(\mathbf{X}'\mathbf{X})_{rs}^{-1}\mathbf{X}'$), die alle den Wert $0,666 \hat{\sigma}_i^2$ besitzen. Damit ergeben sich die studentisierten homogenisierten Residuenvektoren $\boldsymbol{\tau}_1$ und $\boldsymbol{\tau}_2$ und wegen (442.12) die Testgrößen für den Ausreißertest zu

$$\begin{aligned}\boldsymbol{\tau}'_1 &= | 1,58; -0,32; -1,27; -0,95; 0,95; 0,00 | \\ \boldsymbol{\tau}'_2 &= | 0,27; -1,35; 1,08; -1,21; 1,21; 0,00 | .\end{aligned}$$

Bei einem Signifikanzniveau von $\alpha = 0,05$ beträgt $c_{0,95;6,4} = 1,93$, so daß die Hypothese, unter den Beobachtungen befinde sich kein Ausreißer, anzunehmen ist.

444 Zuverlässigkeitsmaße

Die Fähigkeit einer Parameterschätzung, grobe Fehler aufzudecken, wird nach BAARDA (1968) als Zuverlässigkeit bezeichnet, so daß als Maß für die Zuverlässigkeit der durch den Ausreißertest in der Beobachtung y_k zu entdeckende Ausreißer Δy_k dienen kann. Er wird als Maß für die *innere Zuverlässigkeit* benutzt, während die *äußere Zuverlässigkeit* den Einfluß eines Ausreißers auf die unbekannt Parameter angibt. Für die Analyse einer Parameterschätzung ist der aufzudeckende Ausreißer Δy_k besonders instruktiv, denn je kleiner Δy_k erhalten wird, desto besser kontrollieren die Beobachtungen sich gegenseitig.

Der in der Beobachtung y_k aufzudeckende Ausreißer Δy_k berechnet sich wie folgt. Nach (443.4) wird ein Ausreißer gefunden, falls der Absolutbetrag des transformierten Residuums e_k in (442.4) den Wert Δe überschreitet, für den gilt

$$\Delta e = \hat{\sigma} q_k c_{1-\alpha; n, n-q} . \quad (444.1)$$

Mit (345.4) und (345.5) erhält man aus (442.2) für das transformierte Residuum

$$e_k = -\mathbf{e}'_k \mathbf{P} \mathbf{Q}_e \mathbf{P} \mathbf{y} .$$

Befindet sich lediglich ein Ausreißer in den Beobachtungen y_1 bis y_n , und zwar der Ausreißer Δy_k in der Beobachtung y_k , dann verursacht er die Abweichung Δe in e_k mit

$$\Delta e = -\Delta y_k q_k^2 \quad (444.2)$$

und bei einer diagonalen Gewichtsmatrix \mathbf{P} mit (442.8) und (442.9) die Abweichung $\Delta \hat{e}$ in \hat{e}_k mit

$$\Delta \hat{e} = -\Delta y_k r_k , \quad (444.3)$$

so daß es von der Teilredundanz r_k abhängt, wie stark ein Ausreißer Δy_k sich im Residuum \hat{e}_k niederschlägt. Hierauf wurde bereits im Zusammenhang mit (385.11) hingewiesen.

Substituiert man (444.2) in (444.1), ergibt sich der Absolutwert des durch den Hypothesentest aufzudeckenden Ausreißers Δy_k zu

$$\Delta y_k = \hat{\sigma} c_{1-\alpha; n, n-q} / q_k \quad (444.4)$$

oder für eine diagonale Gewichtsmatrix \mathbf{P} mit (442.8) und (442.9)

$$\Delta y_k = \hat{\sigma} c_{1-\alpha; n, n-q} / (p_k \sqrt{q_{ekk}}) = \hat{\sigma} c_{1-\alpha; n, n-q} / \sqrt{p_k r_k} \quad (444.5)$$

oder mit (442.11)

$$\Delta y_k = \hat{\sigma} c_{1-\alpha; n, n-q} / \sqrt{p_k q_{hkk}} . \quad (444.6)$$

Ein im Betrag kleiner Ausreißer Δy_k wird also aufgedeckt, wenn die Standardabweichung des transformierten Residuums e_k , des Residuums \hat{e}_k oder des homogenisierten Residuums \hat{e}_{hk} groß wird, was durch die aus (345.5) folgende Beziehung $D(\hat{\mathbf{y}}) = D(\mathbf{y}) - D(\hat{\mathbf{e}})$ zu erklären ist. Ein großer Redundanzanteil r_k führt wegen (442.9) und (442.11) zu einer großen Standardabweichung und damit zu einem kleinen aufzudeckenden Ausreißer, was auch aus (444.3) ersichtlich ist.

Bezeichnet man mit $\Delta \mathbf{y}_k$ den $n \times 1$ Vektor, dessen Komponenten gleich Null sind außer der k -ten Komponente, die den Ausreißer Δy_k enthält, ergibt sich die durch Δy_k verursachte Änderung $\Delta \beta_k$ des Vektors β_b der unbekannt projizierten Parameter nach (345.3) zu

$$\Delta \beta_k = (\mathbf{X}' \mathbf{P} \mathbf{X})_{rs}^{-1} \mathbf{X}' \mathbf{P} \Delta \mathbf{y}_k . \quad (444.7)$$

Als Maß λ_k der äußeren Zuverlässigkeit wird die quadratische Form

$$\lambda_k = \Delta \beta_k' \mathbf{X}' \mathbf{P} \mathbf{X} \Delta \beta_k / \hat{\sigma}^2 \quad \text{für } k \in \{1, \dots, n\} \quad (444.8)$$

definiert, da sich die Matrix $\mathbf{X}' \mathbf{P} \mathbf{X}$ als Gewichtsmatrix der unbekannt Parameter interpretieren läßt. Zur Veranschaulichung von λ_k sei der Fehler Δf_k einer beliebigen linearen Funktion $f = \mathbf{a}' \beta_b$ der projizierten Parameter berechnet, der durch den Ausreißer Δy_k verursacht wird. Man erhält mit (444.7)

$$\Delta f_k = \mathbf{a}' \Delta \beta_k = \mathbf{a}' (\mathbf{G}')^{-1} \mathbf{G}' \Delta \beta_k$$

und mit der Schwarzschen Ungleichung (BLATTER 1974, Vol.I, S.62)

$$\Delta f_k^2 \leq \hat{\sigma}^2 \mathbf{a}' (\mathbf{G}')^{-1} \mathbf{G}^{-1} \mathbf{a} \Delta \beta_k' \mathbf{G} \mathbf{G}' \Delta \beta_k / \hat{\sigma}^2 ,$$

worin \mathbf{G} die untere Dreiecksmatrix der Cholesky-Faktorisierung der wegen (143.8) positiv definiten Matrix $\mathbf{X}'\mathbf{P}\mathbf{X} + \mathbf{B}'\mathbf{B} = \mathbf{G}\mathbf{G}'$ und \mathbf{B} die Matrix in (155.15) zur Berechnung von $(\mathbf{X}'\mathbf{P}\mathbf{X})_{rs}^-$ in der Projektion (345.2) bedeuten. Substituiert man $(\mathbf{G}')^{-1}\mathbf{G}^{-1} = (\mathbf{X}'\mathbf{P}\mathbf{X} + \mathbf{B}'\mathbf{B})^{-1}$ durch (155.17) und beachtet, daß in $f = \mathbf{a}'\boldsymbol{\beta}_b$ wegen der Schätzbarkeit nach (332.2) $\mathbf{a}' = \mathbf{c}'\mathbf{X}$ gilt, erhält man mit (155.11) und (155.14)

$$\Delta f_k^2 \leq \hat{\sigma}^2 \mathbf{a}' (\mathbf{X}'\mathbf{P}\mathbf{X})_{rs}^- \mathbf{a} \Delta \boldsymbol{\beta}'_k \mathbf{X}'\mathbf{P}\mathbf{X} \Delta \boldsymbol{\beta}_k / \hat{\sigma}^2$$

und schließlich

$$\Delta f_k \leq \hat{\sigma}_f \sqrt{\lambda_k} \quad \text{mit} \quad \hat{\sigma}_f^2 = \hat{\sigma}^2 \mathbf{a}' (\mathbf{X}'\mathbf{P}\mathbf{X})_{rs}^- \mathbf{a} . \quad (444.9)$$

Eine obere Schranke für den Fehler Δf_k in einer beliebigen Funktion f der unbekanntem Parameter, verursacht durch den Ausreißer Δy_k , läßt sich also mit der durch $\hat{\sigma}^2$ berechneten Standardabweichung $\hat{\sigma}_f$ von f und dem Maß λ_k der äußeren Zuverlässigkeit angeben.

Nimmt man an, daß die Alternativhypothese $\Delta y_k = \Delta \bar{y}_k$ in (442.1) richtig ist und daß $\Delta \bar{y}_k$ durch (444.4) bestimmt wird, erhält man den Nichtzentralitätsparameter λ in der Verteilung (442.5) zu

$$\lambda = (\hat{\sigma}/\sigma_\epsilon)^2 c_{1-\alpha;n,n-q}^2$$

und mit $\hat{\sigma}/\sigma_\epsilon \approx 1$

$$\lambda = c_{1-\alpha;n,n-q}^2 . \quad (444.10)$$

Hiermit läßt sich (422.10) entsprechend die Wahrscheinlichkeit $1 - \beta$ berechnen, daß ein Ausreißer nicht entdeckt wird, beziehungsweise die Wahrscheinlichkeit β , einen Ausreißer aufzudecken,

$$1 - \beta = \int_0^{c_{1-\alpha;n,n-q}} \tau'(1, n - q - 1, \lambda) d\tau . \quad (444.11)$$

Die Wahrscheinlichkeit β variiert in Abhängigkeit von n und $n - q$. Wünscht man dagegen einen Ausreißer derart zu bestimmen, daß die Wahrscheinlichkeit β seines Aufdeckens konstant und zum Beispiel $\beta = 0,80$ beträgt, ist (444.11) nach λ und mit (442.5) nach $\Delta \bar{y}_k$ aufzulösen. Dies ist der von BAARDA (1968) gewählte Weg, der aber rechenintensiver ist (GAIDA und KOCH 1985) als die Ermittlung eines Ausreißers nach (444.4) bis (444.6).

Literatur

- ABRAMOWITZ, M. und I.A. STEGUN (1972) *Handbook of Mathematical Functions*. Dover, New York.
- ACKERMANN, F., H. EBNER und H. KLEIN (1970) Ein Programm-Paket für die Aerotriangulation mit unabhängigen Modellen. *Bildmessung und Luftbildwesen*, 38:218–224.
- AHRENS, H. (1968) *Varianzanalyse*. Akademie-Verlag, Berlin.
- ALBERT, A. (1972) *Regression and the Moore-Penrose Pseudoinverse*. Academic Press, New York.
- ANDERSON, T.W. (1958) *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.
- ANDERSON, T.W. (1962) The choice of the degree of a polynomial regression as a multiple decision problem. *Annals Mathematical Statistics*, 33:255–265.
- ANDREWS, D.F., R. GNANADESIKAN und J.L. WARNER (1973) Methods for assessing multivariate normality. In: KRISHNAIAH, P.R. (Hrsg.), *Multivariate Analysis III*. Academic Press, New York, 95–116.
- BAARDA, W. (1968) *A testing procedure for use in geodetic networks*. Publications on Geodesy, Vol. 2, Nr. 5. Netherlands Geodetic Commission, Delft.
- BAARDA, W. (1973) *S-transformations and criterion matrices*. Publications on Geodesy, Vol. 5, Nr. 1. Netherlands Geodetic Commission, Delft.
- BÄHR, H.-G. und R. RICHTER (1975) Über die Wahl von a-priori-Korrelationen. *Z Vermessungswesen*, 100:180–188.
- BARKER, V.A. (Hrsg.) (1977) *Sparse Matrix Techniques*. Springer, Berlin.
- BARNETT, V. und T. LEWIS (1978) *Outliers in Statistical Data*. Wiley, New York.
- BEN-ISRAEL, A. und T.N.E. GREVILLE (1974) *Generalized Inverses: Theory and Applications*. Wiley, New York.

- BENNING, W. (1982) Zur Auswertung von Katastermessungen mithilfe automatisierter Fehlersuche – Das Programmsystem KAROLA –. *Allgemeine Vermessungs-Nachrichten*, 89:2–18.
- BENNING, W. (1995) Vergleich dreier L_p -Schätzer zur Fehlersuche in hybriden Lagenetzen. *Z Vermessungswesen*, 120:606–617.
- BJERHAMMAR, A. (1973) *Theory of Errors and Generalized Matrix Inverses*. Elsevier, Amsterdam.
- BLATTER, C. (1974) *Analysis I, II, III*. Springer, Berlin.
- BOCK, R.D. (1975) *Multivariate Statistical Methods in Behavioral Research*. McGraw-Hill, New York.
- BÖHME, G. (1974) *Anwendungsorientierte Mathematik, Bd. 1: Algebra*. Springer, Berlin.
- BOSSLER, J.D. (1972) *Bayesian Inference in Geodesy*. Dissertation, The Ohio State University, Columbus.
- BOULLION, T.L. und P.L. ODELL (1971) *Generalized Inverse Matrices*. Wiley, New York.
- BRAMMER, K. und G. SIFFLING (1975) *Kalman-Bucy-Filter*. Oldenbourg, München.
- BROEMELING, L.D. (1985) *Bayesian Analysis of Linear Models*. Dekker, New York.
- BUKY, R.S. und P.D. JOSEPH (1968) *Filtering for Stochastic Processes with Applications to Guidance*. Interscience, New York.
- CAROSIO, A. (1995) Ausgleichung geodätischer Netze mit Verfahren der robusten Statistik. *Vermessung, Photogrammetrie, Kulturtechnik*, 93:188–191.
- CARTA, D.G. (1975) Low-order approximations for the normal probability integral and the error function. *Mathematics of Computation*, 29:856–862.
- CASPARY, W. (1978) Zur Lösung singulärer Ausgleichungsmodelle durch Bedingungsgleichungen. *Allgemeine Vermessungs-Nachrichten*, 85:81–87.
- CASPARY, W. und K. WICHMANN (1994) *Lineare Modelle*. Oldenbourg, München.

- CONSTANTINE, A.G. (1963) Some non-central distribution problems in multivariate analysis. *Annals Mathematical Statistics*, 34:1270–1285.
- CONSUL, P.C. (1969) The exact distributions of likelihood criteria for different hypotheses. In: KRISHNAIAH, P.R. (Hrsg.), *Multivariate Analysis II*. Academic Press, New York, 171–181.
- CRAMÉR, H. (1946) *Mathematical Methods of Statistics*. Princeton University Press, Princeton.
- DOKSUM, K.A. (1976) Some remarks on the development of nonparametric methods and robust statistical inference. In: OWEN, D.B. (Hrsg.), *On the History of Statistics and Probability*. Dekker, New York, 237–263.
- DRAPER, N.R. und H. SMITH (1966) *Applied Regression Analysis*. Wiley, New York.
- DRYGAS, H. (1977) Best quadratic unbiased estimation in variance-covariance component models. *Mathematische Operationsforschung und Statistik, Series Statistics*, 8:211–231.
- EBNER, H. (1972) A posteriori Varianzschätzungen für die Koordinaten unabhängiger Modelle. *Z Vermessungswesen*, 97:166–172.
- EHLERT, D. (1977) *Speicherplatz sparende EDV-Programme zur Auflösung von Gleichungssystemen mit symmetrischen Koeffizientenmatrizen*. Reihe B, 222. Deutsche Geodätische Kommission, München.
- FADDEEV, D.K. und V.N. FADDEEVA (1963) *Computational Methods of Linear Algebra*. Freeman, San Francisco.
- FAHRMEIR, L. und A. HAMERLE (Hrsg.) (1984) *Multivariate statistische Verfahren*. Walter de Gruyter, Berlin.
- FISHER, R.A. und F. YATES (1963) *Statistical Tables*. Longman, Edinburgh.
- FISZ, M. (1976) *Wahrscheinlichkeitsrechnung und mathematische Statistik*. Deutscher Verlag der Wissenschaften, Berlin.
- FÖRSTNER, W. (1979A) Das Programm TRINA zur Ausgleichung und Gütebeurteilung geodätischer Lagenetze. *Z Vermessungswesen*, 104:61–72.
- FÖRSTNER, W. (1979B) Ein Verfahren zur Schätzung von Varianz- und Kovarianzkomponenten. *Allgemeine Vermessungs-Nachrichten*, 86:446–453.

- FRITSCH, D. (1985) Some additional informations on the capacity of the linear complementarity algorithm. In: GRAFAREND, E.W. und F. SANZO (Hrsg.), *Optimization and Design of Geodetic Networks*. Springer, Berlin, 169–184.
- FRÖHLICH, H. und H. DUDDEK (1983) Zur Ermittlung der äußeren Genauigkeit elektrooptisch gemessener Strecken mittlerer Reichweite auf der Eichlinie Jackerath. *Markscheidewesen*, 90:182–184.
- GAIDA, W. und K.R. KOCH (1985) Solving the cumulative distribution function of the noncentral F -distribution for the noncentrality parameter. *Scientific Bulletins of the Stanislaw Staszic University of Mining and Metallurgy*, Geodesy b.90(1024):35–44.
- GAUSS, C.F. (1809) *Theoria Motus Corporum Coelestium*. Perthes und Besser, Hamburg.
- GAUSS, C.F. (1823) *Theoria Combinationis Observationum*. Dieterich, Göttingen.
- GEORGE, A. und J.W. LIU (1981) *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall, Englewood Cliffs.
- GNEDENKO, B.W. (1957) *Lehrbuch der Wahrscheinlichkeitsrechnung*. Akademie-Verlag, Berlin.
- GOTTHARDT, E. (1978) *Einführung in die Ausgleichsrechnung*. Wichmann, Karlsruhe.
- GRAFAREND, E. und A. D'HONE (1978) *Gewichtsschätzung in geodätischen Netzen*. Reihe A, 88. Deutsche Geodätische Kommission, München.
- GRAFAREND, E. und B. SCHAFFRIN (1976) Equivalence of estimable quantities and invariants in geodetic networks. *Z Vermessungswesen*, 101:485–491.
- GRAFAREND, E. und B. SCHAFFRIN (1979) Kriterium-Matrizen I – zweidimensionale homogene und isotrope geodätische Netze. *Z Vermessungswesen*, 104:133–149.
- GRAFAREND, E., A. KLEUSBERG und B. SCHAFFRIN (1980) An introduction to the variance-covariance-component estimation of Helmert type. *Z Vermessungswesen*, 105:161–180.
- GRAFAREND, E.W. und B. SCHAFFRIN (1993) *Ausgleichsrechnung in linearen Modellen*. B.I. Wissenschaftsverlag, Mannheim.

- GRAYBILL, F.A. (1969) *Introduction to Matrices with Applications in Statistics*. Wadsworth, Belmont.
- GRAYBILL, F.A. (1976) *Theory and Application of the Linear Model*. Duxbury Press, North Scituate.
- GREGORY, R.T. und D.L. KARNEY (1969) *A Collection of Matrices for Testing Computational Algorithms*. Wiley, New York.
- GREPEL, U. (1987) *Effiziente Rechenverfahren für umfangreiche geodätische Parameterschätzungen*. Nr. 75. Mitteilungen aus den Geodätischen Instituten der Rheinischen Friedrich-Wilhelms-Universität, Bonn.
- GROSSMANN, W. (1969) *Grundzüge der Ausgleichsrechnung*. Springer, Berlin.
- GROTEMEYER, K.P. (1970) *Lineare Algebra*. Bibliographisches Institut, Mannheim.
- GRÜN, A. (1978) Progress in photogrammetric point determination by compensation of systematic errors and detection of gross errors. *Nachrichten aus dem Karten- und Vermessungswesen*, Reihe II(36):113–140.
- HAGEN, G. (1837) *Grundzüge der Wahrscheinlichkeitsrechnung*. Dümmler, Berlin.
- HAMPEL, F.R., E.M. RONCHETTI, P.R. ROUSSEEUW und W.A. STAHEL (1986) *Robust Statistics*. Wiley, New York.
- HARVILLE, D. (1976) Extension of the Gauss-Markov theorem to include the estimation of random effects. *Annals of Statistics*, 4:384–395.
- HARVILLE, D.A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J American Statistical Association*, 72:320–338.
- HAWKINS, D.M. (1980) *Identification of Outliers*. Chapman and Hall, London.
- HECK, B., E. KUNTZ und B. MEIER-HIRMER (1977) Deformationsanalyse mittels relativer Fehlerellipsen. *Allgemeine Vermessungs-Nachrichten*, 84:78–87.
- HEIN, G. (1978) Multivariate Analyse der Nivellementsdaten im Oberrheingraben und Rheinischen Schild. *Z Vermessungswesen*, 103:430–436.

- HEITZ, S. (1968) *Geoidbestimmung durch Interpolation nach kleinsten Quadraten aufgrund gemessener und interpolierter Lotabweichungen*. Reihe C, 124. Deutsche Geodätische Kommission, München.
- HELMERT, F.R. (1872) *Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. Teubner, Leipzig.
- HELMERT, F.R. (1924) *Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. 3. Auflage. Teubner, Leipzig.
- HENRICI, P. (1977) *Applied and Computational Complex Analysis, Vol. 2*. Wiley, New York.
- HINDERER, K. (1972) *Grundbegriffe der Wahrscheinlichkeitstheorie*. Springer, Berlin.
- HIRVONEN, R.A. (1962) On the statistical analysis of gravity anomalies. *Annales Academiae Scientiarum Fennicae*, 58. Series A, III, Helsinki.
- HOAGLIN, D.C., F. MOSTELLER und J.W. TUKEY (1983) *Understanding Robust and Exploratory Data Analysis*. Wiley, New York.
- HOLLANDER, M. und D.A. WOLFE (1973) *Nonparametric Statistical Methods*. Wiley, New York.
- HÖPCKE, W. (1980) *Fehlerlehre und Ausgleichsrechnung*. Walter de Gruyter, Berlin.
- HOTELLING, H. (1951) A generalized T test and measure of multivariate dispersion. In: NEYMAN, J. (Hrsg.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley and Los Angeles, 23–41.
- HOUSEHOLDER, A.S. (1964) *The Theory of Matrices in Numerical Analysis*. Blaisdell, New York.
- HUANG, Y. (1995) On the design of estimators with high breakdown points for outlier identification in triangulation networks. *Bulletin Géodésique*, 69:292–299.
- HUBER, P.J. (1964) Robust estimation of a location parameter. *Annals Mathematical Statistics*, 35:73–101.
- HUBER, P.J. (1981) *Robust Statistics*. Wiley, New York.
- HUMAK, K.M.S. (1977) *Statistische Methoden der Modellbildung, Bd. 1*. Akademie-Verlag, Berlin.

- ITO, K. (1962) A comparison of the powers of two multivariate analysis of variance tests. *Biometrika*, 49:455–462.
- JAGLOM, A.M. (1959) *Einführung in die Theorie der stationären Zufallsfunktionen*. Akademie-Verlag, Berlin.
- JAMES, A.T. (1964) Distributions of matrix variates and latent roots derived from normal samples. *Annals Mathematical Statistics*, 35:475–501.
- JENKINS, G.M. und D.G. WATTS (1968) *Spectral Analysis and its Applications*. Holden-Day, San Francisco.
- JENNINGS, A. (1977) *Matrix Computation for Engineers and Scientists*. Wiley, New York.
- JOHNSON, N.L. und S. KOTZ (1970) *Distributions in Statistics: Continuous Univariate Distributions, Vol. 1, 2*. Houghton Mifflin, Boston.
- JOHNSON, N.L. und F.C. LEONE (1977) *Statistics and Experimental Design, Vol. 1, 2*. Wiley, New York.
- KAMPMANN, G. (1993) Auswertetechniken bei der überbestimmten Koordinatentransformation. *BDVI-Forum*, 19:139–152.
- KAMPMANN, G. (1994) Robuste Deformationsanalyse mittels balancierter Ausgleichung. *Allgemeine Vermessungs-Nachrichten*, 101:8–17.
- KELM, R. (1978) Ist die Varianzschätzung nach Helmert MINQUE? *Allgemeine Vermessungs-Nachrichten*, 85:49–54.
- KLEFFE, J. (1978) Simultaneous estimation of expectation and covariance matrix in linear models. *Mathematische Operationsforschung und Statistik, Series Statistics*, 9:443–478.
- KLEFFE, J. und R. PINCUS (1974) Bayes and best quadratic unbiased estimators for parameters of the covariance matrix in a normal linear model. *Mathematische Operationsforschung und Statistik*, 5:43–67.
- KOCH, K.R. (1973) Höheninterpolation mittels gleitender Schrägebene und Prädiktion. *Vermessung, Photogrammetrie, Kulturtechnik, Mitteilungsblatt*, 71:229–232.
- KOCH, K.R. (1977) Least squares adjustment and collocation. *Bulletin Géodésique*, 51:127–135.
- KOCH, K.R. (1978) Hypothesentests bei singulären Ausgleichungsproblemen. *Z Vermessungswesen*, 103:1–10.

- KOCH, K.R. (1981) Varianz- und Kovarianzkomponentenschätzung für Streckenmessungen auf Eichlinien. *Allgemeine Vermessungs-Nachrichten*, 88:125–132.
- KOCH, K.R. (1982A) Kalman filter and optimal smoothing derived by the regression model. *Manuscripta Geodaetica*, 7:133–144.
- KOCH, K.R. (1982B) *Optimization of the configuration of geodetic networks*. Reihe B, 258/III: 82–89. Deutsche Geodätische Kommission, München.
- KOCH, K.R. (1982C) S-transformations and projections for obtaining estimable parameters. In: “Forty Years of Thought”. *Anniversary Volume on the Occasion of Prof. Baarda’s 65th Birthday*, 1:136–144. Delft.
- KOCH, K.R. (1983A) Ausreißertests und Zuverlässigkeitsmaße. *Vermessungswesen und Raumordnung*, 45:400–411.
- KOCH, K.R. (1983B) Die Wahl des Datums eines trigonometrischen Netzes bei Punkteinschaltungen. *Z Vermessungswesen*, 108:104–111.
- KOCH, K.R. (1983C) Rechenverfahren bei der Einschaltung von Punkten in ein trigonometrisches Netz. *Allgemeine Vermessungs-Nachrichten*, 90:99–107.
- KOCH, K.R. (1984) Statistical tests for detecting crustal movements using Bayesian inference. *NOAA Technical Report NOS NGS 29*, US Department of Commerce, National Geodetic Survey, Rockville, Maryland.
- KOCH, K.R. (1985A) Ein statistisches Auswerteverfahren für Deformationsmessungen. *Allgemeine Vermessungs-Nachrichten*, 92:97–108.
- KOCH, K.R. (1985B) Test von Ausreißern in Beobachtungspaaren. *Z Vermessungswesen*, 110:34–38.
- KOCH, K.R. (1986) Maximum likelihood estimate of variance components; ideas by A.J. Pope. *Bulletin Géodésique*, 60:329–338.
- KOCH, K.R. (1990) *Bayesian Inference with Geodetic Applications*. Springer, Berlin.
- KOCH, K.R. (1994) Bayessche Inferenz für die Prädiktion und Filterung. *Z Vermessungswesen*, 119:464–470.
- KOCH, K.R. (1995) Statistische Grundlagen zur Untersuchung von Immobilienwerten. In: SCHMALGEMEIER, H. (Hrsg.), *Statistische Methoden in der Grundstückswertermittlung*. Schriftenreihe des DVW, 16, Wittwer, Stuttgart, 7–12.

- KOCH, K.R. (1996) Robuste Parameterschätzung. *Allgemeine Vermessungs-Nachrichten*, 103:1–18.
- KOCH, K.R. (2000) *Einführung in die Bayes-Statistik*. Springer, Berlin.
- KOCH, K.R. und J. KUSCHE (2002) Regularization of geopotential determination from satellite data by variance components. *J Geodesy*, 76:259–268.
- KOCH, K.R. und A.J. POPE (1969) Least squares adjustment with zero variances. *Z Vermessungswesen*, 94:390–393.
- KOCH, K.R. und K. RIESMEIER (1985) Bayesian inference for the derivation of less sensitive hypothesis tests. *Bulletin Géodésique*, 59:167–179.
- KOCH, K.R. und M. SCHMIDT (1994) *Deterministische und stochastische Signale*. Dümmler, Bonn.
- KOCH, K.R. und Y. YANG (1998) Konfidenzbereiche und Hypothesentests für robuste Parameterschätzungen. *Z Vermessungswesen*, 123:20–26.
- KOK, J.J. (1984) On data snooping and multiple outlier testing. *NOAA Technical Report NOS NGS 30*, US Department of Commerce, National Geodetic Survey, Rockville, Maryland.
- KORIN, B.P. (1968) On the distribution of a statistic used for testing a covariance matrix. *Biometrika*, 55:171–178.
- KOWALSKY, H.J. (1977) *Lineare Algebra*. Walter de Gruyter, Berlin.
- KRARUP, T. (1969) *A contribution to the mathematical foundation of physical geodesy*. Geodaetisk Institut, Meddelelse No.44, Kopenhagen.
- KRARUP, T., J. JUHL und K. KUBIK (1980) Götterdämmerung over least squares adjustment. *14th Congress ISP Hamburg, International Archives of Photogrammetry*, XXIII, B3, Commission III:369–378.
- KRAUS, K. (1971) Automatische Berechnung digitaler Höhenlinien. *Z Vermessungswesen*, 96:233–239.
- KRES, H. (1983) *Statistical Tables for Multivariate Analysis*. Springer, Berlin.
- KUBIK, K. (1970) The estimation of the weights of measured quantities within the method of least squares. *Bulletin Géodésique*, 95:21–40.
- KUBIK, K., W. WENG und P. FREDERIKSEN (1985) Oh, gross errors! *Australian Journal of Geodesy, Photogrammetry and Surveying*, 42:1–18.

- LAMOTTE, L.R. (1973A) On non-negative quadratic unbiased estimation of variance components. *J American Statistical Association*, 68:728–730.
- LAMOTTE, L.R. (1973B) Quadratic estimation of variance components. *Biometrics*, 29:311–330.
- LÄUTER, J. (1974) Approximation des Hotellingschen T^2 durch die F -Verteilung. *Biometrische Zeitschrift*, 16:191–202.
- LAWLEY, D.N. (1938) A generalization of Fisher's z -test. *Biometrika*, 30:180–187, 467–469.
- LAWSON, C.L. und R.J. HANSON (1974) *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs.
- LEHMANN, E.L. (1959) *Testing Statistical Hypotheses*. Wiley, New York.
- LINKWITZ, K. (1977) Über eine neue Anwendung der Gaußschen Methode der kleinsten Quadrate: Die Formfindung und statische Analyse von räumlichen Seil- und Hängernetzen. *Abhandlungen der Braunschweigischen Wissenschaftlichen Gesellschaft*, Göttingen, 27:121–153.
- LINNIK, J.W. (1961) *Methode der kleinsten Quadrate in moderner Darstellung*. Deutscher Verlag der Wissenschaften, Berlin.
- LUND, R.E. (1975) Tables for an approximate test for outliers in linear models. *Technometrics*, 17:473–476.
- MARKOFF, A.A. (1912) *Wahrscheinlichkeitsrechnung*. Teubner, Leipzig.
- MEIER, S. und W. KELLER (1990) *Geostatistik*. Springer, Wien.
- MEISSL, P. (1969) Zusammenfassung und Ausbau der inneren Fehlertheorie eines Punkthaufens. In: RINNER, K., K. KILLIAN und P. MEISSL (Hrsg.), *Beiträge zur Theorie der geodätischen Netze im Raum*. Reihe A, 61: 8–21. Deutsche Geodätische Kommission, München.
- MEISSL, P. (1976) Hilbert spaces and their applications to geodetic least squares problems. *Bollettino di Geodesia e Scienze Affini*, 35:49–80.
- MEISSL, P. (1980) A-priori prediction of roundoff error accumulation during the direct solution of a superlarge geodetic normal equation system. *NOAA Professional Paper 12*, US Department of Commerce, National Geodetic Survey, Rockville, Maryland.
- MESCHKOWSKI, H. (1962) *Hilbertsche Räume mit Kernfunktion*. Springer, Berlin.

- MIERLO, J. VAN (1981) A testing procedure for analytic geodetic deformation measurements. In: HALLERMANN, L. (Hrsg.), *Beiträge zum II. Internationalen Symposium über Deformationsmessungen mit geodätischen Methoden*. Wittwer, Stuttgart, 321–353.
- MIKHAIL, E.M. und F. ACKERMANN (1976) *Observations and Least Squares*. Dun-Donnelley, New York.
- MITTERMAYER, E. (1972) Zur Ausgleichung freier Netze. *Z Vermessungswesen*, 97:481–489.
- MONIN, A.S. und A.M. YAGLOM (1975) *Statistical Fluid Mechanics, Vol. 2*. MIT Press, Cambridge.
- MOOD, A.M., F.A. GRAYBILL und D.C. BOES (1974) *Introduction to the Theory of Statistics*. McGraw-Hill Kogakusha, Tokyo.
- MORITZ, H. (1973) *Least-squares collocation*. Reihe A, 75. Deutsche Geodätische Kommission, München.
- MORITZ, H. (1980) *Advanced Physical Geodesy*. Wichmann, Karlsruhe.
- MUDHOLKAR, G.S., Y.P. CHAUBEY und C.C. LIN (1976) Some approximations for the noncentral F -distribution. *Technometrics*, 18:351–358.
- MUIRHEAD, R.J. (1972) The asymptotic noncentral distribution of Hotelling's generalized T_0^2 . *Annals Mathematical Statistics*, 43:1671–1677.
- MÜLLER, P.H. (Hrsg.) (1975) *Lexikon der Stochastik*. Akademie-Verlag, Berlin.
- NAGARSENKER, B.N. (1977) On the exact non-null distributions of the LR criterion in a general MANOVA model. *Sankya*, 39(Series A):251–263.
- NEISS, F. und H. LIERMANN (1975) *Determinanten und Matrizen*. Springer, Berlin.
- NEUBURGER, E. (1972) *Einführung in die Theorie des linearen Optimalfilters*. Oldenbourg, München.
- PAPO, H.B. und A. PERELMUTER (1983) Reparametrization of deformation analysis. *Manuscripta Geodaetica*, 8:41–58.
- PAPOULIS, A. (1965) *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York.

- PATNAIK, P.B. (1949) The non-central χ^2 - and F -distributions and their applications. *Biometrika*, 36:202–232.
- PEARSON, E.S. und H.O. HARTLEY (1976) *Biometrika Tables for Statisticians, Vol.I, II*. Biometrika Trust, London.
- PELZER, H. (1971) *Zur Analyse geodätischer Deformationsmessungen*. Reihe C, 164. Deutsche Geodätische Kommission, München.
- PELZER, H. (1978) Ein indirektes Vergleichswertverfahren und Anwendung statistischer Methoden. *Z Vermessungswesen*, 103:245–254.
- PILLAI, K.C.S. und K. JAYACHANDRAN (1967) Power comparisons of tests of two multivariate hypotheses based on four criteria. *Biometrika*, 54:195–210.
- PILLAI, K.C.S. und D.L. YOUNG (1971) On the exact distribution of Hotelling's generalized T_0^2 . *J Multivariate Analysis*, 1:90–107.
- PODER, K. und C.C. TSCHERNING (1973) *Cholesky's method on a computer*. Internal Report No. 8. The Danish Geodetic Institute, Kopenhagen.
- POPE, A.J. (1971) Transformation of covariance matrices due to changes in minimal control (Zusammenfassung). *EOS, Transactions, American Geophysical Union*, 52:820.
- POPE, A.J. (1976) The statistics of residuals and the detection of outliers. *NOAA Technical Report NOS65 NGS1*, US Department of Commerce, National Geodetic Survey, Rockville, Maryland.
- PRESS, S.J. (1972) *Applied Multivariate Analysis*. Holt, Rinehart and Winston, New York.
- PRICE, R. (1964) Some non-central F -distributions expressed in closed form. *Biometrika*, 51:107–122.
- PUKELSHEIM, F. (1979) *On the existence of unbiased nonnegative estimates of variance covariance components*. Institut für Mathematische Stochastik der Universität, Freiburg.
- RAO, C.R. (1973) *Linear Statistical Inference and its Applications*. Wiley, New York.
- RAO, C.R. und S.K. MITRA (1971) *Generalized Inverse of Matrices and its Applications*. Wiley, New York.

- REISSMANN, G. (1976) *Die Ausgleichungsrechnung*. Verlag für Bauwesen, Berlin.
- RIESMEIER, K. (1984) *Test von Ungleichungshypothesen in linearen Modellen mit Bayes-Verfahren*. Reihe C, 292. Deutsche Geodätische Kommission, München.
- ROHDE, C.A. (1965) Generalized inverses of partitioned matrices. *J Soc Indust Appl Math*, 13:1033–1035.
- ROUSSEEUW, P.J. (1984) Least median of squares regression. *J American Statistical Association*, 79:871–880.
- ROUSSEEUW, P.J. und A.M. LEROY (1987) *Robust Regression and Outlier Detection*. Wiley, New York.
- ROY, S.N. (1957) *Some Aspects of Multivariate Analysis*. Wiley, New York.
- ROY, S.N., R. GNANADESIKAN und J.N. SRIVASTAVA (1971) *Analysis and Design of Certain Quantitative Multiresponse Experiments*. Pergamon Press, Oxford.
- RUMMEL, R. (1976) A model comparison in least squares collocation. *Bulletin Géodésique*, 50:181–192.
- RUTISHAUSER, H. (1976) *Vorlesungen über numerische Mathematik, Bd. 1, 2*. Birkhäuser, Basel.
- SACHS, L. (1978) *Angewandte Statistik*. Springer, Berlin.
- SCHACH, S. und T. SCHÄFER (1978) *Regressions- und Varianzanalyse*. Springer, Berlin.
- SCHAFFRIN, B. (1981) Ausgleichung mit Bedingungs-Ungleichungen. *Allgemeine Vermessungs-Nachrichten*, 88:227–238.
- SCHAFFRIN, B. (1983) *Varianz-Kovarianz-Komponenten-Schätzung bei der Ausgleichung heterogener Wiederholungsmessungen*. Reihe C, 282. Deutsche Geodätische Kommission, München.
- SCHAFFRIN, B. (1985) *Das geodätische Datum mit stochastischer Vorinformation*. Reihe C, 313. Deutsche Geodätische Kommission, München.
- SCHEFFÉ, H. (1959) *The Analysis of Variance*. Wiley, New York.
- SCHEK, H.J. und P. MAIER (1976) Nichtlineare Normalgleichungen zur Bestimmung der Unbekannten und deren Kovarianzmatrix. *Z Vermessungswesen*, 101:149–159.

- SCHEK, H.J., F. STEIDLER und U. SCHAUER (1977) *Ausgleichung großer geodätischer Netze mit Verfahren für schwach besetzte Matrizen*. Reihe A, 87. Deutsche Geodätische Kommission, München.
- SCHENDEL, U. (1977) *Sparse-Matrizen*. Oldenbourg, München.
- SCHLOSSMACHER, E.J. (1973) An iterative technique for absolute deviations curve fitting. *J American Statistical Association*, 68:857–859.
- SCHWARZ, C.R. (1978) TRAV10 horizontal network adjustment program. *NOAA Technical Memorandum NOS NGS 12*, US Department of Commerce, National Geodetic Survey, Rockville, Maryland.
- SCHWARZ, H.R., H. RUTISHAUSER und E. STIEFEL (1972) *Numerik symmetrischer Matrizen*. Teubner, Stuttgart.
- SEARLE, S.R. (1971) *Linear Models*. Wiley, New York.
- SEBER, G.A.F. (1977) *Linear Regression Analysis*. Wiley, New York.
- SJÖBERG, L.E. (1985) Adjustment and variance-covariance component estimation with a singular covariance matrix. *Z Vermessungswesen*, 110:145–151.
- SMIRNOW, W.I. (1975) *Lehrgang der höheren Mathematik, Teil I, II*. Deutscher Verlag der Wissenschaften, Berlin.
- SNAY, R.A. (1976) Reducing the profile of sparse symmetric matrices. *NOAA Technical Memorandum NOS NGS 4*, National Geodetic Survey, Rockville, Maryland.
- SOMOGYI, J. und J. ZÁVOTI (1993) Robust estimation with iteratively reweighted least-squares method. *Acta Geod Geoph Mont Hung*, 28:413–420.
- SPÄTH, H. (1974) *Algorithmen für multivariable Ausgleichsmodelle*. Oldenbourg, München.
- SPÄTH, H. (1987) *Mathematische Software zur linearen Regression*. Oldenbourg, München.
- SRIVASTAVA, J.N. (1966) Some generalizations of multivariate analysis of variance. In: KRISHNAIAH, P.R. (Hrsg.), *Multivariate Analysis*. Academic Press, New York, 129–145.
- STIEFEL, E. (1970) *Einführung in die numerische Mathematik*. Teubner, Stuttgart.

- STRANG, G. (1980) *Linear Algebra and its Applications*. Academic Press, New York.
- SUGIURA, N. und Y. FUJIKOSHI (1969) Asymptotic expansions of the non-null distributions of the likelihood ratio criteria for multivariate linear hypothesis and independence. *Annals Mathematical Statistics*, 40:942–952.
- TEUNISSEN, P.J.G. (1985) *The geometry of geodetic inverse linear mapping and non-linear adjustment*. Publications on Geodesy, Vol. 8. Netherlands Geodetic Commission, Delft.
- TWARSON, R.P. (1973) *Sparse Matrices*. Academic Press, New York.
- TIKU, M.L. (1967) Tables of the power of the F -test. *J American Statistical Association*, 62:525–539.
- TIKU, M.L. (1972) More tables of the power of the F -test. *J American Statistical Association*, 67:709–710.
- TOUTENBURG, H. (1975) *Vorhersage in linearen Modellen*. Akademie-Verlag, Berlin.
- TSCHERNING, C.C. (1978) Collocation and least squares methods as a tool for handling gravity field dependent data obtained through space research techniques. In: HIEBER, S. und T.D. GUYENNE (Hrsg.), *European Workshop on Space Oceanography, Navigation and Geodynamics*. European Space Agency, Paris, 141–149.
- WELSCH, W. (1978) A posteriori Varianzenschätzung nach Helmert. *Allgemeine Vermessungs-Nachrichten*, 85:55–63.
- WENTZEL, E.S. und L.A. OWTSCHAROW (1975) *Aufgabensammlung zur Wahrscheinlichkeitsrechnung*. Akademie-Verlag, Berlin.
- WERNER, H. (1975) *Praktische Mathematik I*. Springer, Berlin.
- WICKI, F. (1992) *Robuste M -Schätzer und Zuverlässigkeit*. Bericht 190, Institut für Geodäsie und Photogrammetrie, Eidgenössische Technische Hochschule, Zürich.
- WILKS, S.S. (1932) Certain generalizations in the analysis of variance. *Biometrika*, 24:471–494.
- WILKS, S.S. (1962) *Mathematical Statistics*. Wiley, New York.

- WITTING, H. und G. NÖLLE (1970) *Angewandte Mathematische Statistik*. Teubner, Stuttgart.
- WOLF, H. (1968) *Ausgleichsrechnung nach der Methode der kleinsten Quadrate*. Dümmler, Bonn.
- WOLF, H. (1973) Die Helmert-Inverse bei freien geodätischen Netzen. *Z Vermessungswesen*, 98:396–398.
- WOLF, H. (1975) *Ausgleichsrechnung, Formeln zur praktischen Anwendung*. Dümmler, Bonn.
- WOLF, H. (1978A) Das geodätische Gauß-Helmert-Modell und seine Eigenschaften. *Z Vermessungswesen*, 103:41–43.
- WOLF, H. (1978B) The Helmert block method – its origin and development. In: *Proceedings Second International Symposium on Problems Related to the Redefinition of North American Geodetic Networks*. US Department of Commerce, Washington, 319–326.
- WOLF, H. (1979) *Ausgleichsrechnung II, Aufgaben und Beispiele zur praktischen Anwendung*. Dümmler, Bonn.
- WROBEL, B. (1974) *Zur Steigerung der Auflösungsgenauigkeit von Normalgleichungen durch Konditionsverbesserung mittels additiver Modifikation*. Reihe C, 199. Deutsche Geodätische Kommission, München.
- XU, P. (1989) On robust estimation with correlated observations. *Bulletin Géodésique*, 63:237–252.
- YANG, Y. (1994) Robust estimation for dependent observations. *Manuscripta geodaetica*, 19:10–17.

Sachverzeichnis

- Ablehnungsbereich, 301, 303, 305
- Absolutbetrag eines Vektors, 12
- Absolutglied, 30, 59, 180
- affine Transformation, 45
- allgemeine Lösung, 60, 61, 68
- Allgemeinfall der Ausgleichsrechnung, 231
- Alternativhypothese, 301, 303, 305, 313, 315, 329
- Annahmebereich, 301, 303, 305
- Assoziativgesetz, 6, 8, 17
- Auffelderung, 207
- Ausgleichung nach bedingten Beobachtungen, 239
 - - vermittelnden Beobachtungen, 167
- Ausgleichsrechnung, 165, 167, 172, 231, 239
- Ausreißer, 226, 277, 282, 287, 289, 329, 331, 335
 - test, 226, 277, 329, 331, 333
- äußere Zuverlässigkeit, 335, 337
- Axiome der Wahrscheinlichkeit, 85, 91, 93

- b**alancierte Ausgleichung, 293
 - r Versuchsplan, 219
- Basis, 10, 11, 38, 45
 - des Nullraums, 64, 203, 205, 222
- Bayes-Statistik, 87, 307
- Bayessche Formel, 87
- bedingte Verteilung, 99, 130
 - -sfunktion, 99, 100
 - Wahrscheinlichkeit, 87, 99
- Bedingungsgleichung, 240
- Beobachtung, 89, 162, 164, 166, 197, 228, 231, 246, 260, 296
 - sfehler, 120
 - sgleichung, 167, 278, 291
- Bereichsschätzung, 162, 322
- beste erwartungstreue Schätzung, 163, 277
 - invariante quadratische erwartungstreue Schätzung, 248, 253, 259, 267
 - lineare erwartungstreue Schätzung, 169, 171, 185, 200, 210, 228, 233, 262

- quadratische erwartungstreue Schätzung, 247
- r Test, 302
- Schätzung, 163
- Betaverteilung, 125, 155
- bilineare Form, 48, 145, 263
- Binomialverteilung, 93, 124
- Block-Diagonal-Struktur, 35
- Blockmatrix, 19, 36, 62, 180
- Borelsche Menge, 89
- Bruchpunkt, 287, 292

- Charakteristische Funktion**, 115
 - Gleichung, 49
- χ^2 -Anpassungstest, 296
 - Verteilung, 134, 137, 148, 297, 313, 329
- Cholesky-Faktorisierung, 33, 51, 68, 167, 188

Datum, 203

- stransformation, 208
- Determinante, 40, 78, 154
- Diagonalelement, 16, 26, 51
- Diagonalmatrix, 16, 30, 49, 113, 168
- Dichte, 91, 92, 97, 117, 121, 125, 127, 134, 139, 148
 - funktion, 91
- Differentialquotient, 74
- differentielle Drehung, 47, 204
- Differenzmenge, 4, 84
- Dimension eines Vektorraums, 11
 - en einer Matrix, 15
- direkte Summe, 9, 69
- disjunkt, 4
- diskrete Verteilung, 93
 - Zufallsvariable, 91
- Dispersion, 105
 - smatrix, 108
- Distributivgesetz, 6, 8, 17
- doppelt nichtzentrale F -Verteilung, 143
- Drehmatrix, 47
- dreidimensionaler Raum, 5, 203
- Dreiecksmatrix, 16, 26, 30, 33, 51, 135, 167
- Durchschnittsmenge, 4, 84

- E**ffekt, 217, 221, 225, 310
effiziente Schätzfunktion, 162
Eigenvektor, 48, 197, 325
Eigenwert, 48, 50, 53, 128, 197, 317, 325
eindimensionale Zufallsvariable, 89
einfache Klassifikation, 218
Einflußfunktion, 279
Einheits-Dreiecksmatrix, 16, 27, 30
Einheitsmatrix, 16
einseitiger Hypothesentest, 313
Element einer Matrix, 15
- einer Menge, 3
elementare Matrix, 23, 26
- Umformung, 23, 26, 41
Elementarereignis, 83, 85, 88, 96
empirischer multipler Korrelationskoeffizient, 312
Ereignis, 83
Ergebnisraum, 83
ergodischer Zufallsprozeß, 273
erschöpfende Schätzfunktion, 162
erwartungstreue Schätzung, 163, 169, 171, 185, 200, 210, 228, 233, 262, 276
erwartungstreu schätzbare Funktion, 199
Erwartungswert, 102, 104, 118, 144, 161, 166
Euklidischer Raum, 13, 172, 203
Extremum, 72
- F**aktor, 217, 221, 225, 310
Falksches Schema, 17
Fehler, 166, 241, 245
- 1. Art, 301, 303, 305
- 2. Art, 301, 303, 305, 313, 315
-fortpflanzungsgesetz, 109
-gleichung, 167
-hyperellipsoid, 325
-wahrscheinlichkeit, 302
fester Effekt, 218
- Parameter, 163, 165, 231
Filterung, 240
Fisher-Verteilung, 138
Fraktal, 120, 136, 140, 144, 157, 159, 305, 308, 313, 315, 334
freies Netz, 203

Freiheitsgrad, 134, 137, 139, 141, 143
führende Hauptuntermatrix, 28
Funktionaldeterminante, 78, 128
 F -Verteilung, 138, 141, 144, 156, 158, 298, 305, 330

Gamma-Funktion, 122

-verteilung, 121, 124, 135

Gauß-Helmert-Modell, 231, 254

Gauß-Jordan-Methode, 31

Gauß-Markoff-Modell, 166, 184, 192, 196, 209, 226, 246, 296, 304

- mit nicht vollem Rang, 196, 209, 226, 296, 304

- - und Restriktionen, 209, 304

- mit Restriktionen, 184

Gaußsche Elimination, 26, 37, 179, 190, 236, 268

-r Algorithmus, 26

gegenseitig sich ausschließende Ereignisse, 82, 84, 85, 91

gemischtes Modell, 231, 233, 239, 240, 245, 296

generalisierte Inverse, 54, 57, 60, 62, 66, 68, 197, 201, 251, 297

geordnetes Paar, 5

Gewicht, 113, 168, 177, 182, 278

-skoeffizient, 167

-smatrix, 113, 166, 195, 226, 246, 259, 332

-sreziproke, 167

gleichmäßig bester unverzerrter Test, 302

- bestes Konfidenzintervall, 323

Gleichungssystem, 30, 59, 68

gon, 114, 206

Grundgesamtheit, 162

Grundraum, 83

Güte eines Tests, 302

Halbwertsbreite, 274

Hebelpunkt, 287, 289, 292

hierarchische Klassifikation, 220

Histogramm, 82

höhere Klassifikation, 221

homogener Zufallsprozeß, 273

homogenes Gleichungssystem, 59, 60

homogenisiertes Residuum, 333

Homogenisierung, 168

Homoskedastizität, 167

Hypothese, 301, 304, 307, 308, 314, 329

-ntest, 304, 312, 329

idempotente Matrix, 53, 55, 70, 146, 174, 198, 297

Identifizierbarkeit, 202

Infimum, 72

inneres Produkt, 12

innere Zuverlässigkeit, 335

Invarianzeigenschaft der Eigenwerte, 49, 318

inverse Matrix, 18, 26, 37, 54, 188

-s Element, 6

isotroper Zufallsprozeß, 273

iterierte Schätzung, 254

Jacobische Determinante, 78

- Matrix, 78, 101

Kalman–Bucy–Filter, 193

kartesisches Produkt, 5

Klassifikation, 218

Koeffizientenmatrix, 30, 59, 166, 196, 231, 259

Kofaktor, 167

Kollokation, 244

Kolmogoroff–Smirnow–Test, 296

Kombination, 39

Kommutativgesetz, 6, 16

komplementäres Ereignis, 84, 85, 94

Komplementärmenge, 4

Komponente, 7

Konfidenzgrenze, 323

Konfidenzhyperellipsoid, 325, 328

Konfidenzintervall, 323, 328

Konfidenzniveau, 323, 328

konsistente Schätzfunktion, 162

-s Gleichungssystem, 59, 166, 251

Koordinaten, 5, 7, 12, 182, 203

-achsen, 5, 12, 46

-transformation, 45

Körper der reellen Zahlen, 6

Korrelationskoeffizient, 106, 110, 230, 312

Korrelationsmatrix, 110, 230

korreliert, 106, 271, 273

Kosten, 305

- Kovariable, 225
- Kovarianz, 105, 107, 111, 145, 252, 263, 267
 - analyse, 225
 - funktion, 274
 - komponente, 246, 253, 264, 272
 - matrix, 107, 128, 144, 152, 165, 166, 260, 301
- kritischer Bereich, 301
- Kronecker-Produkt, 20

- L**agrangesche Funktion, 73, 170, 186, 235, 249
 - r Multiplikator, 73, 170, 186, 235, 249
- Länge eines Vektors, 12
- Laplacescher-Entwicklungssatz, 40
- leere Menge, 3, 83
- Likelihoodfunktion, 165, 174, 262, 278, 303
- Likelihood-Quotientenkriterium, 314, 320
 - Quotiententest, 303, 314, 321
- lineare Abhängigkeit, 9, 48, 65
 - Regression, 228
 - Schätzung, 169, 185, 200, 210, 228, 233, 262, 276
 - Transformation, 45, 108, 133
 - Unabhängigkeit, 10, 13, 21, 61, 63
 - r Vektorraum, 8
 - s Gleichungssystem, 30, 59, 68
- Linearisierung, 108, 168
- Linksinverse, 52, 58, 61, 64
- LMS-Methode, 277, 292
- lokal beste Schätzung, 249, 253
 - es Extremum, 72
- Lösung eines Gleichungssystems, 30, 59
 - sraum, 60, 64
- L_1 -Norm-Schätzung, 285, 293
- L_p -Norm-Schätzung, 284

- M**acht eines Tests, 302
- Maßstabsfaktor, 203
- Matrix, 15
- Matrizenaddition, 16
- Matrizengleichung, 44
- Matrizenidentität, 36
- Matrizenmultiplikation, 16
- Maximalwurzel-Kriterium, 319

- Maximum, 72
 - Likelihood-Methode, 165, 174, 187, 197, 235, 262, 278
- Median, 286
- mehrdimensionale Zufallsvariable, 89, 96
- Menge, 3
- Merkmal, 217, 259, 316
 - svektor, 260
- Methode der kleinsten Quadrate, 165, 171, 187, 197, 234, 262, 277, 285
- Minimum, 72
- MINQUE-Verfahren, 249
- mittlerer Fehler, 105, 109
- Modell, 161, 166, 184, 196, 214, 226, 246, 304
 - I, 218
 - II, 218
- modifizierte M-Schätzung, 289
- Moment, 104, 115, 126, 155
 - erzeugende Funktion, 115, 122, 129, 135, 137
- Moore-Penrose-Inverse, 57
- M-Schätzung, 279
 - nach Huber, 281, 289
- multipler Korrelationskoeffizient, 230, 312
- multivariate Normalverteilung, 118, 127
 - Varianzanalyse, 259
 - Verteilung, 97
 - s Gauß-Markoff-Modell, 260, 275, 314
 - - mit Restriktionen, 275
 - s Moment, 104, 115
- N**äherungswert, 168
- natürliche Zahl, 4
- n -dimensionale Zufallsvariable, 96
 - r Raum, 5, 13
- nichtzentrale χ^2 -Verteilung, 137, 297
 - F -Verteilung, 141, 298
 - Wishart-Verteilung, 148, 153, 159, 300, 315
- Norm eines Vektors, 12, 284
- Normalgleichungen, 173, 179, 187, 197, 217, 235, 270
- Normalgleichungsmatrix, 179, 187, 216, 235, 268
- Normalverteilung, 116, 127, 134, 145, 148, 175, 235, 247, 261, 281, 296, 299
- n -Tupel, 5
- Null, 6
 - hypothese, 301, 304, 308, 315, 329

- matrix, 15
- raum, 38, 60, 64, 203, 222
- vektor, 7

- O**bere Dreiecksmatrix, 16, 26, 30
 - r α -Prozentpunkt, 120
- Orientierungsunbekannte, 205
- orthogonal, 13, 38, 46, 49, 70, 216
 - basis, 13
 - polynom, 216, 310
 - e Matrix, 46, 49, 128, 325
 - e Projektion, 70, 172
 - e Transformation, 46, 50, 318
 - er Projektionsoperator, 70, 173, 201, 250
 - er Unterraum, 14
 - es Komplement, 14, 38, 70
 - es Koordinatensystem, 13, 47, 325
- Orthonormalbasis, 13, 47

- P**arameter, 161, 162, 166, 184, 197, 214, 217, 228, 231, 260, 278, 304, 314
 - raum, 162, 301, 314
 - schätzung, 161, 162, 165, 192, 196, 226, 259, 277
- partieller Korrelationskoeffizient, 230
- Permutation, 39
- Pivotstrategie, 28, 66
- Poisson-Verteilung, 95, 124
- Polynommodell, 214, 310
- positiv definit, 50, 68, 109, 127, 149, 166, 194, 210, 246, 260, 300, 322
 - semidefinit, 50, 54, 68, 109, 151, 194, 300
- Posteriori-Wahrscheinlichkeit, 87
- Prädiktion, 240
- Priori-Wahrscheinlichkeit, 87
- Profil einer Matrix, 34
- Projektion, 70, 172, 200
 - soperator, 70, 173, 200
- projizierter Parameter, 199, 200, 222, 280, 308
- Pseudoinverse, 57, 63, 67, 68, 201, 222
- Punktschätzung, 162

- Q**uadratische Form, 48, 50, 144, 146, 152, 172, 248, 297, 304
 - Matrix, 15
 - s Programmierproblem, 185

Quadratsumme der Residuen, 172, 180, 190, 197, 211, 297
Quadrupel, 5
qualitative Aussage, 217, 225
Quantil, 120
quantitative Aussage, 166, 225

Randomisierung, 220
Randverteilung, 98, 130, 139, 141, 143
-sfunktion, 98
Rang einer Matrix, 21, 26, 43
-defekt, 22, 196, 204, 222
-faktorisierung, 26, 53, 58
-raum, 37
Rechtsinverse, 52, 58
reelle Zahl, 4, 6
reflexive generalisierte Inverse, 57, 62, 66, 68, 201
Regression, 166
-sanalyse, 166
-smodell, 228
reguläre Matrix, 18, 26, 36, 43, 50, 53, 63, 187, 216
rekursive Parameterschätzung, 192
Relation, 5
relative Häufigkeit, 82, 102, 121
Reparameterisierung, 200
reproduzierende Eigenschaft, 123, 134, 151
Residuum, 172, 175, 189, 197, 210, 228, 263, 280, 296, 331
Richtungskosinus, 47, 325
robuste M-Schätzung, 278, 283
- Parameterschätzung, 277, 286
Rückrechnung, 27, 37, 179, 268

Schätzbare Funktion, 199, 200, 209, 222, 308
Schätzung einer Kovarianzmatrix, 263, 272
Scheinvariable, 225
schiefsymmetrische Matrix, 18
schwach besetzte Matrix, 35
sicheres Ereignis, 84, 85, 91
Signal, 241
Signifikanzniveau, 302, 305, 311, 315, 333
Simplex-Algorithmus, 285
singuläre Matrix, 18, 42, 174, 189, 196, 235, 239
Skalar, 8

- multiplikation, 8, 16
- produkt, 12, 48, 71, 250
- Spalten einer Matrix, 15
 - raum, 37, 59, 71, 172, 250
 - vektor, 15
- Sparse-Matrix, 35
- Spur einer Matrix, 44, 68, 171
 - kriterium, 318, 320
- Standardabweichung, 105, 115, 258, 270, 328, 332
- standardisierte Normalverteilung, 118
 - s Residuum, 334
- stationärer Zufallsprozeß, 273
 - Punkt, 73
- stetige Zufallsvariable, 92, 97, 116
- Stichprobe, 162, 301
- stochastischer Prozeß, 273
- S-Transformation, 208
- Student-Verteilung, 143
- studentisiertes Residuum, 332
- Stufe, 218, 221, 310
- suffiziente Schätzfunktion, 162
- Supremum, 72
- symmetrische Matrix, 18, 32, 48, 50, 53, 62, 75, 247
 - reflexive generalisierte Inverse, 57, 62, 66, 68, 201
 - s Gleichungssystem, 61
- systematischer Anteil, 240

- T**aylor-Entwicklung, 72, 108, 169
 - Karman-Struktur, 273
- Teilmenge, 4
- Teilredundanz, 288, 291, 332
- Test, 301
 - größe, 301, 304, 308, 314, 321
 - statistik, 301, 312, 318
 - verfahren, 301
 - verteilung, 134, 148
- Transformation, 45, 50, 108, 133, 318
- Translation, 204
 - s-Invarianz, 248
 - svektor, 45
- Transponierung einer Matrix, 18
- Trendanteil, 240

Trennschärfe, 302, 305, 311
trennscharfer unverzerrter Test, 302
Tripel, 5
Tschebyscheff–Approximation, 285
 t -Verteilung, 143

Umfang eines Tests, 302
unabhängige Ereignisse, 88, 100
- Zufallsvariable, 100, 105, 113, 132, 147
unbekannter Parameter, 161, 162, 166, 184, 197, 214, 217, 228, 231, 260
univariate Verteilung, 92, 116, 121, 125
-s Gauß–Markoff–Modell, 259, 304
unkorreliert, 106, 167, 192, 215, 221, 260
unmögliches Ereignis, 84, 85, 91
untere Dreiecksmatrix, 16, 30, 33, 51, 135, 167
-r α -Prozentpunkt, 120
Unterraum, 8, 13, 38, 69
unverzerrte Schätzung, 163
-r Test, 302
unvollständige Betafunktion, 126, 139
-r Versuchsplan, 219
-s multivariates Modell, 271

Vandermonde–Matrix, 29
Variablentransformation, 78, 101, 128, 139, 141, 150
Varianz, 105, 107, 118, 164, 167, 170, 200, 253, 260, 267, 331
-analyse, 218, 221, 310
-komponente, 246, 253, 264, 272
- der Gewichtseinheit, 167, 175, 189, 198, 212, 258, 263, 297, 312, 328
Vektor, 6
-addition, 7
-raum, 8, 10, 12, 14, 70
vektorieller Zufallsprozeß, 273
Venn–Diagramm, 4, 86
verallgemeinertes Skalarprodukt, 48, 71, 250
Verbesserung, 172
-sgleichung, 167
Vereinigungsmenge, 4, 84
Verknüpfung von Mengen, 4
Versuchsplan, 218, 222, 225
-matrix, 218, 222, 225
Verteilung, 91, 93, 97, 101, 116, 120, 127, 134, 138, 143, 148, 154

- sfunktion, 90, 92, 97, 118, 123, 126, 135, 138, 139, 142
- Vertrauensintervall, 323
- Verzerrung, 163
- voller Spaltenrang, 22, 51, 52, 146, 166
 - Zeilenrang, 22, 52
- vollständiger Versuchsplan, 219
- W**ahrscheinlichkeit, 81, 85, 88, 91, 97, 109, 302, 305, 311, 315, 322, 333
 - sdichte, 91
 - sraum, 85, 88, 96, 162
 - sverteilung, 91
- Wechselwirkung, 220, 222
- Wiener–Hopf–Integralgleichung, 244
- wirksamste Schätzfunktion, 162
- Wishart–Matrix, 148, 152, 154, 158
 - Verteilung, 148, 152, 154, 159, 299, 315, 321
- Z**eilen einer Matrix, 15
 - vektor, 15
- zentraler Grenzwertsatz, 118
- zentrales Moment, 105
- zufälliges Ereignis, 84, 88, 90, 96
- Zufallsparameter, 164, 228, 231, 241
- Zufallsprozeß, 273
- Zufallsvariable, 88, 90, 91, 93, 96, 99, 100, 102, 104, 117, 121, 125
- Zufallsvektor, 97, 101, 105, 111, 127, 134, 148, 162, 166, 197, 228, 245, 260
- Zuverlässigkeitsmaß, 335
- zweifache Klassifikation, 220
- zweiseitiger Hypothesentest, 313