Friedrich Pukelsheim

Punktschätzungen
in linearen statistischen Modellen

*

Dissertation
zur Erlangung des Doktorgrades
der Mathematischen Fakultät der Albert-Ludwigs-Universität

Freiburg im Breisgau
Februar 1977

Dekan: Professor Dr. H. Klingen
Referent: Professor Dr. H. Witting
Korreferent: Professor Dr. O. Krafft
Datum der Promotion:

Ich danke Herrn Professor Dr. H. Witting für die anregende und hilfreiche Betreuung, mit der er mich bei der Anfertigung dieser Arbeit unterstützt hat.

Mein Dank gilt auch den Herren Professoren Dr. G.P.H. Styan, Montreal, und Dr. O. Krafft, Aachen, mit denen ich über Teil III beziehungsweise Teil IV dieser Arbeit diskutieren konnte.

Schließlich danke ich der Friedrich-Ebert-Stiftung, Bonn, die die Promotion mit einem Stipendium gefördert hat.

Friedrich Pukelsheim

Albert-Ludwigs-Universität
Institut für Mathematische Stochastik
Hermann-Herder-Straße 10
D-7800 Freiburg im Breisgau

Inhaltsverzeichnis

Seite

Einleitende Zusammenfassung 4

I. LINEARE MODELLE
1. Definition und Beispiele 13
2. Die Streuungs-Mittelwert-Korrespondenz 18
3. Gemeinsame Problemstellung linearer und multilinearer Schätzverfahren 22

II. SPEZIELLE LINEARE UND MULTILINEARE SCHÄTZVERFAHREN
1. Schätzung des Mittelwerts 31
2. Schätzung der Varianz-Kovarianz-Komponenten 34
3. Schätzung der Schiefe 40
4. Schätzung des Exzeß 46

III. NICHTNEGATIV-DEFINITE SCHÄTZUNG DER STREUUNGSMATRIX
1. Der Fundamentaldefekt linearer Modelle 53
2. Eine hinreichende Bedingung zur nichtnegativ-definiten Schätzung der Streuungsmatrix 55

IV. NICHTNEGATIVE SCHÄTZER IN VARIANZKOMPONENTEN-MODELLEN ALS LÖSUNGEN KONVEXER PROGRAMME
1. Erwartungstreue nichtnegative Schätzbarkeit 63
2. Negativität eliminierende Projektoren 70
3. Minimierungsprogramme 75
4. Beste nichtnegative Schätzer als Lösungen konvexer Programme 79
5. Abschliessender Ausblick 89

Literatur- und Autorenverzeichnis 93
Symbolliste 96
Sachregister 98

EINLEITENDE ZUSAMMENFASSUNG

Über lineare statistische Modelle gibt es umfangreiche und bekannte Lehrbücher wie die von Scheffé (1959), Searle (1971) oder C.R. Rao (1975): dieses Gebiet beschäftigt in gleichem Maße die Mathematische wie die Angewandte Statistik, und wird seinerseits von dem so gezeugten Interesse in Bewegung gehalten.

In der vorliegenden Arbeit beschränken wir uns auf die Punktschätzung von Mittelwert, Varianz-Kovarianz-Komponenten, Schiefe und Exzeß, und darunter auf solche Verfahren, die durch ihre finiten Eigenschaften: Erwartungstreue, Nichtnegativ-Definitheit, minimale Norm etc. charakterisiert werden; ausgeklammert haben wir asymptotische Fragestellungen, Testtheorie und weitgehend auch eine Normalverteilungsannahme. Dementsprechend entstammen unsere Hilfsmittel der linearen Algebra und der konvexen Analysis.

Der inhaltliche Aufbau ist darauf ausgerichtet, das Gemeinsame der Problemstellungen und die Einheitlichkeit der Problemlösungen deutlicher hervorzuheben als in den genannten Lehrbüchern: die untersuchten Schätzverfahren sind im Grunde Variationen ein und desselben Themas.

TEIL 1: LINEARE MODELLE

Kapitel 1 (***Definition und Beispiele***) beginnt mit einem anschaulichen Beispiel, das zum allgemeinen Begriff eines linearen Modells hinführt; anschließend machen wir deutlich, daß die gewählte Definition alle in diesem Rahmen üblichen Modelle erfaßt. Gemäß unserer Einstellung werden Mittelwertvektor und Streuungsmatrix schon in der Definition gleichartig behandelt: Ein lineares Modell $Y \sim (\sum_{\pi=1}^{p} b_\pi x_\pi;\ \sum_{\kappa=1}^{k} t_\kappa V_\kappa)$ besteht aus einem $\mathbb{R}^n$-wertigen Zufallsvektor Y, für den lineare Zerlegungen sowohl für den Erwartungswert wie auch für die Streuungsmatrix gelten:

$$\mathcal{E}\ Y = \sum b_\pi x_\pi \ ; \quad \mathcal{D}\ Y = \sum t_\kappa V_\kappa \ ;$$

dabei werden die p zerlegenden Vektoren x_π und die k zerlegenden symmetrischen Matrizen V_κ als bekannt vorausgesetzt; die Vektorparameter $b := (b_1,\dots,b_p)' \in \mathbb{R}^p$ und $t := (t_1,\dots,t_k) \in \mathbb{R}^k$ gilt es zu schätzen.

In Kapitel 2 (*Die Streuungs-Mittelwert-Korrespondenz*) erläutern wir, wie die quadratische Schätzung des Streuungsparameters t und die lineare Schätzung des Mittelwertparameters b als verschiedene Ansichten desselben Problems erscheinen; der Tensorkalkül der multilinearen Algebra gibt mit dem Kroneckerprodukt das Mittel an die Hand, quadratische Schätzfunktionen mit linearen zu identifizieren: ***invariante quadratische Streuungsschätzung im ursprünglichen Modell ist dasselbe wie lineare Mittelwertschätzung in einem geeignet abgeleiteten Modell.***

Kapitel 3 (*Gemeinsame Problemstellung linearer und multilinearer Schätzverfahren*) resümiert diejenigen Eigenschaften, die etwas über die Güte von Schätzverfahren aussagen: Gauss'sche Risikofunktion der mittleren quadratischen Abweichung, Reduktionen durch Linearität, Invarianz und Erwartungstreue, Minimierung von Norm oder Varianz. In §§6f werden die hier benutzten Optimalitätskonzepte und ihre Abkürzungen aufgelistet und mit den Bezeichnungen anderer Autoren verglichen.

Wichtig zum Verständnis der folgenden Teile sind besonders die Definition 1.3, Kapitel 2 und die Bezeichnungen 1.2 und 3.13.

Wir erwähnen nun die wesentlichsten Verbindungen zur Literatur. Kapitel 3 wurde in Anlehnung an Kapitel 1 in Witting (1966) formuliert. Die Definition 1.3 eines linearen Modells enthält inhaltlich das Übliche, formal fördert sie die durch die Streuungs-Mittelwert-Korrespondenz begründete Anschauung. Auch Seely (1970) führt mit einer "koordinatenfreien" Schätztheorie die Streuungsschätzung auf die Mittelwertschätzung zurück, einen uns ähnlicheren Ansatz macht Mitra (1971); unabhängig davon habe ich in meiner Diplomarbeit (1974) die Streuungs-Mittelwert-Korrespondenz entwickelt. Beide Formulierungen – die Seely'sche wie die unsrige – eröffnen grundsätzlich dieselben Möglichkeiten; für deren Verwirklichung erscheint mir aber die Streuungs-Mittelwert-Korrespondenz handlicher und überschaubarer. Beispielsweise läßt sie sich eher fortsetzen – wir tun dies in Teil II –, um auch die Schätzung von Schiefe und Exzeß zu erfassen.

TEIL II: SPEZIELLE LINEARE UND MULTILINEARE SCHÄTZVERFAHREN

Ziemlich ausführlich begründen wir in Kapitel 1(*Schätzung des Mittelwerts*) die Schätzverfahren für den Mittelwertparameter b, da diese Aussagen den Unterbau für die folgenden drei Kapitel hergeben.

Zusammen mit der Streuungs-Mittelwert-Korrespondenz führen sie in Kapitel 2 (*Schätzung der Varianz-Kovarianz-Komponenten*) auf einfache Weise zu einer abgerundeten Theorie der Streuungsschätzung. Deren Wirksamkeit wird an vier Modellen der Varianzanalyse verdeutlicht; schließlich vertreten wir die Überlegenheit des hier vorgetragenen Ansatzes gegenüber der Art, wie die klassischen Henderson'schen Verfahren gerechtfertigt werden.

Für die Kapitel 3 und 4(*Schätzung der Schiefe und des Exzess*) wählen wir das einfachere Modell gleicher Varianzen und unabhängiger Beobachtungskomponenten $Y \sim (\sum b_\pi x_\pi \;;\; \sigma^2 I_n)$. Der Gebrauch transformierter Modelle – in der Streuungs-Mittelwert-Korrespondenz erfolgreich erprobt – führt auch hier auf raschem Weg zur Bestimmung bester Schätzer unter den homogenen Polynomen dritten beziehungsweise vierten Grades.

Die Kapitel 1 und 2 bereiten Teil III vor; Teil IV stützt sich nur auf §§1.1f und 2.1f.

Für Beweise zu den Ergebnissen in den Kapiteln 1 und 2 wird auf die Literatur verwiesen. Bis zu §2.4 folgen wir meiner Diplomarbeit (1974); Seely's Satz über die Existenz von erwartungstreuen Streuungsschätzern mit gleichmäßig kleinster Varianz wird angeschlossen. Mit Kapitel 2 endet der expositorische Teil der Arbeit.

Die Reduktion multilinearer Schätzverfahren auf lineare ist so allgemein, daß sich ihre Verwendung zur Schätzung auch von Schiefe und Exzeß geradewegs anbietet; trotzdem ist diese Idee neu. Sie führt zu einfachen Kriterien und gestattet, die von

Anscombe (1961) heuristisch angegebenen Schätzer in die allgemeine Theorie einzugliedern.

TEIL III: NICHTNEGATIV-DEFINITE SCHÄTZUNG DER STREUUNGSMATRIX

Ausgehend von unserer Definition linearer Modelle schildern wir in Kapitel 1 (***Der Fundamentaldefekt linearer Modelle***) einen Makel der herkömmlichen Theorie, den wir mit dem Begriff ***Fundamentaldefekt*** umschreiben. Beispielsweise können bei der Streuungsschätzung nur solche Schätzwerte $\hat{t}$ hingenommen werden, die zu einer nichtnegativ-definiten Schätzung $\sum\hat{t}_\kappa V_\kappa$ der Streuungsmatrix führen; diese Einschränkung wird von der bisher entwickelten Theorie einfach mißachtet.

In Kapitel 2(***Eine hinreichende Bedingung zur nichtnegativ-definiten Schätzung der Streuungsmatrix***) führt eine zeitweilige Normalverteilungsannahme zu einer Teillösung: Gemäß Seely (1971) existiert bei Normalverteilungsannahme eine vollständige suffiziente Statistik für t, falls die k zerlegenden Matrizen V_κ einen quadratischen Teilraum symmetrischer Matrizen aufspannen. Wie wir ergänzend beweisen, fallen dann "alle" Schätzverfahren zusammen: der Varianzanalysen-Schätzer, C.R. Rao’s MINQUE, Seely's erwartungstreuer Schätzer mit gleichmäßig kleinster Varianz und Corbeil & Searle's Restricted Maximum Likelihood Schätzer. In dieser paradiesischen Situation erhält man auch fast sicher eine nichtnegativ-definite Schätzung der Streuungsmatrix.

Unser Ergebnis liefert für die von Corbeil & Searle (1976b) durchgerechneten Beispiele die theoretische Begründung nach, warum die beobachtete Gleichheit der Schätzer notwendig eintreten muß. Ebenso charakterisieren wir eine Lage, in der die Likelihood-Gleichung gelöst und auf iterative Verfahren verzichtet werden kann (Anderson 1970; Corbeil & Searle 1976a). Den Begriff "Fundamentaldefekt" haben wir schon in der Diplomarbeit (1974) geprägt, ohne ihn dort näher zu untersuchen.

TEIL IV: NICHTNEGATIVE SCHÄTZER IN VARIANZKOMPONENTEN-MODELLEN ALS LÖSUNGEN KONVEXER PROGRAMME

Von einem ***Varianzkomponenten-Modell*** sprechen wir, wenn in einem linearen Modell die k zerlegenden Matrizen V_κ nichtnegativ-definit sind und t – als Vektor von Varianzen $t_\kappa = \sigma^2_\kappa$ – im abgeschlossenen positiven Quadranten $\mathbb{R}^k_+$ variiert. Auch hier droht der Fundamentaldefekt: Schätzwerte $\hat{t}$ außerhalb des $\mathbb{R}^k_+$ sind Unsinn. Wir machen uns deshalb auf die Suche nach erwartungstreuen nichtnegativ-definiten quadratischen Schätzfunktionen mit minimaler Norm (***beste nichtnegative Schätzer***}, um die – möglicherweise defekten – erwartungstreuen invarianten quadratischen Schätzfunktionen mit minimaler Norm (***beste defekte Schätzer, MINQUE***) abzulösen; wir werden dieses Ziel erreichen und uns dabei hauptsächlich auf die Theorie konvexer Programme stützen.

Das Thema von Kapitel 1 (***Erwartungstreue nichtnegative Schätzbarkeit***) lautet in der Sprache solcher Programme: Wann existieren zulässige Lösungen? Wir zeigen, daß für eine einzelne Varianzkomponente t_κ genau dann eine erwartungstreue nichtnegativ-definite quadratische Schätzfunktion existiert, wenn die zugehörige zerlegende Matrix V_κ einen echten Beitrag zur Erklärung der Streuung liefert.

In Kapitel 2 (***Negativität eliminierende Projektoren***) veranschaulichen wir einige Begriffe, um das Primal- und Dualprogramm und deren optimale Lösungen anzugeben und zu interpretieren: Q-reduzierte Modelle, Negativität eliminierende Projektoren, Positiv- und Negativteil einer symmetrischen Matrix.

Mit Kapitel 3 (***Minimierungsprogramme***) bieten wir einen kurzen Einstieg in die Theorie mathematischer Programme. Wir möchten damit einerseits einen zu abrupten Übergang von linearen zu konvexen Methoden vermeiden und andererseits ein ähnliches Vorgehen in der Testtheorie erfassen, das den hier eingeschlagenen Weg wesentlich bestimmt hat.

Nach diesen Vorbereitungen gelingt uns in Kapitel 4 (***Beste nichtnegative Schätzer als Lösungen konvexer Programme***) die konstruktive Bestimmung bester nichtnegativer Schätzer: zum ersten erhält man sie aus dem besten defekten Schätzer durch eine passende Korrektur von dessen defektem Teil, und zum zweiten ergeben sie sich als beste defekte Schätzer, wenn man das ursprüngliche Modell mit einem Negativität eliminierenden Projektor reduziert. Die erste Art gründet ausschließlich auf den Formulierungen des Kapitels 3, für die zweite ziehen wir zusätzlich die Fenchel'sche Dualitätstheorie konvexer Programme zu Rate.

Im rückschauenden Kapitel 5 (*Abschließender Ausblick*) gehen wir noch einmal der uns leitenden Anschauung nach, daß die Streuungs-Mittelwert-Korrespondenz die volle Parität von Mittelwert- und Streuungsschätzung sichert: Dies führt zu besten kegelerhaltenden Schätzfunktionen und damit zu einer Klasse linearer Modelle, für die Teil IV den Weg weisen dürfte, den Fundamentaldefekt vollständig zu beheben.

Konvexe Programme sind bei linearen Modellen bisher lediglich verwendet worden, um Algorithmen zur Berechnung einzelner Schätzwerte zu finden; in diese Richtung gehen auch C.R. Rao's Worte (1972 pp.113f): *"If non-negative estimators are desired, one may have to minimize (the norm) subject to the condition that (the estimator) is non-negative definite in addition to (invariance and unbiasedness), if possible. The computational problem is likely to be difficult, but a suitable computer program may be written."*

Der Einsatz mathematischer Programme, um beste nichtnegative Schätzer als optimale Lösungen konvexer Programme darzustellen, ist neu. Dabei haben wir nur zusammengesetzt, was verschiedene Lehrbücher schon bereitstellen: Die Bedeutung des Problems unterstreicht Searle (1971 pp.406-408), C.R. Rao's MINQUE (1973 pp. 303-305) legt seine Formalisierung nahe; die Lösungsidee stammt aus der Testtheorie (Witting 1966 pp.69-73), Rockafellar (1970) ebnet den Weg zu ihrer Verwirklichung.

Um den Formelkalkül so gering wie möglich zu halten, bevorzugen wir einen redundanten verbalen Stil. Wir halten uns (mit Ausnahmen) an die folgende Summationskonvention: Bei einem Summationszeichen ohne Summationsindizes wird über *alle* griechischen Indizes summiert, zwar von 1 bis zu ihrem lateinischen Äquivalent. In dieser Arbeit sind alle Vektorräume reelle Vektorräume, alle Matrizen reelle Matrizen und alle Projektoren orthogonale Projektoren. Eckige Klammern bei Blockvektoren und -matrizen erinnern daran, daß die Einträge selbst wieder Vektoren und Matrizen sind. Nichtnegativ-definit wird mit NND abgekürzt. Das Zeichen := bedeutet definierende Gleichheit. Ein fetter Punkt kennzeichnet das Ende eines Beweises:

●

Teil I

LINEARE MODELLE

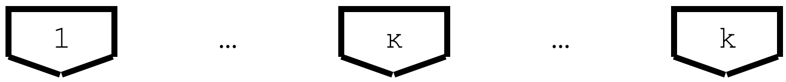
Kapitel 1

DEFINITION UND BEISPIELE

Nach einem einfachen Beispiel und einigen Bezeichnungen werden in diesem Kapitel lineare Modelle allgemein definiert. An Hand klassischer Beispiele wird danach deutlich gemacht, daß die Definition alle in diesem Rahmen üblichen Modelle erfaßt.

1 BEISPIEL. ***Heteroskedastische Varianzen***. Gegeben seien k Produktionseinheiten. Jede Einheit κ erzeuge n_κ viele Erträge (yields) $y_{\kappa\nu}$ $(\nu = 1,\dots,n_\kappa)$, die durch reelle Zahlen repräsentiert werden: $y_{\kappa\nu} \in \mathbb{R}$. Um die dem Produktionsprozeß innewohnenden Schwankungen wiederzugeben, werden die $y_{\kappa\nu}$ als Realisierung einer Zufallsgröße $Y_{\kappa\nu}$ aufgefaßt.

Die Produktionseinheiten seien so eingerichtet, daß ihre Erträge im Mittel gleich ausfallen: $\mathcal{E}\, Y_{\kappa\nu} = \mu$; ferner sei gestattet anzunehmen, daß die Varianz nur von der Produktionseinheit κ abhängt: $\mathrm{Var}\, Y_{\kappa\nu} = \sigma_\kappa^2$; $\mathrm{COV}(Y_{\kappa\nu}, Y_{\alpha\beta}) = 0$ falls $(\kappa,\nu) \neq (\alpha,\beta)$. Dieses Modell können wir so skizzieren:



$Y_{1\nu} \sim (\mu;\sigma_1^2)$ $\quad$ $Y_{\kappa\nu} \sim (\mu;\sigma_\kappa^2)$ $\quad$ $Y_{k\nu} \sim (\mu;\sigma_k^2)$

$\nu = 1,\dots,n_1$ $\quad$ $\nu = 1,\dots,n_\kappa$ $\quad$ $\nu = 1,\dots,n_k$

Es ist bequem, die $n := \sum n_\kappa$ Zufallsgrößen $Y_{\kappa\nu}$ lexikographisch anzuordnen zu $Y := (Y_{11}, Y_{12}, \dots, Y_{1n_1}, \dots, Y_{k1}, \dots, Y_{kn_k})'$. Mit dem Vektor $1_n := (1,\dots,1)' \in \mathbb{R}^n$, der $n_\kappa \times n_\kappa$ Einheitsmatrix I_{n_κ} und der suggestiven Notation von k Blockdiagonalmatrizen lautet die Vektorform des vorliegenden Modells:

$$\mathcal{E}\, Y = 1_n\mu \ ; \qquad \mathcal{D}\, Y = \sum \sigma_\kappa^2 \,\mathrm{Diag}[0 : I_{n_\kappa} : 0] \ .$$

Dies ist ein Spezialfall eines gleich zu definierenden linearen Modells und eines Varianzkomponenten-Modells. Wir setzen das Beispiel in §3.12 fort.

2 Für einen $\mathbb{R}^n$-wertigen Zufallsvektor Y bezeichnet $\mathcal{E}$ Y den Erwartungswert und $\mathcal{D}$ Y die Streuungsmatrix (Varianz-Kovarianz-Matrix, dispersion matrix).

$\mathbb{R}^{n\times p}$ bezeichnet den Vektorraum aller n×p-Matrizen. Im Raum $\mathbb{R}^{n\times n}$ der quadratischen Matrizen sei Sym(n) der Teilraum aller symmetrischen Matrizen. Ein Projektor ist eine symmetrische und idempotente Matrix, die Menge aller n×n-Projektoren wird mit Proj(n) bezeichnet. Für eine Matrix A sei $\mathcal{R}$ A (range von A) der von ihren Spalten aufgespannte Bildraum; $\mathcal{N}$ A bezeichne den Nullraum. A' ist die zu A transponierte Matrix.

Für einen Teilraum $\mathcal{L}$ des ist Proj($\mathcal{L}$) der Projektor mit Bildraum $\mathcal{L}$: $\mathcal{R}$ Proj($\mathcal{L}$) = $\mathcal{L}$. Das orthogonale Komplement von $\mathcal{L}$ wird mit $\mathcal{L}^\perp$ notiert; für die orthogonalen Komplemente von $\mathcal{R}$ A ($\mathcal{N}$ A) schreiben wir auch $\mathcal{R}^\perp$A ($\mathcal{N}^\perp$A). Für eine Teilmenge $\mathcal{M}$ von Vektoren bezeichnet span $\mathcal{M}$ den von $\mathcal{M}$ aufgespannten Teilraum; für eine Matrix $X = [x_1:\ldots:x_p]$ mit Spalten x_π bedeutet das: $\mathcal{R}\,X = \text{span}\,\{x_1,\ldots,x_p\}$. Die Summe zweier Teilräume $\mathcal{L}_1$ und $\mathcal{L}_2$ ist der von ihnen erzeugte Unterraum: $\mathcal{L}_1 + \mathcal{L}_2 = \text{span}\,\mathcal{L}_1 \cup \mathcal{L}_2$.

Die folgende Definition betont die Gleichartigkeit der Annahmen für den Mittelwertvektor und die Streuungsmatrix. Die meisten Autoren benutzen statt der p zerlegenden Vektoren x_π gleich die von ihnen gebildete Matrix $X = [x1:\ldots:x_p]$.

3 DEFINITION. Ein ***lineares Modell*** $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ besteht aus einem $\mathbb{R}^n$-wertigen Zufallsvektor Y, bekannten, den Erwartungswert zerlegenden Vektoren $x_1,\ldots,x_p$ des $\mathbb{R}^n$ und bekannten, die Streuungsmatrix zerlegenden symmetrischen n×n-Matrizen $V_1,\ldots,V_k$, so daß für (unbekannte) Vektorparameter $b = (b_1,\ldots,b_p)' \in \mathbb{R}^p$ und $t = (t_1,\ldots,t_k) \in \mathbb{R}^k$ gilt:

$$\mathcal{E}\ Y = \sum b_\pi x_\pi; \quad \mathcal{D}\ Y = \sum t_\kappa V_\kappa\ .$$

Ein lineares Modell heißt ***Varianzkomponenten-Modell***, falls die k zerlegenden Matrizen V_κ NND und die k Komponenten t_κ nichtnegativ s1nd.

4 Um mit Recht von Y als Zufallsvektor sprechen zu können, sei ein (fiktiver) Meßraum $(\Omega,\mathcal{A})$ angenommen und Y als meßbare

Funktion von $(\Omega,\mathcal{A})$ in den Borelschen Stichprobenraum $(\mathbb{R}^n,\mathcal{B}^n)$ vorausgesetzt, vergleiche die ***Grundannahme der Mathematischen Statistik*** in Witting (1966 p.13). Auf dem Meßraum $(\Omega, \mathcal{A})$ sei ferner eine Klasse $\mathcal{P}$ von Wahrscheinlichkeitsmaßen gegeben; auf dem Stichprobenraum $(\mathbb{R}^n,\mathcal{B}^n)$ induziert Y dann die Verteilungsklasse $\mathcal{P}^Y$. In der Definition wurde nun gefordert

$$\mathcal{E}_{\mathcal{P}}Y \subset \text{span}\ \{x_1,\dots,x_p\}, \quad \mathcal{D}_{\mathcal{P}}Y \subset \text{span}\ \{V_1,\dots,Vk\}\ ,$$

ohne deutlich zu machen, wie diese Teilmengenbeziehungen genauer aussehen sollen. Wir wollen die folgende, in allen Arbeiten über lineare Modelle stillschweigend gemachte Präzisierung nachholen:

$$\text{span}\ \mathcal{E}_{\mathcal{P}}Y = \text{span}\ \{x_1,\dots,x_p\}\ ;$$

$$\text{span}\ \mathcal{D}_{\mathcal{P}}Y = \text{span}\ \{V_1,\dots,V_k\}\ .$$

Dies ist bei festgelegter Verteilungsklasse $\mathcal{P}^Y$ immer erreichbar, indem die zerlegenden Vektoren x_π und Matrizen V_κ passend gewählt werden. Zum Beispiel ist es sinnvoll und ohne Einschränkung möglich, die zerlegenden Matrizen V_κ nicht nur als quadratisch, sondern gleich als symmetrisch vorauszusetzen, wie dies in Definition 3 geschehen ist.

5 BEISPIEL. *Quasinormale lineare Modelle*. Im klassischen Idealfall ist Y ein normalverteilter Zufallsvektor mit unkorrelierten Komponenten von gleicher Varianz: $Y \sim N_n(\sum b_\pi x_\pi;\ \sigma_n^2 I_n)$. Die genaue Festlegung einer Verteilung wird aber erst bei der Angabe von Tests und Konfidenzbereichen benötigt, da dann die Verteilungen einer oder mehrerer Teststatistiken berechnet werden müssen (Witting 1966 pp.26-29).

In der vorliegenden Arbeit beschäftigen wir uns dagegen nur mit linearen und multilinearen Schätzfunktionen; in dieser Klasse können erwartungstreue Schätzfunktionen und darunter solche mit minimaler Varianz allein aus der Kenntnis der Momente bestimmt werden. Demzufolge werden wir (bis auf Satz III.2.4) eine Normalverteilungsannahme nur benutzen, um höhere Momente zu bestimmen. Ein lineares Modell mit derart berechneten Momenten nennen wir ein *quasinormales lineares Modell*.

Die folgenden vier Beispiele zeigen, daß alle Modelle der Varianzanalyse lineare Modelle in unserem Sinne sind. Wir verweisen pauschal auf Scheffé (1959 p.6) und die vergleichenden Erläuterungen von Searle (1971 pp.376-381).

6 BEISPIEL. ***Modelle mit festen Effekten.*** Diese Modelle werden durch die Vektorgleichung $Y = Xb+e$ definiert, wobei b der unbekannte, nichtstochastische Mittelwertparameter ist, und der stochastische Störterm e Mittelwert 0 und eine bis auf positives Vielfaches σ^2 bekannte Streuungsmatrix V besitzt. Der Störterm e wird dabei als hypothetische, nicht beobachtbare Zufallsvariable aufgefaßt; in unserer Schreibweise $Y \sim (Xb;\sigma^2 V)$ kommt er nicht vor.

7 BEISPIEL. ***Modelle mit zufälligen Effekten.*** In diesem Fall lautet die grundlegende Vektorgleichung $Y = 1_n\mu+\sum U_\kappa\xi_\kappa$: neben dem gemeinsamen Mittelwert μ beeinflussen zufällige, $\mathbb{R}^{c_\kappa}$ -wertige Effekte über bekannte $n\times c_\kappa$-Matrizen U_κ die Beobachtungen. Mit $c := \sum c_\kappa$ und der Blockschreibweise $U := [U_1:\ldots:U_k] \in \mathbb{R}^{n\times c}$ und $\xi \coloneqq [\xi_1':\ldots:\xi_k']'$ kann gleichwertig notiert werden: $Y = 1_n\mu + U\xi$. Als weitere Voraussetzung habe ξ Mittelwert 0 und eine unbekannte Streuungsmatrix T. In unserer Schreibweise ergibt sich demnach $Y \sim (1_n\mu;\ UTU')$. Um zu zeigen, daß ein lineares Modell in unserem Sinn vorliegt, schreiben wir $U = [u_1:\ldots:u_c]$ als Blockmatrix seiner Spalten u_γ; dann gilt

$$UTU' = \sum T_{\gamma\gamma}u_\gamma u_\gamma' + \sum_{\beta<\gamma} T_{\beta\gamma}(u_\beta u_\gamma'+u_\gamma u_\beta') \quad .$$

In diesem Fall gibt es also neben dem unbekannten Mittelwert μ noch unbekannte ***Varianzkomponenten*** $T_{\gamma\gamma}$ und ***Kovarianzkomponenten*** $T_{\beta\gamma}$, $\beta \neq \gamma$.

8 BEISPIEL. ***Modelle mit gemischten Effekten.*** Bei einem Modell mit festen Effekten wird eine beliebige lineare Zerlegung des Mittelwerts und eine einfache Gestalt des Störterms vorausgesetzt, bei Modellen mit zufälligen Effekten ist es umgekehrt. Trifft beides zusammen, spricht man von einem Modell mit gemischten Effekten: $Y = Xb+\sum U_\kappa\xi_\kappa$, das heißt $Y \sim (Xb;\ UTU')$.

Die Begriffe "lineares Modell" und "Modell mit gemischten Effekten" sind gleichwertig: jedes lineare Modell (im Sinne unserer Definition) kann durch geeignete Wahl von U und T zu einem Modell mit gemischten Effekten gemacht werden. Umgekehrt haben wir vorher gezeigt, daß solche Modelle lineare Modelle sind.

9 BEISPIEL. *Varianzkomponenten-Modelle.* Eine wichtige Teilklasse bilden diejenigen Modelle mit gemischten Effekten, deren k Effekte ξ_κ paarweise unkorreliert sind und Streuungsmatrizen $\mathcal{D}\xi_\kappa = \sigma_\kappa^2 I_{c_\kappa}$ haben. In solchen Modellen $Y \sim (Xb; \sum \sigma_\kappa^2 U_\kappa U_\kappa')$ stellen die zerlegenden Matrizen $U_\kappa U_\kappa'$ selbst wieder Streuungsmatrizen dar, und die k Komponenten σ_κ^2 des Streuungsparameters sind Varianzen und als solche nichtnegativ. Abweichend von Scheffé (1959 p.221) und C.R. Rao (1973 pp.258,302) haben wir den Namen "Varianzkomponenten-Modelle" auf diese Fälle beschränkt.

Kapitel 2

DIE STREUUNGS-MITTELWERT-KORRESPONDENZ

In linearen Modellen sind quadratische Streuungsschätzung und lineare Mittelwertschätzung zwei Ansichten desselben Problems, das ist das Hauptergebnis dieses Kapitels. Der Weg dahin beginnt mit einigen Aussagen der multilinearen Algebra und führt zu Schätzern, die invariant sind gegenüber Verschiebungen des Mittelwerts.

1 Quadratische Funktionen $Q(y)$ entstehen per Definition aus bilinearen Funktionen $B(x,y)$, indem beide Argumente gleichgesetzt werden: $Q(y) = B(y,y)$. Die Zurückführung quadratischer Schätzfunktionen auf lineare stellt daher dasselbe Problem wie die Reduktion bilinearer Abbildungen auf lineare: dies leistet der Tensorkalkül der multilinearen Algebra.

Für den algebraischen Teil dieses Kapitels verweisen wir pauschal auf Kapitel 1 in Greub (1967), für die Streuungs-Mittelwert-Korrespondenz auf meine Diplomarbeit (1974) und die Folgearbeiten (Pukelsheim 1976, 1977).

2 Ein ***Tensorprodukt*** für zwei Vektorräume V und W ist eine bilineare Abbildung φ in einen Vektorraum T, so daß für jeden Vektorraum U und jede bilineare Abbildung $B: V\times W\to U$ genau eine lineare Abbildung $L: T\to U$ existiert mit $B = L\circ\varphi$. Für T wird auch $V\otimes W$ geschrieben.

Im Fall $V = \mathbb{R}^n$ und $W = \mathbb{R}^p$ gibt es zwei kanonische Tensorprodukte: das ***Kroneckerprodukt*** $x\otimes y : \mathbb{R}^n\times\mathbb{R}^p \to \mathbb{R}^{np}$, und das ***äußere Produkt*** $xy' : \mathbb{R}^n\times\mathbb{R}^p \to \mathbb{R}^{n\times p}$. Der dann existierende tensorprodukttreue Vektorraumisomorphismus definiert für eine Matrix A den Vektor vec A durch lexikographische Anordnung der Einträge:

$$\operatorname{vec} A := (A_{11},A_{12},\dots,A_{1p},\dots,A_{np})' : \mathbb{R}^n\times\mathbb{R}^p \to \mathbb{R}^{np} ;$$

$$\operatorname{vec} xy' = x \otimes y .$$

Die konsistente Erweiterung des euklidischen inneren Produkts zweier Vektoren x und y des $\mathbb{R}^n$

$$<x,y> := \sum x_\nu y_\nu = x'y$$

ist das euklidische Produkt zweier Matrizen A und B des $\mathbb{R}^{n\times p}$

$$<A,B:> := \sum A_{\nu\pi}B_{\nu\pi} = \text{trace } AB' \ .$$

Die Funktion vec erhält diese inneren Produkte:

$$(\text{vec } A)'(\text{vec } B) = <\text{vec } A, \text{ vec } B> = <A,B> = \text{trace } AB' \ .$$

Zur Klammerersparnis werden wir häufig vec'A statt (vec A)' schreiben.

3 Das Kroneckerprodukt zweier Vektoren induziert auf natürliche Weise das Kroneckerprodukt $A \otimes B = [[A_{\nu\pi}B]]$ zweier Matrizen (Ben-Israel & Greville 1974 pp.41f,119). Wir werden von den folgenden Rechenregeln Gebrauch machen:

$$A \otimes B \cdot C \otimes D = AC \otimes BD \ ; \quad (A \otimes B)' = A' \otimes B' \ ;$$

$$\text{vec } ABC = A \otimes C' \cdot \text{vec } B \ ; \quad (A \otimes B)^+ = A^+ \otimes B^+ \ ;$$

$$\text{rank } A \otimes B = \text{rank } A \cdot \text{rank } B \ ; \quad \mathcal{R}(A \otimes B) = \mathcal{R}A \otimes \mathcal{R}B \ .$$

4 Mit diesen Hilfen führen wir nun einige weitere Bezeichnungen für lineare Modelle ein. Für ein lineares Modell $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ bilden wir analog zur Blockmatr1x $X = [x_1:\ldots:x_p]$ die Blockmatrix

$$D := [\text{vec } V_1:\ldots:\text{vec } V_k] \in \mathbb{R}^{n^2\times k}$$

Nach §2 kann jede quadratische Schätzfunktion Q(Y,Y) für den Streuungsparameter t faktorisiert werden mittels einer linearen Funktion und des Kroneckerprodukts $Y \otimes Y$. Für dessen Erwartungswert gilt: $\mathcal{E}\, Y \otimes Y = \text{vec } \mathcal{E}\, YY' = \text{vec}(\mathcal{D}\, Y + \mathcal{E}\, Y(\mathcal{E}\, Y)')$, und weiter

$$\mathcal{E}\, Y \otimes Y = Dt + X \otimes X \cdot b \otimes b \ .$$

Demnach hängt der Erwartungswert einer quadratischen Schätzfunktion für den Streuungsparameter t im allgemeinen auch vom Mittelwertparameter b ab, was aus mehreren Gründen nicht befriedigen kann.

Denn teilt man den Stichprobenraum $\mathbb{R}^n$ auf in den Teilraum $\mathcal{R}X$ und dessen orthogonales Komplement (Scheffé 1959 p.23), dann kann offensichtlich derjenige Anteil der Beobachtung Y, der nach $\mathcal{R}X$ fällt, als Mittelwert gedeutet werden; mithin sollte sich eine Schätzung des Streuungsparameters nur auf den dazu komplementären Anteil in $\mathcal{R}^{\perp}X$ stützen.

Des weiteren hat man häufig ein zweistufiges Vorgehen im Sinn: Zunächst finde man eine Schätzung $\hat{t}$ für den Streuungsparameter t, die den unbekannten Mittelwertparameter nicht benutzt, und danach definiere man mit $V = \sum \hat{t}_\kappa V_\kappa$ ein neues Modell $Y \sim (Xb; V)$, um b zu schätzen.

5 Schließlich beschreibe die Teilmenge $\overline{G}$ des $\mathbb{R}^k$ diejenigen Werte $t \in \mathbb{R}^k$, für die $\sum \hat{t}_\kappa V_\kappa$ NND ist. Betrachtet man die Mittelwert-Translationen $y \to y+Xb$, $b \in \mathcal{R}^p$, dann ist sowohl der Parameter

$$\mathcal{D}Y = \mathcal{D}(Y+Xb) = \sum t_\kappa V_\kappa$$

als auch – bei Normalverteilungsannahme – die zugrundeliegende Verteilungsklasse

$$\{ N_n(\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa) \mid b \in \mathbb{R}^p,\ t \in \overline{G} \subset \mathbb{R}^k \}$$

invariant gegenüber solchen Transformationen.

6 Die Forderung nach "Unabhängigkeit von b" wird folglich formalisiert, indem man nur Schätzfunktionen zuläßt, die invariant sind gegenüber der Gruppe der Mittelwert-Translationen $y \to y+Xb$, $b \in \mathbb{R}^p$. Eine maximalinvariante Statistik bezüglich dieser Gruppe ist MY, wobei

$$M = I_n - XX^+ = \mathrm{Proj}\ \mathcal{R}^{\perp}X \in \mathrm{Proj}(n)$$

orthogonal auf das orthogonale Komplement von $\mathcal{R}^{\perp}X$ projiziert (Seely 1971 p.718; Kleffe 1976 p.707).

Da jede invariante Statistik notwendig über die Maximalinvariante MY faktorisiert, führt ***Reduktion durch Invarianz*** auf das M-***reduzierte Modell*** $MY \sim (0; \sum t_\kappa MV_\kappa M)$. Es gilt

$$\mathcal{E}\ MY \otimes MY = M \otimes M \cdot Dt = D_M t \text{ , mit}$$

$$D_M := M \otimes M \cdot D = [\mathrm{vec}\ MV_1M{:}\ldots{:}\mathrm{vec}\ MV_kM] \in \mathbb{R}^{n^2 \times k} \ .$$

Also wird die Klasse der ***invarianten quadratischen Schätzfunktionen*** für den Streuungsparameter t gegeben durch

$$\{ L \cdot MY \otimes MY \mid L \in \mathbb{R}^{k \times n^2} \} \ ,$$

das sind die linearen Schätzfunktionen zum Zufallsvektor $MY \otimes MY$. Sei $F := \mathcal{D}(Y-\mathcal{E}Y)\otimes(Y-\mathcal{E}Y)$ die $n^2{\times}n^2$-Matrix aller zentrierten vierten Momente von Y. Setzt man $F_M := M{\otimes}M \cdot F \cdot M{\otimes}M$, dann ist F_M die Streuungsmatrix von $MY{\otimes}MY$.

7 DEFINITION. Das zum Zufallsvektor $MY{\otimes}MY$ gehörende lineare Modell $MY{\otimes}MY \sim (D_M t;\ F_M)$ heißt das ***abgeleitete Modell*** des ursprünglichen Modells $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$.

8 Schätzung des Streuungsparameters ist also eine Frage des Standpunktes: entweder man faßt t als Streuungsparameter im ursprünglichen Modell auf und sucht hier nach invarianten quadratischen Schätzfunktionen, oder man interpretiert t als Mittelwertparameter im abgeleiteten Modell und bestimmt dort lineare Schätzfunktionen. Dies nennen wir die ***Streuungs-Mittelwert-Korrespondenz***.

Die wichtigste Auswirkung der Streuungs-Mittelwert-Korrespondenz ist die, daß für lineare Modelle eine Schätztheorie nur für den Mittelwert hergeleitet zu werden braucht. Über das abgeleitete Modell hat man sogleich Terminologie und Methoden zur Hand, um auch den Streuungsparameter t zu schätzen.

Andererseits kann jedes Schätzverfahren für t über das abgeleitete Modell auf die Schätzung eines Mittelwertparameters zurückgespielt werden; wir werden das etwa bei den Henderson-Schätzern (§II.2.7) und bei der nichtnegativen Schätzung von Varianzkomponenten (§IV.5.2) vormachen und gerade in diesen Fällen zu nützlicher Einsicht gelangen.

Das M-reduzierte Modell oder das abgeleitete Modell sind Beispiele für ***transformierte Modelle***; in solchen Modellen zu denken, wird weiterhin eine bedeutende Rolle spielen: Das von $V^{\frac{1}{2}+}Y$ erzeugte Modell bestimmt den Aitken-Schätzer; die von den Kroneckerpotenzen $MY^{\otimes 3}$ und $MY^{\otimes 4}$ und den Hadamardpotenzen MY^{*3} und MY^{*4} erzeugten Modelle dienen zur Schätzung von Schiefe und Exzeß; und Q-reduzierte Modelle führen zu nichtnegativen Schätzern von Varianzkomponenten.

Kapitel 3

GEMEINSAME PROBLEMSTELLUNG LINEARER UND MULTILINEARER SCHÄTZVERFAHREN

In diesem Kapitel beschreiben wir die verschiedenen Schätzverfahren für den Mittelwertparameter b und den Streuungsparameter t: Ausgehend von der Gauss'schen Risikofunktion der mittleren quadratischen Abweichung erörtern wir die Reduktionen durch Linearität, Erwartungstreue und Invarianz und die Minimierung von Norm und Varianz.

1 Als Beispiel schätzen wir eine Linearform c'b des Mittelwertparameters b; analoges gilt dann über das abgeleitete Modell für die Schätzung des Streuungsparameters t, wobei noch die in §2.6 beschriebene Reduktion durch Invarianz hinzukommt. Unsere Darstellung folgt der in Witting (1966 pp.25f,36f,44).

Gegeben sei ein lineares Modell $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ und ein Vektor c des $\mathbb{R}^p$, der eine zu schätzende Linearform c'b des Mittelwertparameters b bestimmt. Eine Schätzfunktion $\hat{c}(Y)$ ist per Definition eine meßbare Abbildung $\hat{c}$ des Borelschen Stichprobenraumes $(\mathbb{R}^n, \mathcal{B}^n)$ in die Zahlengerade $(\mathbb{R},\mathcal{B})$. Die Güte einer Schätzfunktion $\hat{c}(Y)$ wird durch die ***Gauss'sche Risikofunktion (mean square error)*** gemessen, die die mittlere quadratische Abweichung des geschätzten vom zu schätzenden Wert widergibt, wenn $P \in \mathcal{P}$ zugrundeliegt:

$$R(P;\hat{c}) := \mathcal{E}_P\, |\hat{c}(Y)-c'b|^2 = $$
$$= \mathrm{Var}_P\, \hat{c}(Y) + |\mathcal{E}_P\hat{c}(Y)-c'b|^2 .$$

Ein lineares Modell arbeitet statt mit den einzelnen Verteilungen $P\in\mathcal{P}$ mit deren Parametern b und t; die Existenz einer solchen Parametrisierung reicht im allgemeinen nicht, um für einen Schätzer $\hat{c}(Y)$ die Varianz $\mathrm{Var}_P\, \hat{c}(Y)$ und den Mittelwert $\mathcal{E}_P\, \hat{c}(Y)$ zu berechnen.

Lineare Modelle werden charakterisiert durch lineare Zerlegungen von Mittelwertvektor und Streuungsmatrix. Dazu tritt an dieser Stelle die ebenso kennzeichnende ***Reduktion durch Linearität***: In der Klasse aller Schätzfunktionen $\hat{c}(Y)$ wird nur die Teilklasse aller linearen Schätzfunktionen (LE ≡ linear estimator) $\hat{c}'Y$ ($\hat{c} \in \mathbb{R}^n$) betrachtet. Mit dieser brutalen, aber wirksamen Einschränkung gilt für die Risikofunktion:

$$R(b,t;\hat{c}) = \hat{c}'(\sum t_\kappa V_\kappa)\hat{c} + [(X'\hat{c}-c)'b]^2 \ .$$

2 Eine häufig gebilligte zweite Einschränkung bietet die ***Reduktion durch Erwartungstreue***: untersucht werden lediglich erwartungstreue lineare Schätzfunktionen (UB-LE ≡ unbiased linear estimator). Für diese gilt per Definition $\mathcal{E}Y = Xb \Rightarrow \mathcal{E}\hat{c}'Y := c'b$, das heißt $X'\hat{c} = c$. Damit verschwindet in der Risikofunktion mit dem zweiten Summanden auch die Abhängigkeit von b:

$$R(t;\hat{c}) := \mathrm{Var}_t\ \hat{c}'Y = \hat{c}'(\sum t_\kappa V_\kappa)\hat{c} \ .$$

3 Dieses Risiko läßt sich nach oben abschätzen, indem man die Matrixnormungleichung $\|AB\| \leq \|A\| \cdot \|B\|$ verwendet:

$$R(t;\hat{c}) = \|(\sum t_\kappa V_\kappa)^{½}\hat{c}\|^2 \leq \|(\sum t_\kappa V_\kappa)^{½}\|^2 \cdot \|\hat{c}\|^2 \ .$$

Der Streuungsparameter t ist nicht bekannt, und es liegt nahe, den von t unabhängigen Ausdruck $\|\hat{c}\|^2$ zu minimieren. Diese Lösungen nennen wir ***erwartungstreue lineare Schätzfunktionen von kleinster Norm*** (MN-UB-LE ≡ minimum norm unbiased linear estimator). Mindestens für $\sum t_\kappa V_\kappa = I_n$ sind solche Schätzer auch von kleinster Varianz.

4 Allgemein heißen erwartungstreue lineare Schätzfunktionen, die in ihrer Klasse das Risiko $R(t;\cdot)$ minimieren, ***erwartungstreue lineare Schätzfunktionen von kleinster Varianz unter t*** (MV-UB-LE ≡ minimum variance unbiased linear estimator). Wenn sinnvoll, wird der Zusatz "unter t" weggelassen. Wird die Varianz für alle t gleichzeitig minimiert, heißt der Schätzer ***von gleichmäßig kleinster Varianz*** (UMV ≡ uniform minimum variance).

5 Statt wie in §2 die Klasse aller linearen Schätzfunktionen auf die aller erwartungstreuen einzuschränken, kann man direkt das Risiko $R(b,t;\hat{c})$ minimieren. Um dabei einer Abhängigkeit von dem zu schätzenden Parameter b zumindest halbwegs zu entgehen, wird ein Minimaxverfahren gewählt, das das maximale Risiko auf der abgeschlossenen Kugel $\|b\| \leq \beta$ $(\beta > 0)$

$$\max \{ R(b,t;\hat{c}) \mid \|b\| \leq \beta \} = \hat{c}'(\sum t_\kappa V_\kappa)\hat{c} + \beta^2 \|X'\hat{c}-c\|^2$$

minimiert. Solche Schätzer nennen wir ***Minimax Mean Square Error lineare Schätzfunktionen*** (MMSE-LE).

6 Zur Übersicht stellen wir die beschriebenen Schätzverfahren für eine Linearform c'b, $c \in \mathbb{R}^p$, des Mittelwertparameters b noch einmal zusammen. Ein Bindestrich deutet an, daß die Klasse aller Schätzfunktionen, das heißt aller meßbaren Funktionen, auf eine Teilklasse eingeschränkt wird.

MN-UB-LE	Min. Norm erw.-treue lin. Schätzfkt.
MV-UB-LE	lok. Min. Var. erw.-treue lin. Schätzfkt.
UMV-UB-LE	gleichm. Min. Var. erw.-treue lin. Schätzfkt.
MMSE-LE	Minimax Mean Square Error lin. Schätzfkt.

Lineare Schätzfunktionen im abgeleiteten Modell sind invariante quadratische Schätzfunktionen (IQE) im ursprünglichen Modell (§§2.6f). Zur Bestimmung der Varianz solcher Schätzer müssen die vierten Momente des ursprünglichen Modells gegeben sein. Durch einfache Übertragung erhalten wir dann zur Schätzung einer Linearform q't, $q \in \mathbb{R}^k$, des Streuungsparameters t die folgenden Verfahren:

MN-UB-IQE	Min. Norm erw.-treue invar. quadr. Schätzfkt.
MV-UB-IQE	lok.Min.Var.erw.-treue invar. quadr. Schätzfkt.
UMV-UB-IQE	glm.Min.Var.erw.-treue invar. quadr. Schätzfkt.
MMSE-IQE	Minimax Mean Square Error inv. quadr. Schätzfkt.

7 In den hier gewählten Namen spiegeln sich sowohl die einzelnen Reduktionen wie auch die genauen Eigenschaften der Schätzverfahren wider. In der Literatur werden häufig andere Bezeichnungen benutzt; der Zusammenhang ist der folgende.

MN-UB-LE	Kleinste-Quadrate-Schätzer, ordinary (simple) least squares estimator, OLSE, SLSE
MV-UB-LE	best linear unbiased estimator, BLUE
UMV-UB-LE	Gauss-Markoff-Schätzer, GME
MMSE-LE	best linear estimator, BLE, Ridge estimator, minimax estimator
MN-UB-IQE	min.norm quadr.unbiased estimator, MINQUE
MV-UB-IQE	best quadratic unbiased estimator, BQUE

8 In Teil II werden die besten Schätzer für die genannten Verfahren angegeben. Es wird sich herausstellen, daß diese Schätzer formal auch dann existieren, wenn keine Erwartungstreue möglich ist; man kann zeigen, daß sie allgemein die Verzerrung (Bias) $\|X'\hat{c}-c\|^2$ minimieren und somit die Eigenschaft der Erwartungstreue ersetzt wird durch die ***Minimierung des Bias*** (MB). Die oben angegebenen Schätzer werden dadurch zu MN-MB-LE etc.; dabei entspricht der MV-MB-LE dem best linear minimum bias estimator (BLIMBE) von C.R. Rao (1973 pp.306f).

Für Bayes-Verfahren wird auf die Arbeit von Kleffe & Pincus (1974) verwiesen. Gnot, Klonecki & Zmyślony(1976) untersuchen erwartungstreue Streuungsschätzer mit gleichmäßig kleinster Varianz, ohne vorher durch Invarianz zu reduzieren. C.R. Rao (1976) gibt eine Übersicht über Minimax-Verfahren in linearen Modellen.

Plackett (1972) dokumentiert mit Briefen von Gauss und Legendre, was aus heutiger Sicht über die Entstehung der Methode der kleinsten Quadrate rekonstruiert werden kann.

9 Eine quadratische Schätzfunktion $\hat{q}' \cdot MY \otimes MY$ für eine Linearform $q't$ des Streuungsparameters wird üblicherweise als quadratische Form $(MY)'A(MY)$ geschrieben. Aus den Rechenregeln für das Kroneckerprodukt (§§2.2f) folgt $\hat{q}' \cdot MY \otimes MY = \langle \hat{q}, MY \otimes MY \rangle$ und $Y'MAMY = \operatorname{trace} AMY(MY)' = \langle A, MY(MY)' \rangle$. Wegen $\hat{q}' = \operatorname{vec} A$ wird die Übertragung von Eigenschaften zwischen $\hat{q}$ und A dann besonders einfach, wenn diese Eigenschaften durch innere Produkte beschrieben werden; als Beispiel führen wir ohne Beweis den folgenden Satz über Erwartungstreue an.

10 SATZ. *(Erwartungstreue)*

Gegeben sei ein lineares Modell $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ *mit den Bezeichnungen von* §2.6.

(a) Sei $c \in \mathbb{R}^p$. *Es ist* $\hat{c}'Y$ *eine erwartungstreue lineare Schätzfunktion für* $c'b$ *genau dann, wenn* $X'\hat{c} = c$, *oder äquivalent: für alle* $\pi = 1,\dots,p$: $\langle \hat{c}, x_\pi \rangle = c_\pi$.

(b) Sei $c \in \mathbb{R}^p$. *Es ist* $c'b$ *erwartungstreu linear schätzbar genau dann, wenn* $c \in \mathcal{R}\, X'$, *oder äquivalent:* c *liegt im Bildraum der zu den* p *zerlegenden Vektoren* x_π *gehörenden Gramschen Matrix* $X'X = ((\langle x_\pi, x_\rho \rangle))$.

(c) Alle Linearformen von b *sind erwartungstreu linear schätzbar genau dann, wenn die Gramsche Matrix* $X'X$ *nicht singulär ist, oder äquivalent: die zerlegenden Vektoren* $x_1,\dots,x_p$ *sind linear unabhängig.*

(d) Sei $q \in \mathbb{R}^k$. *Es ist* $(MY)'A(MY)$ *eine erwartungstreue invariante quadratische Schätzfunktion für* $q't$ *genau dann, wenn* $D_M'\cdot \mathrm{vec}\, A = q$, *oder äquivalent: für alle* $\kappa=1,\dots,k$: $\langle A, MV_\kappa M\rangle = q_\kappa$.

(e) Sei $q \in \mathbb{R}^k$. *Es ist* $q't$ *erwartungstreu invariant quadratisch schätzbar genau dann, wenn* $q \in \mathcal{R}\, D_M'$, *oder äquivalent:* q *liegt im Bildraum der zu den* k *zerlegenden Matrizen* $MV_\kappa M$ *gehörenden Gramschen Matrix* $D_M'D_M = ((\ \langle MV_\kappa M, MV_\lambda M\rangle\ ))$.

(f) Alle Linearformen von t *sind erwartungstreu invariant quadratisch schätzbar genau dann, wenn die Gramsche Matrix* $D_M'D_M$ *nicht singulär ist, oder äquivalent: die zerlegenden Matrizen* $MV_1M,\dots, MV_kM$ *sind linear unabhängig.*

11 Neben den Invarianzüberlegungen, die in §2.6 zur Definition des Projektors M führten, bestimmt M auch die Menge aller erwartungstreuen linearen Nullschätzer. Denn nach Satz 10(a) ist $\hat{c}'Y$ genau dann erwartungstreu für Null, wenn $X'\hat{c} = 0$, das heißt $\hat{c} \in \mathcal{R}X' = \mathcal{N}^\perp X = \mathcal{R}M$ (Ben-Israel & Greville 1974 p.64). Der lineare Raum aller erwartungstreuen linearen Nullschätzer ist also gleich $\{\ \hat{c}'Y \mid \hat{c} \in \mathcal{R}M\ \}$. Definieren wir analog zu M die Projektoren

$$N := I_{n^2} - D_M D_M^+ = \mathrm{Proj}\ \mathcal{R}^\perp D_M\ ; \quad N_M := M \otimes M \cdot N\ ;$$

dann ist der lineare Raum aller erwartungstreuen invarianten quadratischen Nullschätzer gleich $\{\ Y'AY \mid \mathrm{vec}\, A \in \mathcal{R}N_M\ \}$.

12 BEISPIEL. ***Heteroskastische Varianzen.*** In diesem Modell (§1.1) wie in solchen mit zufälligen Effekten (§1.7) tritt nur die einfache Mittelwertzerlegung $1_n\mu$ auf. Es lohnt sich, die zugehörigen Größen besonders zu kennzeichnen, und wir tun dies mit dem Index n:

$$J_n := 1_n 1_n'\ ; \quad \overline{J_n} := n^{-1} J_n\ ; \quad M_n := I_n - \overline{J_n}\ .$$

$\overline{J_n}$ und M_n sind Projektoren, die einen Vektor y "mitteln": $\overline{J_n}y = \overline{y.}1_n$, beziehungsweise "zentrieren": $M_n y = y - \overline{y.}1_n$.

Für das Modell mit heteroskedastischen Varianzen sei der Vektor der relativen Beobachtungshäufigkeiten: $a_\kappa = n_\kappa/n$. Dann gilt für die Gramsche Matrix der k zerlegenden Matrizen $M_n \cdot \mathrm{Diag}\,[0:I_{n_\kappa}:0] \cdot M_n$:

$$D_M'D_M = (n-2)\,\mathrm{Diag}\; a + aa' \;,$$

und diese Matrix ist genau dann nicht singulär, wenn $k = 1$ und $n > 1$, oder wenn $n > 2$ und alle $n_\kappa \geq 1$. Der für mehrere Varianzkomponenten ($k > 1$) zutreffende zweite Fall wird im folgenden zugrundegelegt, man erhält dafür

$$(D_M'D_M)^{-1} = \frac{1}{n-2}\,(\mathrm{Diag}\; a)^{-1} - \frac{1}{(n-1)(n-2)}\,J_k \;.$$

Wir setzen das Beispiel in §II.1.9 fort.

13 Insgesamt haben wir an ein lineares Modell $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ die folgenden gebundenen Variablen geknüpft:

$$\begin{array}{ll}
X : [x_1:\ldots:x_p] \;, & \\
D = [\mathrm{vec}\; V_1:\ldots:\mathrm{vec}\; V_k] \;; & D_M = M \otimes M \cdot D \;; \\
M = \mathrm{Proj}\; \mathscr{R}^\perp X = I_n - X(X'X)^- X' \;, & \\
N = \mathrm{Proj}\; \mathscr{R}^\perp D_M = I_{n^2} - D_M(D_M'D_M)^- D_M' ; & N_M = M \otimes M \cdot N \;; \\
X'X := ((\, x_\pi' x_\rho \,)) \;, & \\
D_M'D_M = ((\; \mathrm{trace}\; MV_\kappa MV_\lambda \;)) \;. &
\end{array}$$

Die paarige Anordnung läßt augenfällig zu Tage treten, daß der jeweils erste Begriff der Mittelwert- und der zweite der Streuungsschätzung zugeordnet ist. Diese Parallelität wird im zweiten Teil weiter ausgebaut.

Teil II

SPEZIELLE LINEARE UND MULTILINEARE SCHÄTZVERFAHREN

Kapitel 1

SCHÄTZUNG DES MITTELWERTS

In diesem Kapitel geben wir beste Schätzfunktionen für die in Kapitel I.3 begründeten Verfahren der erwartungstreuen linearen Schätzung von kleinster Norm oder Varianz und Minimax Mean Square Error linearen Schätzung an.

1 Vorausgesetzt sei ein lineares Modell $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ und eine erwartungstreu linear schätzbare Linearform $c'b$ des Mittelwertparameters b. Die Varianz wird minimiert für einen festen Wert $t \in \mathbb{R}^k$ des Streuungsparameters; setzen wir $V := \sum t_\kappa V_\kappa$, dann beschränken wir uns also auf das Modell $Y \sim (\sum b_\pi x_\pi;\ V)$, oder etwas allgemeiner $Y \sim (\sum b_\pi x_\pi;\ \sigma^2 V)$.

Eine zusammenfassende und vergleichende Diskussion zur Mittelwertschätzung führt Zyskind (1975 pp.653-661). Die folgenden beiden Sätze können bewiesen werden mit dem Projektionssatz in Drygas (1970 p.37), dem Satz von Lehmann-Scheffé (C.R. Rao 1973 p.317) oder unter Verwendung generalisierter Matrixinverser.

2 SATZ. (*Minimum Norm erwartungstreue Schätzer für* $c'b$)

Gegeben sei ein lineares Modell $Y \sim (\sum b_\pi x_\pi;\ \sigma^2 V)$ *und eine erwartungstreu linear schätzbare Linearform* $c'b$, $c \in \mathbb{R}^p$, *von* b.

Es gibt genau eine erwartungstreue lineare Schätzfunktion von kleinster Norm (MN-UB-LE) $\hat{c}'Y$ *für* $c'b$, *und dafür gilt:*

(a) $\hat{c}'Y = c'X^+Y$;

(b) $\hat{c}'Y$ $c'\hat{b}$ *mit beliebiger Lösung* $\hat{b}$ *der "Normalgleichung"*

$$X'X\hat{b} = X'Y \ ;$$

(c) $\hat{c} = \sum \lambda_\pi x_\pi$ *mit beliebiger Lösung* λ *der Gleichung* $X'X\lambda = c$.

3 SATZ. (*Minimum Varianz erwartungstreue Schätzer für* $c'b$)

Unter denselben Voraussetzungen wie im vorigen Satz setze

$B := X^+ - X^+V(MVM)^+$.

Für beliebige $z \in \mathbb{R}^n$ *ist* $c'BY + z'(M-MVM(MVM)^+)Y$ *eine erwartungstreue lineare Schätzfunktion von kleinster Varianz* (MV-UB-LE) *für* $c'b$. *Der Vielfachheit erzeugende Term* $M-MVM(MVM)^+$ *verschwindet genau dann, wenn* $y \in \mathcal{R}X+\mathcal{R}V$ *für alle* $y \in \mathbb{R}^n$*; in jedem Fall ist* $(M-MVM(MVM)^+)Y$ *fast sicher* 0.

4 Die ausschließliche Verwendung von Moore-Penrose Inversen wie in Albert (1973) erleichtert es, die Ergebnisse geometrisch zu interpretieren. Die Bedingung $y \in \mathcal{R}X+\mathcal{R}V$ besagt, daß jede mögliche Stichprobe $y \in \mathbb{R}^n$ durch die Annahme des Modells $Y \sim (\sum b_\pi x_\pi; V)$ erfaßt wird. - Die drei Summanden $c'X^+$, $c'X^+V(MVM)^+$, $z'(M - MVM(MVM)^+)$ sind orthogonal zueinander, so daß $c'BY$ unter allen MV-UB-LEn zusätzlich die Norm minimiert; im folgenden nennen wir $c'BY$ "die" erwartungstreue lineare Schätzfunktion von kleinster Varianz. - Die Zerlegung $c'X^+Y-c'X^+V(MVM)^+Y$ kann so gesehen werden, daß der erwartungstreue Schätzer von kleinster Norm $c'X^+Y$ um einen mit der Streuung V gewichteten Term korrigiert wird (Zyskind 1975 pp.660f).

5 Nach dem Satz von Lehmann-Scheffé (C.R. Rao 1973 p.517) ist eine erwartungstreue lineare Schätzfunktion $\hat{c}'Y$ genau dann von kleinster Varianz, wenn sie unkorreliert ist mit allen Nullschätzern, das heißt $MV\hat{c}=0$, oder äquivalent $V\hat{c} \in \mathcal{R}X$ (Zyskind 1975 p.653).

Im Spezialfall $V = I_n$ ist demnach die erwartungstreue lineare Schätzfunktion von kleinster Norm auch von kleinster Varianz. Allgemein ist das genau dann der Fall, wenn $\mathcal{R}VX \subset \mathcal{R}X$, oder äquivalent $\mathcal{R}VM \subset \mathcal{R}M$ dies läßt sich so interpretieren, daß die mit V gewichtete Korrektur auf $\mathcal{R}M$ nichts zur Mittelwertschätzung beiträgt: $V(\mathcal{R}M) \subset \mathcal{R}^\perp X$.

6 SATZ. (*Spezielle Darstellungen für* B)

Unter denselben Voraussetzungen wie in Satz 2 gilt für die Matrix B *aus Satz 3:*

(a) $B = X^+ \iff \mathcal{R}VX \subset \mathcal{R}X$.

(b) $B = (X'V^+X)^+X'V^+ \iff \mathcal{R}X \subset \mathcal{R}V$.

7 Zur Deutung der Aussage (b) betrachtet man das mit der

Inversen der Quadratwurzel (Ben-Israel & Greville 1974 p. 121) von V transformierte Modell $V^{\frac{1}{2}+}Y \sim (V^{\frac{1}{2}+}Xb;\sigma^2 VV^+)$ (C.R. Rao 1973 p. 221). Dieses Modell nennen wir ***Aitken-Modell***, da hier die erwartungstreue lineare Schätzfunktion von kleinster Norm dieselbe ist wie der ***Aitken-Schätzer***:

$$(V^{\frac{1}{2}+}X)^+ V^{\frac{1}{2}+}Y = (X'V^+X)^+X'V^+Y \ .$$

Den rechten Ausdruck erhält man dabei auch als Lösung der mit V ***gewichteten Normalgleichung:***

$$X'V^+X\hat{b} = X'V^+Y \ .$$

Minimax Mean Square Error lineare Schätzfunktionen werden in dieser Arbeit nicht weiter auftreten, doch seien sie der Vollständigkeit halber angegeben (vergleiche C.R. Rao 1976).

8 SATZ. (*Minimax Mean Square Error Schätzer für* c'b)

Unter denselben Voraussetzungen wie in Satz 2 und mit der Matrix B *aus Satz 3 gilt für alle* $\beta > 0$:

Eine lineare Schätzfunktion für c'b, *die unter* V *den maximalen Mean Square Error auf der Kugel* $\|b\| \leq \beta$ *minimiert* (MMSE-LE), *ist* $c'(I_p+\beta^{-2}BVB')^{-1}BY$; ***sie ist fast*** *sicher eindeutig.*

9 BEISPIEL. ***Heteroskedastische Varianzen***. Nach §I.3.12 ist μ erwartungstreu linear schätzbar. Die erwartungstreue lineare Schätzfunktion von kleinster Norm ist gerade das Stichprobenmittel $\overline{Y..}$, die von kleinster Varianz unter $\sigma_1^2,\dots,\sigma_k^2$ der Aitken-Schätzer $d^{-1}\sum Y_{\kappa.}/\sigma_\kappa^2$ mit dem Divisor $d := \sum n_\kappa/\sigma_\kappa^2$. Schrumpft man den Aitken-Schätzer um den Faktor $d/(\beta^{-2}+d) < 1$, erhält man den MMSE-LE $(\beta^{-2}+d)^{-1} \sum Y_\kappa /\sigma_\kappa^2$.

Nach Satz 6(a) ist das Stichprobenmittel genau dann von kleinster Varianz, wenn alle k Varianzkomponenten gleich sind:. $\sigma_1^2 = \cdots = \sigma_k^2$ Ist eine solche Annahme unberechtigt, dann liegt es nahe, die unbekannten Varianzkomponenten σ_k^2 durch Schätzungen zu ersetzen.

Das Beispiel wird in §IV.1.2 fortgesetzt.

Kapitel 2

SCHÄTZUNG DER VARIANZ-KOVARIANZ-KOMPONENTEN

In diesem Kapitel wenden wir die Streuungs-Mittelwert-Korrespondenz an, übersetzen damit das vorige Kapitel über Mittelwertschätzung und erhalten so auf einfache Weise eine abgerundete Theorie der Streuungsschätzung.

1 Vorausgesetzt sei ein lineares Modell $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ mit M-reduziertem Modell $MY \sim (0;\ \sum t_\kappa M V_\kappa M)$ und abgeleitetem Modell $MY \otimes MY \sim (D_M t;\ F_M)$, siehe §§I.2.6f. Ferner bestimme $q \in \mathbb{R}^k$ eine erwartungstreu invariant quadratisch schätzbare Linearform $q't$ des Streuungsparameters t. Wie schon für Satz I.3.10 (*Erwartungstreue*) identifizieren wir für eine quadratische Schätzfunktion $Y'AY$ den Vektor vec A mit $\hat{q}$; die Streuungs-Mittelwert-Korrespondenz I.2.8 überträgt dann die Ergebnisse der Mittelwertschätzung über das abgeleitete Modell in eine Theorie der Streuungsschätzung.

2 SATZ. (*Minimum Norm erwartungstreue Schätzer für* $q't$)

Gegeben sei ein lineares Modell $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ *und eine erwartungstreu invariant quadratisch schätzbare Linearform* $q't$, $q \in \mathbb{R}^k$, *von* t.

Es gibt genau eine erwartungstreue invariante quadratische Schätzfunktion von kleinster Norm (MN-UB-IQE) $Y'\hat{A}Y$ *für* $q't$, *und dafür gilt:*

(a) $Y'\hat{A}Y = q'D_M^{+} \cdot Y \otimes Y$;

(b) $Y'\hat{A}Y = q'\hat{t}$ *mit beliebiger Lösung* $\hat{t}$ *der "Normalgleichung"*

$$D_M'D_M\hat{t} = D_M' \cdot Y \otimes Y \ ;$$

(c) $\hat{A} = \sum t_\kappa M V_\kappa M$ *mit beliebiger Lösung* λ *der Gleichung*

$$D_M'D_M\lambda = q \ .$$

3 SATZ. (*Minimum Varianz erwartungstreue Schätzer für* q't)

Unter denselben Voraussetzungen wie im vorigen Satz setze $T := D_M^+ - D_M^+ F_M (NF_M N)^+$ *mit bekannten vierten Momenten* F_M *von* $MY \otimes MY$.

Für beliebiges $z \in \mathbb{R}^{n^2}$ *ist* $q'T \cdot MY \otimes MY + z'(N - NF_M N(NF_M N)^+) \cdot MY \otimes MY$ *eine erwartungstreue invariante quadratische Schätzfunktion von kleinster Varianz* (MV-UB-IQE) *unter* F_M *für* q't. *Der Vielfachheit erzeugende Term* $N - NF_M N(NF_M N)^+$ *verschwindet genau dann, wenn* $My \otimes My \in \mathcal{R}D_M + \mathcal{R}F_M$ *für alle* $y \in \mathbb{R}^n$; *in jedem Fall ist* $(N - NF_M N(NF_M N)^+) \cdot MY \otimes MY$ *fast sicher* 0.

4 Die Sätze 1.6 und 1.8 führen offensichtlich zu weiteren Analoga bei der Streuungsschätzung (Pukelsheim 1976). Als grundlegender Unterschied zur Mittelwertschätzung erscheint jedoch die Annahme funktionaler Unabhängigkeit von Mittelwertvektor und Streuungsmatrix hier nicht mehr gerechtfertigt. Unter Normalverteilungsannahme sind für eine invariante quadratische Schätzfunktion Y'AY Erwartungswert und Varianz über die Streuungsmatrix $\mathcal{D}Y$ gekoppelt:

$$\mathcal{E}\, Y'AY = \operatorname{trace} A(\mathcal{D}Y) \; ; \quad \operatorname{Var} Y'AY = 2 \operatorname{trace} A(\mathcal{D}Y)A(\mathcal{D}Y).$$

Für den folgenden Satz übernehmen wir die so berechneten vierten Momente

$$\mathcal{D}Y = V \quad \Rightarrow \quad F_M = 2\, MVM \otimes MVM,$$

ohne weiter von der Normalverteilungsannahme Gebrauch zu machen. Dies haben wir in §I.1.5 "Quasinormalität" genannt. Bei der Angabe von F_M wurde eine Vereinfachung vorweggenommen, die wir erst in §3.6 rechtfertigen werden.

Andererseits eröffnet die Koppelung von Mittelwert und Streuungsmatrix die Möglichkeit, die Existenz einer erwartungstreuen invarianten quadratischen Schätzfunktion mit gleichmäßig kleinster Varianz (UMV-UB-IQE) an Hand der k zerlegenden Matrizen $MV_\kappa M$ zu charakterisieren. Eine elegante Formulierung fand Seely (1971); ihm folgend heißt ein Teilraum $\mathcal{B}$ des linearen Raumes Sym(n) der symmetrischen Matrizen ***quadratisch***, wenn $\mathcal{B}$ mit einer Matrix V auch das Quadrat V^2 enthält. An einem erzeugenden System $\mathcal{B}_1$ von $\mathcal{B}$ kann man diese Eigenschaft leicht erkennen: $\mathcal{B}$ ist genau dann quadratisch, wenn $(V+W)^2 \in \mathcal{B}$ für alle $V,W \in \mathcal{B}_1$.

5 SATZ. (*Erw.-treue Schätzer von glm. kl. Var. für* t)
Gegeben bei ein quasinormales lineares Modell $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$. *Die* k *zerlegenden Matrizen* $MV_\kappa M$ *seien linear unabhängig, und der von ihnen aufgespannte Teilraum* $\mathcal{B}$ *enthalte* M.

(a) Wenn eine erwartungstreue invariante quadratische Schätzfunktion von gleichmäßig kleinster Varianz (UMV—UB—IQE) *für* t *existiert, dann ist sie eindeutig bestimmt und gleich der erwartungstreuen invarianten quadratischen Schätzfunktion von kleinster Norm* (MN—UB—IQE) *für* t.

(b) Sie existiert genau dann, wenn $\mathcal{B}$ *quadratisch ist.*

BEWEIS. (a) Da die $MV_\kappa M$ linear unabhängig sind, ist jede Linearform von t (I.3.10.f) und damit auch der Vektor t erwartungstreu invariant quadratisch schätzbar. $M\in\mathcal{B}$ ist eine zulässige Streuungsmatrix; unter ihr ist die MN—UB-IQE als einziger Schätzer von kleinster Varianz (Satz 3; Satz 1.6.a im abgeleiteten Modell) und folglich der einzige Kandidat für die UMV-UB-IQE.

(b) Der von den $MV_\kappa M$ aufgespannte Raum $\mathcal{B}$ ist im wesentlichen der Bildraum von D_M, genauer: $\mathcal{R}D_M = \text{vec}\ \mathcal{B}$. Als $\mathcal{B}$ aufspannende Teilmenge $\mathcal{B}_1$ wählen wir den konvexen Kegel aller NND Matrizen in $\mathcal{B}$. Die MN—UB-IQE ist per Definition von gleichmäßig kleinster Varianz, wenn sie unter jedem $V \in \mathcal{B}_1$ die Varianz lokal minimiert. Wenden wir Satz 1.6(a)auf das abgeleitete Modell an: $\mathcal{R}\ 2\ V\otimes V \cdot D_M \subset \mathcal{R}D_M$, dann erhalten wir als notwendige und hinreichende Bedingung:

$$(*) \qquad \forall V\in\mathcal{B}_1\ \forall W\in\mathcal{B}: \quad VWV \in \mathcal{B}\,.$$

Ist einerseits $\mathcal{B}$ quadratisch, so folgt (*). Gilt andererseits (*), dann liegt für jedes $V\in\mathcal{B}_1$ das Quadrat $V^2 = VMV$ in $\mathcal{B}$; als konvexer Kegel enthält $\mathcal{B}_1$ mit zwei Matrizen V und W auch deren Summe V+W: $\mathcal{B}$ ist quadratisch. ●

Mit der jetzt verfügbaren Theorie können gängige Schätzer auf einfache und schnelle Art gerechtfertigt werden.

6 BEISPIEL. *Varianzkomponenten-Modelle*. In den folgenden vier Modellen (Corbeil & Searle 1976b pp.782—785) werden die Effekte

pro Gruppe und Untergruppe etc. gleich häufig beobachtet; dies ermöglicht die erfolgreiche Verwendung des Kroneckerprodukts. Außerdem benutzen wir die Matrizen J_n, $\overline{J_n}$ und M_n aus §I.3.12.
Für die vier Modelle geben wir der Reihe nach an
- die Komponentengleichungen,
- die Vektorgleichung,
- die Darstellung als lineares Modell,
- den Projektor M,
- die k zerlegenden Matrizen $MV_\kappa M$ und
- den Raum $\mathcal{B}$, aufgespannt durch paarweise orthogonale Projektoren.

(a) *Einfache balancierte Zerlegung mit zufälligem Effekt.*

$Y_{ij} = \mu + \alpha_i + e_{ij}$; $i=1,\dots,a$; $j=1,\dots,n$.

$Y = 1_{an}\mu + I_a \otimes 1_n \cdot \alpha + e$.

$Y \sim (1_{an}\mu;\ \sigma_\alpha^2\ I_a \otimes J_n + \sigma^2\ I_a \otimes I_n)$.

$M = M_{an} = I_a \otimes M_n + M_a \otimes \overline{J_n}$.

$M_{an} \cdot I_a \otimes J_n \cdot M_{an} = M_a \otimes J_n$.

$\mathcal{B} = \text{span}\ \{ I_a \otimes M_n;\ M_a \otimes \overline{J_n} \}$.

(b) *Zweifache balancierte hierarchische Zerlegung mit zufälligem Effekt.*

$Y_{ij\ell} = \mu + \alpha_i + \beta_{ij} + e_{ij\ell}$; $i=1,\dots,a$; $j=1,\dots,b$; $\ell=1,\dots,n$.

$Y = 1_{abn}\mu + I_a \otimes 1_b \otimes 1_n \cdot \alpha + I_a \otimes I_b \otimes 1_n \cdot \beta + e$.

$Y \sim (1_{abn}\mu;\ \sigma_\alpha^2\ I_a \otimes J_b \otimes J_n + \sigma_\beta^2\ I_a \otimes I_b \otimes J_n + \sigma^2\ I_a \otimes I_b \otimes I_n)$.

$M = M_{abn} = I_a \otimes I_b \otimes M_n + I_a \otimes M_b \otimes \overline{J_n} + M_a \otimes \overline{J_b} \otimes \overline{J_n}$.

$M_{abn} \cdot I_a \otimes J_b \otimes J_n \cdot M_{abn} = M_a \otimes J_b \otimes J_n$.

$M_{abn} \cdot I_a \otimes I_b \otimes J_n \cdot M_{abn} = I_a \otimes M_b \otimes J_n + M_a \otimes \overline{J_b} \otimes J_n$.

$\mathcal{B} = \text{span}\ \{ M_a \otimes \overline{J_b} \otimes \overline{J_n};\ I_a \otimes M_b \otimes \overline{J_n};\ I_a \otimes I_b \otimes M_n \}$.

(c) *Zweifache balancierte gekreuzte Zerlegung mit gemsichten Effekten ohne Wechselwirkung.*

$Y_{ij\ell} = \mu + \alpha_i + \beta_j + e_{ij\ell}$; $i=1,\dots,a$; $j=1,\dots,b$; $\ell=1,\dots,n$.

$Y = 1_{abn}\mu + I_a \otimes 1_b \otimes 1_n \cdot \alpha + 1_a \otimes I_b \otimes 1_n \cdot \beta + e$.

$Y \sim ([1_{abn} : I_a \otimes 1_{bn}]\begin{bmatrix}\mu\\ \alpha\end{bmatrix};\ \sigma_\beta^2\ J_a \otimes I_b \otimes J_n + \sigma^2\ I_a \otimes I_b \otimes I_n)$.

$M = I_a \otimes M_{bn} = I_a \otimes I_b \otimes M_n. + I_a \otimes M_b \otimes \overline{J_n}$.

$M \cdot J_a \otimes I_b \otimes J_n \cdot M = J_a \otimes M_b \otimes J_n$.

$\mathcal{B} = \text{span}\ \{ \overline{J_a} \otimes M_b \otimes \overline{J_n};\ I_a \otimes I_b \otimes M_n + M_a \otimes M_b \otimes \overline{J_n} \}$.

(d) *Zweifache balancierte gekreuzte Zerlegung mit gemsichten Effekten mit Wechselwirkung.*

$Y_{ij\ell} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ij\ell};\quad i=1,\dots,a;\quad j=1,\dots,b;\quad \ell=1,\dots,n$.

$Y = 1_{abn}\mu + I_a \otimes 1_b \otimes 1_n \cdot \alpha + 1_a \otimes I_b \otimes 1_n \cdot \beta + I_a \otimes I_b \otimes 1_n \cdot (\alpha\beta) + e$.

$Y \sim ([1_{abn} : I_a \otimes 1_{bn}]\begin{bmatrix}\mu\\ \alpha\end{bmatrix};\quad \sigma_\beta^2 J_a \otimes I_b \otimes J_n + \sigma_{\alpha\beta}^2 I_a \otimes I_b \otimes J_n + \sigma^2 I_a \otimes I_b \otimes I_n)$.

$M = I_a \otimes I_b \otimes M_n + I_a \otimes M_b \otimes \overline{J_n}$.

$M \cdot J_a \otimes I_b \otimes J_n \cdot M = J_a \otimes M_b \otimes J_n$.

$M \cdot I_a \otimes I_b \otimes J_n \cdot M = I_a \otimes M_b \otimes J_n$.

$\mathcal{B} = \text{span}\ \{ \overline{J_a} \otimes M_b \otimes \overline{J_n};\ I_a \otimes I_b \otimes M_n;\ M_a \otimes M_b \otimes \overline{J_n} \}$.

In allen vier Modellen ist $\mathcal{B}$ ein quadratischer Teilraum symmetrischer Matrizen (Seely 1971 p.714). Nach Satz 5 ist die erwartungstreue invariante quadratische Schätzfunktion

$$\hat{t} = (D_M' D_M)^{-1} D_M' \cdot Y \otimes Y$$

von gleichmäßig kleinster Varianz. Bei zusätzlicher Normalverteilungsannahme ist die Statistik $D_M' \cdot Y \otimes Y$ vollständig und suffizient für t (Seely 1971 p.716) und $\hat{t}$ von gleichmäßig kleinster Varianz unter allen erwartungstreuen invarianten Schätzfunktionen, ohne Beschränkung auf die quadratischen. Zudem werden wir mit Satz III.2.4 beweisen, daß $\hat{t}$ gleich dem invarianten Maximum Likelihood Schätzer (REML) von t ist; Corbeil & Searle (1976b) zeigen, daß letzterer derselbe ist wie der klassische Varianzanalysen-Schätzer:

Ein quadratischer Teilraum $\mathcal{B} = \text{span}\ \{ MV_1M, \dots, MV_kM \}$ sichert paradiesische Verhältnisse zum Schätzen der Varianz-Kovarianz-Komponenten: Alle wichtigen Schätzverfahren fallen zusammen und stimmen mit dem einfachsten überein, bei dem im abgeleiteten Modell die Methode der kleinsten Quadrate angewendet wird.

Ähnlich wie hier werden wir in den beiden nächsten Kapiteln transformierte Modelle benutzen, um Schiefe und Exzeß zu schätzen.

7 EXKURS. ***Henderson-Schätzer***. Nach herkömmlicher Art werden die Varianz-Kovarianz-Komponenten mit einem der drei Henderson-Verfahren geschätzt, die sich bei balancierten Modellen zum gerade erwähnten Varianzanalysen-Schätzer vereinfachen. Die Henderson-Schätzer werden nach folgendem Rezept gefunden (Searle 1968 p.749): ***"All of Henderson's three methods involve (i) calculating mean squares of some kind, (ii) obtaining their expectations, and (iii) solving linear equations in the unknown variance components, derived from equating calculated mean squares to their expected values."***

Als eine grundlegende Konsequenz der Streuungs-Mittelwert-Korrespondenz besitzt jedes Verfahren zur Mittelwertschätzung ein Gegenstück bei der Streuungsschätzung und umgekehrt. Formuliert man das zitierte Rezept zum Gebrauch für die Mittelwertschätzung, so wird die willkürliche Vorgehensweise und mangelhafte theoretische Fundierung vollends deutlich: (i) Wähle einige Linearformen in Y. (Wir setzen sie zeilenweise zu einer Matrix L zusammen, $L \in \mathbb{R}^{\ell \times n}$.) (ii) Berechne den Mittelwert: $\mathcal{E}LY = LXb$. (iii) Streiche das $\mathcal{E}$ weg, setze einen Hut auf b und löse die Gleichung $LY = LX\hat{b}$ nach $\hat{b}$ auf!

Eine derart verkürzende Faustregel kann einer reifen Theorie wenig helfen: nichts sichert, daß die Gleichung $LY = LX\hat{b}$ – erstens - lösbar ist und - zweitens - erwartungstreue Schätzer $\hat{b}$ liefert. Zwar könnten diese kritischen Punkte genutzt werden, um die anfängliche Wahl der Linearformen zu motivieren, statt dessen werden sie zumeist gar nicht beachtet - wie Seely (1970 p.1744) kritisiert. Und wie die Diskussion zu Searle's Übersichtsartikel (1968 pp.779f) erkennen läßt, bleibt wenig mehr als Willkür, was die Henderson'schen Verfahren und die an ihrem Beginn stehenden quadratischen Formen bestimmt.

Jedoch sollte man nicht verkennen, daß heuristische Überlegungen und durch Erfahrung geschulte Intuition ***"for balanced data the 'obvious' quadratic forms to use"*** (Searle 1971 p.389) gefunden haben, und die Beispiele von §6 betonen, daß für den sich dann ergebenden Schätzer mehrere und wünschenswerte Optimalitätseigenschaften nachgewiesen werden können.

Kapitel 3

SCHÄTZUNG DER SCHIEFE

In diesem Kapitel zeigen wir, wie analog zur Verwendung des Kroneckerquadrats im abgeleiteten Modell der Übergang zu dritten Kroneckerpotenzen Schätzfunktionen für die Schiefe liefert; eine Teilklasse wird nicht mit dem Kronecker-, sondern mit dem Hadamardprodukt behandelt.

1 Anscombe (1961) untersucht für das klassische Normalmodell $Y \sim N_n(Xb;\ \sigma^2 I_n)$ die Frage, wie Abweichungen von der Normalverteilungsannahme entdeckt werden können. Er folgt dem Vorgehen, die Schiefe γ_1 und den Exzeß γ_2 zu schätzen und die erhaltenen Werte mit denen unter einer Normalverteilung, nämlich $\gamma_1 = \gamma_2 = 0$ zu vergleichen. Dabei sind Schiefe und Exzeß für eine reellwertige Zufallsgröße Y_1 so definiert:

$$\gamma_1 := \mathcal{E}\left[\frac{Y_1 - E\ Y_1}{\sqrt{\operatorname{Var} Y_1}}\right]^3 \ ; \qquad \gamma_2 := \mathcal{E}\left[\frac{Y_1 - \mathcal{E}\,Y_1}{\sqrt{\operatorname{Var} Y_1}}\right]^4 - 3\ .$$

2 Für dieses Kapitel und das nächste sei ein lineares Modell $Y \sim (Xb;\ \sigma^2 I_n)$ mit der zusätzlichen Annahme versehen, daß Y unabhängige Komponenten mit gleicher Schiefe γ_1 und gleichem Exzeß γ_2 habe; wir nennen dieses Modell kurz ein ***klassisches Modell.***

In diesem Kapitel beschäftigen wir uns mit der Schätzung der Schiefe. Sie soll mit einem homogenen Polynom dritten Grades und unabhängig vom Mittelwert geschätzt werden. Nach Reduktion durch Invarianz rechnen wir also im M-reduzierten Modell (§I.2.6) $Z \coloneqq (0;\ \sigma^2 M)$. Die Klasse aller homogenen Polynome dritten Grades in Z sei Γ_1, die Teilklasse, bei der Z nur in reinen Kuben $Z_1^3,\dots,Z_n^3$ vorkommt, sei $\mathcal{G}_1$; die Suche nach besten Schätzfunktionen wird auf die Klassen Γ_1 und $\mathcal{G}_1$ beschränkt.

Es wird sich herausstellen, daß die optimalen Schätzfunktionen noch von der unbekannten Varianz σ^2 abhängen. Um zu praktikablen Ergebnissen zu gelangen, bieten sich zwei Auswege:

Entweder man schätzt statt der Schiefe γ_1 das dritte Moment $\mu_3 := \sigma^3\gamma_1$ und kann die gefundenen Optimalitätseigenschaften dann ausnahmslos übernehmen. Oder man besteht auf der Schätzung der Schiefe γ_1 und ersetzt σ^2 durch eine Schätzung, zerstört dann aber die bewiesene Optimalität.

3 Die Komponenten der dritten Kroneckerpotenz $Z^{\otimes 3} := Z\otimes Z\otimes Z$ bestehen gerade aus den n^3 gemischten dritten Potenzen von $Z_1,\dots, Z_n$, so daß ein homogenes Polynom dritten Grades als Linearform $w'Z^{\otimes 3}$ geschrieben werden kann. Kommen nur reine Kuben $Z_1^3,\dots,Z_n^3$ vor, dann liegt eine Linearform der dritten Potenz des Hadamardprodukts (Styan 1973) $Z^{*3} := Z*Z*Z = (Z_1^3,\dots,Z_n^3)'$ vor:

$$\Gamma_1 = \{\ w'Z^{\otimes 3} \mid w\in\mathbb{R}^{n^3}\ \}\ ; \qquad \mathcal{G}_1 = \{\ w'Z^{*3} \mid w\in\mathbb{R}^n\ \}\ .$$

Bei den Erfahrungen mit der Streuungs—Mittelwert-Korrespondenz leuchtet ein, hier die von $Z^{\otimes 3}$ und Z^{*3} erzeugten Modelle zu untersuchen. Dieses Vorhaben wird dadurch erleichtert, daß die Hadamardpotenz mit einer linearen Transformation aus der Kroneckerpotenz hervorgeht; definieren wir nämlich

$$H_3 := \sum e_\nu(e_\nu'^{\otimes 3}) \in \mathbb{R}^{n\times n^3}\ , \qquad H_4 := \sum e_\nu(e_\nu'^{\otimes 4}) \in \mathbb{R}^{n\times n^4}\ ,$$

wobei e_ν der ν—te euklidische Basisvektor $(0,\dots,0,1,0,\dots,0)$ des $\mathbb{R}^n$ ist, dann gilt: $H_3Z^{\otimes 3} = Z^{*3}$, $H_4Z^{\otimes 4} = Z^{*4}$.

Zunächst berechnen wir die Erwartungswerte.

4 HILFSSATZ. *Für ein klassisches Modell gilt:*

(a) $\mathcal{E}Z^{\otimes 3} = M^{\otimes 3}(\sum e_\nu^{\otimes 3})\sigma^3\gamma_1\ ; \quad \mathcal{E}Z^{*3} = M^{*3}1_n\sigma^3\gamma_1\ .$

(b) In Γ_1 *oder* $\mathcal{G}_1$ *existiert eine erwartungstreue Schätzfunktion für* γ_1 *genau dann, wenn* $\sum M_{\nu\mu}^3 \neq 0$.

BEWEIS. (a) Setze $U := \sigma^{-1}(Y-\mathcal{E}Y)$. Dann haben die unabhängigen Komponenten von U Mittelwert 0, Varianz 1 und Schiefe γ_1. Also verschwindet der Erwartungswert für Potenzen der Art $U_1U_2U_3$ und $U_1^2U_2$; für U_1^3 ist er γ_1.

$$\mathcal{E}Z^{\otimes 3} = \mathcal{E}(MY)^{\otimes 3} = \sigma^3M^{\otimes 3}\mathcal{E}U^{\otimes 3} = \sigma^3M^{\otimes 3}(\sum e_\nu^{\otimes 3})\gamma_1\ ,$$

und Transformation mit H_3 liefert den Erwartungswert für Z^{*3}.

(b) M^{*3} ist NND (Styan 1973 p.221). Daher verschwindet $M^{*3}1_n$ genau mit $(M^{*3})^{1/2}1_n$ oder $1_n'M^{*3}1_n = \sum M_{\nu\mu}^3$. ●

5 Wir werden die Varianz lokal unter Normalverteilungsannahme minimieren. Um die dazu notwendigen sechsten und im nächsten Kapitel die achten Momente übersichtlich darzustellen, lohnt die Beobachtung, daß $z^{\otimes 3}$; nicht nur im $\mathbb{R}^{n^3}$, sondern schon im $\binom{n+3-1}{3}$-dimensionalen Teilraum aller symmetrischen dritten Kroneckerpotenzen liegt.

Genauer benutzen wir für p-te Kroneckerpotenzen $\otimes^p\mathbb{R}^n = \mathbb{R}^{n^p}$ die folgende Terminologie (Greub 1967 pp.90f). Der ***Symmetrisator*** π_s ist ein Endomorphismus des Raumes $\otimes^p\mathbb{R}^n$ und für das erzeugenden System der p-fachen Kroneckerprodukte $\otimes_{\pi=1}^p x_\pi$ $(x_\pi \in \mathbb{R}^n)$ definiert durch

$$\pi_s \cdot \otimes_{\pi=1}^p x_\pi := \frac{1}{p!} \sum_{\sigma\in\mathcal{S}_p} \otimes_{\pi=1}^p x_{\sigma(\pi)} \ ;$$

dabei ist $\mathcal{S}_p$ die symmetrische Gruppe (Permutationsgruppe) der Ordnung p. Die so definierte Abbildung π_s ist ein Projektor, sein Bildraum heißt ***Teilraum der symmetrischen p-ten Kroneckerpotenzen***, seine Bilder ***symmetrisch***. Eine lineare Abbildung $L \in \mathbb{R}^{\ell\times n^p}$ wird ***symmetrisch*** genannt, falls $L = L\cdot\pi_s$.

Beispielsweise ist $z^{\otimes 3} = \pi_s\cdot z^{\otimes 3}$ symmetrisch; bei einer Linearform $w'z^{\otimes 3}$ kann daher ohne Beschränkung der Allgemeinheit w als symmetrisch angenommen werden: $w = \pi_s w$. Ein Beispiel für eine symmetrische Abbildung ist die Transformation H auf Hadamardpotenzen.

6 Mit der tensorprodukttreuen Isometrie vec (§I.2.2) macht man sich leicht klar, daß der Symmetrisator π_s die natürliche Verallgemeinerung des Übergangs $A \to \frac{1}{2}(A+A')$ von quadratischen Matrizen A zu symmetrischen ist.

Bei der Darstellung der vierten Momente F_M (§2.4) haben wir den Symmetrisator π_s übergangen. Ausführlich lauten die vierten Momente nicht F_M, sondern $\pi_s F_M \pi_s$ (vergleiche Lemma 4.1 mit $\Delta_2 = 0$ in Pukelsheim 1977). Daß π_s ohne Schaden vernachlässigt werden kann, liegt an dreierlei: Erstens kommutieren π_s und F_M, zweitens sind die k zerlegenden Vektoren vec $MV_\kappa M$ des abgeleiteten Modells symmetrisch, und drittens sucht man ohne Beschränkung der Allgemeinheit nur unter Schätzern, die durch symmetrische $\hat{q}$ bestimmt werden:

$$\pi_s F_M \pi_s = \pi_s F_M = F_M \pi_s \ ; \quad \pi_s D_M = D_M; \quad \pi_s\hat{q} = \hat{q} \ .$$

Das Lehmann-Scheffé'sche Kriterium (§1.5) wie auch die Bedingungen aus Satz 1.6 bleiben damit dieselben, ob man nun den Symmetrisator π_s beachtet oder nicht:

$$N\pi_s F_M \pi_s \hat{q} = 0 \quad \Leftrightarrow \quad NF_M\hat{q} = 0 \;;$$

$$\mathcal{R}D_M \subset \mathcal{R}\pi_s F_M \pi_s \quad \Leftrightarrow \quad \mathcal{R}D_M \subset \mathcal{R}F_M;$$

$$\mathcal{R}\pi_s F_M \pi_s D_M \subset \mathcal{R}D_M \quad \Leftrightarrow \quad \mathcal{R}F_M D_M \subset \mathcal{R}D_M.$$

7 Wir sind nun in der Lage, die sechsten Momente von Z unter Normalverteilungsannahme anzugeben. Dazu definieren wir die (i,j)-ten euklidischen Basismatrizen $E_{ij} := e_i e_j'$, die auf Platz (i,j) eine Eins haben und sonst überall Nullen, und setzen

$$S_\Gamma := 6M^{\otimes 3} + 3M^{\otimes 3}\textstyle\sum_{i,j,k}(E_{ij}\otimes E_{ij}\otimes E_{kk}+E_{ij}\otimes E_{kk}\otimes E_{ij}+E_{kk}\otimes E_{ij}\otimes E_{ij})M^{\otimes 3} \;;$$

$$S_{\mathcal{G}} := 6M^{*3} + 9(\mathrm{Diag}\ M)M(\mathrm{diag}\ M) \;.$$

8 HILFSSATZ. *Für ein quasinormales klassisches Modell gilt:*

$$\mathcal{D}Z^{\otimes 3} = \sigma^6\ \pi_s S_\Gamma \pi_s \;; \quad \mathcal{D}Z^{*3} = \sigma^6\ S_{\mathcal{G}} \;.$$

BEWEIS. Der erste Teil der Behauptung ergibt sich mit den sechsten Momenten eines $N_n(0;\ I_n)$-verteilten Zufallsvektors U aus $\mathcal{D}Z^{\otimes 3} = \sigma^6 M^{\otimes 3}(\mathcal{D}U^{\otimes 3})M^{\otimes 3}$ und der Beobachtung, daß $M^{\otimes 3}$ und π_s kommutieren. Der zweite Teil folgt aus dem ersten mit der Transformation H_3. ●

Mit derselben Begründung wie in §6 wird im weiteren der Symmetrisator π_s weggelassen: S_Γ und π_s kommutieren, der zerlegende Vektor $M^{\otimes 3}(\sum e_\nu^{\otimes 3})$ des Mittelwerts von $Z^{\otimes 3}$ ist symmetrisch, und man kann sich bei den Linearformen $w'Z^{\otimes 3}$ in Γ_1 auf symmetrische w beschränken.

Alle Summanden von S_Γ sind NND, dies erkennt man etwa aus $E_{ij}\otimes E_{ij}\otimes E_{kk} = (e_i\otimes e_i)(e_j\otimes e_j)'\otimes e_k e_k'$. Die Bedingung $\mathcal{R}X \subset \mathcal{R}V$ aus Satz 1.6(b) ist damit verifiziert: $M^{\otimes 3}(\sum e_\nu^{\otimes 3}) \in \mathcal{R}M^{\otimes 3} \subset \mathcal{R}\ S_\Gamma$, denn für NND Matrizen addieren sich die Bildräume (Hilfssatz IV.1.4.c). Dies impliziert Teil (a) des folgenden Satzes, Teil (b) folgt analog; die Teile (e),(f),(g) können mit dem Lehmann-Scheffé-schen Satz (§1.5) hergeleitet werden. Wir verzichten darauf, die Schritte im einzelnen vorzurechnen.

9 SATZ. (*Schätzfunktionen für die Schiefe* γ_1)

Gegeben sei ein lineares Modell $Y \sim (Xb;\sigma^2 I_n)$, *in dem* Y *unabhängige Komponenten mit gleicher Schiefe* γ_1 *besitze. Setze* $m := (M_{11},\dots,M_{nn})'$. *Es sei* $1_n' M^{*3} 1_n \neq 0$. *Dann gilt für die §3 beschriebenen Klassen* Γ_1 *und* $\mathcal{G}_1$ *von Schätzfunktionen mit den in §7 definierten Matrizen* S_Γ *und* $S_{\mathcal{G}}$ *der sechsten Momente und* $Z := MY$:

(a) Der Aitken-Schätzer für γ_1 *im von* $Z^{\otimes 3}$ *erzeugten Modell* $Z^{\otimes 3} \sim (M^{\otimes 3}(\sum e_\nu^{\otimes 3})\sigma^3\gamma_1;\ \sigma^6 S_\Gamma)$ *ist*

$$\widehat{\gamma_1} := (\sigma^3 \sum e_\nu^{\otimes 3}{}' S_\Gamma^{+} e_\mu^{\otimes 3})^{-1} \sum e_\nu^{\otimes 3}{}' S_\Gamma^{+} Z^{\otimes 3} \in \Gamma_1 \ ;$$

$\widehat{\gamma_1}$ *ist ein homogenes Polynom dritten Grades in* $Z_1,\dots,Z_n$, *erwartungstreu für* γ_1 *und hat unter allen diesen Schätzfunktionen kleinste Varianz bei Normalverteilungsannahme.*

(b) Der Aitken-Schätzer für γ_1 *im von* Z^{*3} *erzeugten Modell* $Z^{*3} \sim (M^{*3}1_n\, \sigma^3\gamma_1;\ \sigma^6 S_{\mathcal{G}})$ *ist*

$$\widehat{g_1} := (\sigma^3 1_n' M^{*3} S_G^{+} M^{*3} 1_n)^{-1} 1_n' M^{*3} S_{\mathcal{G}}^{+} Z^{*3} \in \mathcal{G}_1 \ ;$$

$\widehat{g_1}$ *ist ein homogenes Polynom dritten Grades mit nur reinen Kuben* $Z_1^3,\dots,Z_n^3$, *erwartungstreu für* γ_1 *und hat unter allen diesen Schätzfunktionen kleinste Varianz bei Normalverteilungsannahme.*

(c) Der kleinste-Quadrate-Schätzer im Modell von $Z^{\otimes 3}$ *ist*

$$\overline{g_1} := (\sigma^3 1_n' M^{*3} 1_n)^{-1} 1_n' Z^{*3} \in \mathcal{G}_1 \ ;$$

$\overline{g_1}$ *ist ein homogenes Polynom dritten Grades mit nur reinen Kuben* $Z_1^3,\dots,Z_n^3$, *erwartungstreu und von kleinster Norm unter allen erwartungstreuen homogenen Polynomen dritten Grades in* $Z_1,\dots,Z_n$.

Unter Normalverteilungsannahme gelten ferner:

(d) $\operatorname{Var} \overline{g_1} \geq \operatorname{Var} \widehat{g_1} \geq \operatorname{Var} \widehat{\gamma_1}$.

(e) $\widehat{g_1}$ *ist von minimaler Varianz in* Γ_1 *genau dann, wenn*

$$\sum e_\nu' S_{\mathcal{G}}^{+} M^{*3} 1_n \ S_\Gamma e_\nu^{\otimes 3} = \rho (M e_\nu)^{\otimes 3} \ , \qquad \rho \in \mathbb{R} \ .$$

(f) $\overline{g_1}$ *ist von minimaler Varianz in* $\mathcal{G}_1$ *genau dann, wenn*

$$(\operatorname{Diag} M) M m = \rho\ M^{*3} 1_n \ , \qquad \rho = m' M m / 1_n' M^{*3} 1_n \ .$$

(g) $\overline{g_1}$ *ist von minimaler Varianz in* Γ_1 *genau dann, wenn*

$$\pi_s \cdot (\operatorname{vec} M) \otimes M m = \rho \sum (M e_\nu)^{\otimes 3} \ , \qquad \rho = m' M m / 1_n' M^{*3} 1_n \ .$$

10 Anscombe (1961 pp.14f) begnügt sich mit der Schätzung der dritten Momente $\mu_3 = \sigma^3\gamma_1$, für die Satz 9 entsprechend gilt mit $\sigma^3\,\widehat{\gamma_1}$, $\sigma^3\,\widehat{g_1}$ und $\sigma^3\,\overline{g_1}$, die Klassen Γ_1 und $\mathcal{G}_1$ bleiben dieselben. Die von Anscombe (op.cit. p.4) angegebene Schätzfunktion $\sigma^3 g_1$ ist gerade unser Kleinste-Quadrate-Schätzer $\sigma^3\,\overline{g_1}$.

Für sein erstes Beispiel (op.cit. p.14) $Y \sim ((0,1,2)'\mu;\ \sigma^2 I_3)$ verifiziert man mit Satz 9(c), daß

$$\sigma^3\,\overline{g_1} = \frac{125}{174} \sum Z_\nu{}^3$$

die Varianz wohl in $\mathcal{G}_1$, nicht aber in Γ_1 minimiert. Im zweiten Fall (op.cit. p.15) $Y \sim ((0,1,2,3)'\mu;\ \sigma^2 I_4)$ findet man, daß weder $\sigma^3\,\overline{g_1}$ in $\mathcal{G}_1$, noch $\sigma^3\,\widehat{g_1}$ in Γ_1 von kleinster Varianz ist.

Besondere Aufmerksamkeit widmet Anscombe der ***design condition I*** einer gemeinsamen Mittelwert-Komponente: $1_n \in \mathcal{R}X$, und der ***design-condition II*** von gleichen Varianzen der Komponenten von Z: $M_{11}=\cdots=M_{nn}$ (op.cit. p.3). Unter diesen Bedingungen, bemerkt er (op.cit. p.15), ist $\sigma^3\,\overline{g_1}$ von kleinster Varianz in $\mathcal{G}_1$. Dies ist richtig und folgt aus Satz 9(f), denn $Mm = M_{11} \cdot M1_n = 0$; aber nach Satz 9(g) wird die Varianz sogar in der weit größeren Klasse Γ_1 minimiert.

Kapitel 4

SCHÄTZUNG DES EXZEß

In diesem Kapitel geben wir Schätzfunktionen für den Exzeß an; dem Aufbau des vorigen Kapitels folgend, beschränken wir uns im wesentlichen auf die Angabe der Ergebnisse.

1 Wie im vorigen Kapitel sei ein klassisches Modell (§3.2) vorausgesetzt. Da in die Definition des Exzeß die affine Verschiebung -3 eingeht, bestehen die Schätzfunktionen hier aus homogenen Polynomen vierten Grades in $Z_1,\dots,Z_n$ plus Konstante:

$$\Gamma_2 := \{ w'Z^{\otimes 4} + c_\Gamma \mid w\in\mathbb{R}^{n^4} ,\ c_\Gamma \in \mathbb{R} \} ;$$

$$\mathcal{G}_2 := \{ w'Z^{*4} + c_{\mathcal{G}} \mid w\in\mathbb{R}^{n} ,\ c_{\mathcal{G}} \in \mathbb{R} \} .$$

Aus dem folgenden Hilfssatz errechnet man, daß bei Erwartungstreue für γ_2 notwendig gilt:

$$c_\Gamma = -3\ \sigma^4\ w'\pi_s\ M^{\otimes 4} \sum e_\nu\otimes e_\mu\otimes e_\nu\otimes e_\mu ;$$

$$c_{\mathcal{G}} = -3\ \sigma^4\ w'm^{*2} ;$$

dabei ist $m = (M_{11},\dots,M_{nn})'$, der Symmetrisator π_s wurde in §3.5 eingeführt.

2 HILFSSATZ. *Für ein klassisches Modell gilt:*

(a) $\mathcal{E}Z^{\otimes 4} = M^{\otimes 4}\ (\sum e_\nu^{\otimes 4})\ \sigma^4\gamma_2 + 3\ \sigma^4\ \pi_s\ M^{\otimes 4} \sum e_\nu\otimes e_\mu\otimes e_\nu\otimes e_\mu$; $\mathcal{E}Z^{*4} = M^{*4}1_n\ \sigma^4\gamma_2 + 3\ \sigma^4\ m^{*2}$.

(b) In Γ_2 ***oder*** $\mathcal{G}_2$ ***existiert eine erwartungstreue Schätzfunktion für*** γ_2 ***genau dann, wenn der Rang von*** X ***kleiner als*** n ***ist.***

BEWEIS. Die Behauptungen folgen analog wie in Hilfssatz 3.4. Für (b) gilt zunächst $\sum M_{\nu\mu}^4 \neq 0$, oder äquivalent $M\neq 0$, $\mathcal{R}X\neq\mathbb{R}^n$. ●

Bei einem klassischen Modell wird ungeschrieben vorausgesetzt, daß der Rang von X kleiner als n ist; folglich ist der

Exzeß γ_2 "immer" erwartungstreu schätzbar. Hilfssatz 2 führt zu Anscombe's (1961 p.7) Gleichungen 28 und 29.

3 Für die achten Momente definieren wir

$$A_\Gamma := 24\, M^{\otimes 4} + 12\, M^{\otimes 4} \cdot \sum_{i,j,k,\ell} \cdot$$
$$(E_{ij}\otimes E_{ij}\otimes E_{kk}\otimes E_{\ell\ell} + E_{ii}\otimes E_{jk}\otimes E_{jk}\otimes E_{\ell\ell} +$$
$$(E_{ii}\otimes E_{jj}\otimes E_{k\ell}\otimes E_{k\ell} + E_{ik}\otimes E_{jj}\otimes E_{ik}\otimes E_{\ell\ell} +$$
$$(E_{i\ell}\otimes E_{jj}\otimes E_{kk}\otimes E_{i\ell} + E_{ii}\otimes E_{j\ell}\otimes E_{kk}\otimes E_{j\ell})\cdot M^{\otimes 4} ;$$

$$A_{\mathcal{G}} := 24\, M^{*4} + 72\, (\mathrm{Diag}\, M)\, M^{*2}\, (\mathrm{Diag}\, M).$$

4 HILFSSATZ. *Für ein quasinormales Modell gilt:*

$$\mathcal{D} Z^{\otimes 4} = \sigma^8\, \pi_s A_\Gamma \pi_s \; ; \qquad \mathcal{D} Z^{*4} = \sigma^8\, A_{\mathcal{G}} \; .$$

Der Beweis verläuft wie bei Hilfssatz 3.8. Mit der Formel $\mathcal{E}(\sum Z_\nu^4)^2 = \mathrm{Var}\, 1_n' Z^{*4} + (\mathcal{E}\, 1_n' Z^{*4})^2$ und den Hilfssätzen 2 und 4 kann Anscombe's (1961 p.7) Gleichung 33 verifiziert werden.

Parallel zu Satz 3.9 erhalten wir jetzt die folgenden Ergebnisse:

5 SATZ(*Schätzfunktionen für den Exzess* γ_2)

Gegeben sei ein lineares Modell $Y \sim (Xb;\sigma^2 I_n)$, *in dem* Y *unabhängige Komponenten mit gleichem Exzess* γ_2 *besitze. Setze* $m := (M_{11},\dots,M_{nn})'$. *Dann gilt für die* §1 *beschriebenen Klassen* Γ_2 *und* $\mathcal{G}_2$ *von Schätzfunktionen mit den in* §3 *definierten Matrizen* A_Γ *und* $A_{\mathcal{G}}$ *der achten Momente und* $Z := MY$:

(a) Der Aitken-Schätzer für γ_2 *im von* $Z^{\otimes 4}$ *erzeugten Modell* $Z^{\otimes 4} - 3\,\sigma^4\, \pi_s\, M^{\otimes 4} \sum e_\nu\otimes e_\mu\otimes e_\nu\otimes e_\mu \sim (M^{\otimes 4}(\sum e_\nu^{\otimes 4})\sigma^4\gamma_2;\ \sigma^8 A_\Gamma)$ *ist*

$$\widehat{\gamma_2} := \frac{\sigma^{-4}\sum e_\nu^{\otimes 4}{}' A_\Gamma^+ Z^{\otimes 4} - 3\sum e_\nu^{\otimes 4}{}' A_\Gamma^+ (e_\mu\otimes e_\lambda\otimes e_\mu\otimes e_\lambda)}{\sum e_\nu^{\otimes 4}{}' A_\Gamma^+ e_\mu^{\otimes 4}} \in \Gamma_2 \; ;$$

$\widehat{\gamma_2}$ *ist bis auf die Konstante ein homogenes Polynom vierten Grades in* $Z_1,\dots,Z_n$, *erwartungstreu für* γ_2 *und hat unter allen diesen Schätzfunktionen kleinste Varianz bei Normalverteilungsannahme.*

(b) Der Aitken-Schätzer für γ_2 *im von* Z^{*4} *erzeugten Modell* $Z^{*4} - 3\,\sigma^4\, m^{*2} \sim (M^{*4}1_n\, \sigma^4\gamma_2;\ \sigma^8\, A_{\mathcal{G}})$ *ist*

$$\widehat{g_2} := \frac{\sigma^{-4}\, 1_n{}'M^{*4}A_{\mathcal{G}}{}^{+}Z^{*4} - 3\ 1_n{}'M^{*4}A_{\mathcal{G}}{}^{+}m^{*4}}{1_n{}'M^{*4}A_{\mathcal{G}}{}^{+}M^{*4}1_n} \in \mathcal{G}_2\ ;$$

$\widehat{g_2}$ *ist bis auf die Konstante ein homogenes Polynom vierten Grades mit nur reinen vierten Potenzen* $Z_1{}^4,\dots,Z_n{}^4$, *erwartungstreu für* γ_2 *und hat unter allen diesen Schätzfunktionen kleinste Varianz bei Normalverteilungsannahme.*

(c) Der Kleinste-Quadrate-Schätzer im Modell von $Z^{\otimes 4}$ *ist*

$$\overline{g_2} := (1_n{}'M^{*4}1_n)^{-1}\,(\sigma^{-4}1_n{}'Z^{*4} - 3m'm) \in \mathcal{G}_1\ ;$$

$\overline{g_2}$ *ist bis auf die Konstante ein homogenes Polynom vierten Grades mit nur reinen vierten Potenzen* $Z_1{}^4,\dots,Z_n{}^4$, *erwartungstreu und von kleinster Norm unter allen bis auf eine Konstante homogenen Polynomen vierten Grades in* $Z_1,\dots,Z_n$.

Unter Normalverteilungsannahme gelten ferner:

(d) $\mathrm{Var}\ \overline{g_2} \geq \mathrm{Var}\ \widehat{g_2} \geq \mathrm{Var}\ \widehat{\gamma_2}$.

(e) $\widehat{g_2}$ *ist von minimaler Varianz in* Γ_2 *genau dann, wenn*

$$\sum e_\nu{}'A_{\mathcal{G}}{}^{+}M^{*4}1_n\ A_\Gamma e_\nu{}^{\otimes 4} = \rho(\sum Me_\nu)^{\otimes 4}\ ,\qquad \rho \in \mathbb{R}\ .$$

(f) $\overline{g_2}$ *ist von minimaler Varianz in* $\mathcal{G}_2$ *genau dann, wenn* (§3.5)

$$(\mathrm{Diag}\ M)M^{*2}m = \rho\ M^{*4}1_n\ ,\qquad \rho = m'M^{*2}m/1_n{}'M^{*4}1_n\ .$$

(g) $\overline{g_2}$ *ist von minimaler Varianz in* Γ_2 *genau dann, wenn*

$$\pi_s\cdot(\mathrm{vec}\ M)\otimes(\mathrm{vec}\ M(\mathrm{Diag}\ M)M\ ) = \rho \sum (Me_\nu)^{\otimes 4}\ ,\qquad \rho \textit{ wie in (f).}$$

6 Die Aussagen (e) bis (g) der Sätze 5 und 3.9 ähneln dem Satz von Hsu (Pukelsheim 1977 Theorem 4.2). Sie variieren dasselbe formale Problem: Wann ist der Kleinste-Quadrate-Schätzer von minimaler Varianz? etc.

Wir ergänzen nun die in §3.10 genannten Beispiele von Anscombe. Im ersten Fall ist

$$\overline{g_2} = \sigma^{-4}\,\frac{625}{914}\sum Z_\nu{}^4 - \frac{3150}{914}$$

zwar in $\mathcal{G}_2$, nicht aber in Γ_2 von kleinster Varianz. Im zweiten Fall minimiert $\overline{g_2}$ die Varianz in $\mathcal{G}_2$ nicht; für $\widehat{g_2}$ erhalten wir:

$$\widehat{g_2} = \sigma^{-4}(0.484\ Z_1^4 + 0.484\ Z_2^4 + 0.596\ Z_3^4 - 0.400\ Z_4^4) - 3.842\ .$$

Im allgemeinen erfordert die Rechnung im "großen" Modell von $Z^{\otimes 4}$ die Inversion einer $n^4 \times n^4$-Matrix, und im "kleinen" Modell von Z^{*4} die Inversion einer $n\times n$-Matrix.

Als ***design condition III*** fordern wir (vergleiche Anscombe 1961 pp.6,20), daß die Komponenten von Z bis auf Permutationen gleiche Kovarianzen haben, genauer: die Zeilen von M seien bis auf Permutationen einander gleich. Gelten die design-condition II (§3.10) und III, dann hat $\overline{g_2}$ minimale Varianz in $\mathcal{G}_2$; denn für Satz 5(f) ist einerseits $(\mathrm{Diag}\ M)M^{*2}m = M_{11}^2\ M^{*2}\ 1_n$ wegen gleicher Zeilensummen ein Vielfaches von 1_n, andererseits aus demselben Grund auch $M^{*4}1_n$. Im Gegensatz zu §3.10 ist dies ein nichttriviales Beispiel mit $\rho \neq 0$.

7 Wir wollen nun in $\overline{g_2}$ den unbekannten Wert σ^{-4} durch eine Schätzung ersetzen und die so erhaltene Schätzfunktion $\widetilde{g_2}$ mit dem Anscombe'schen Schätzer g_2 vergleichen. Dazu definieren wir

$$\nu := \mathrm{rank}\ M = n - \mathrm{rank}\ X\ ; \quad d := \sum M_{ij}^4\ .$$

Eine übliche und nach Kapitel 2 sinnvolle Schätzfunktion für σ^2 ist $s^2 := \nu^{-1}Y'MY$. Schreibt man in $\overline{g_2}$ einfach s^2 für σ^2, dann geht selbst unter Normalverteilung die Erwartungstreue verloren:

$$\mathcal{E}\ s^{-4}\ 1_n'Z^{*4} - 3m'm = -6(\nu+2)^{-1}\ m'm\ .$$

Die diesen Fehler ausgleichende Korrektur führt zu

$$\widetilde{g_2} := d^{-1}\ (\ s^{-4}\sum Z_i^4 - 3(\nu+2)^{-1}\ \nu\ m'm\ )\ .$$

Anscombe (1961 p.7) verwendet einen anderen Divisor:

$$g_2 = \tfrac{d}{D}\ \widetilde{g_2}\ ; \quad D := d - 3[\nu(\nu+2)]^{-1}[\sum M_{ii}^2]^2\ .$$

Normalverteilungsannahme führt bei beiden Schätzern zum Mittelwert 0, sie sind also lokal erwartungstreu. Wegen $D/d < 1$ hat $\widetilde{g_2}$ kleinere Varianz als g_2. Die asymptotischen Eigenschaften sind dieselben: für eine Folge von Matrizen $X_n \in \mathbb{R}^{n\times p_n}$ mit beschränkten Rängen: $\sup \{\ \mathrm{rank}\ X_n \mid n=1,2,\dots\ \} < p$, gilt:

$$\begin{aligned} 0 &\le 1-D/d = 3\ (1_n'm^{*2})^2/[\nu(\nu+2)\sum M_{ij}^4] \\ &\le 3n \sum M_{ii}^4/[\nu(\nu+2)(\sum M_{ii}^4 + \textstyle\sum_{i\neq j} M_{ij}^4)] \\ &\le 3n/[(n-p)(n-p+2)] \quad \to \quad 0 \quad \text{für} \quad n \to \infty\ . \end{aligned}$$

Dieser Vergleich spricht eher für die Statistik $\widetilde{g_2}$ als für g_2.

8 Wir beschließen diesen Teil mit einer kurzen Rückschau. Die Trennung der Mittelwert- und Streuungsschätzung etwa in Searle's Buch (1971) oder der von C.R. Rao (1973) für seine MINQUE-Theorie eigens bereitgestellte Apparat suggerieren Unterschiede, wo keine sind. Also haben wir Gewicht darauf gelegt, eine gemeinsame Grundlegung von Mittelwert- und Streuungsschätzung vorzuführen. Die Stärke dieses einheitlichen Aufbaus mag man daran messen, daß er ohne Bruch die bisher abgetrennte Schätzung von Schiefe und Exzeß mit einbezieht.

In der zweiten Hälfte dieser Arbeit wenden wir uns einem Problem zu, das an einer Stelle mit alarmierender Deutlichkeit auffällt: der Möglichkeit, negative Schätzwerte zu erhalten für nichtnegative Varianzkomponenten.

Teil III

NICHTNEGATIV-DEFINITE SCHÄTZUNG DER STREUUNGSMATRIX

Kapitel 1

DER FUNDAMENTALDEFEKT LINEARER MODELLE

In diesem Kapitel verstehen wir negative Schätzwerte für Varianzkomponenten als Beispiel des allgemeineren Fundamentaldefekts linearer Modelle. Wir unterscheiden zwei Ausprägungen: die NND Schätzung der gesamten Streuungsmatrix und die nichtnegative Schätzung einzelner Varianzkomponenten.

1 Der Definition I.1.3 linearer Modelle $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ ist als Präzisierung beigefügt, wieviele Parameterwerte $b\in\mathbb{R}^p$ und $t\in\mathbb{R}^k$ bei zugrundeliegender Verteilungsklasse $\mathcal{P}$ existieren sollen:

$$\operatorname{span} \mathcal{E}_{\mathcal{P}}Y = \operatorname{span} \{ x_1,\dots,x_p \} ,$$

$$\operatorname{span} \mathcal{D}_{\mathcal{P}}Y = \operatorname{span} \{ V_1,\dots,V_k \} .$$

Die genaue Parametermenge und ihre lineare Hülle, etwa

$$\mathcal{E}_{\mathcal{P}}Y \quad \text{und} \quad \operatorname{span} \mathcal{E}_{\mathcal{P}}Y$$

wurden im Fortgang der Theorie nicht weiter auseinandergehalten. Demzufolge können die hergeleiteten Schätzfunktionen Werte liefern, die zwar in der linearen Hülle span $\mathcal{E}_{\mathcal{P}}Y$, nicht aber in der Parametermenge $\mathcal{E}_{\mathcal{P}}Y$ liegen. Solche Werte sind mindestens modellwidrig und schlimmstenfalls unrealisierbar. Dies nennen wir den ***Fundamentaldefekt linearer Modelle***.

2 Beim Mittelwert kann man den Fundamentaldefekt noch "wegdefinieren", indem die Verteilungsklasse $\mathcal{P}$ als so groß angesetzt wird, daß ihre Erwartungswerte selbst schon einen linearen Raum bilden: $\mathcal{E}_{\mathcal{P}}Y = \operatorname{span} \mathcal{E}_{\mathcal{P}}Y$. Das ist schon nicht mehr möglich für die Differenz zwischen span $\mathcal{D}_{\mathcal{P}}Y$ und $\mathcal{D}_{\mathcal{P}}Y$: niemals sind alle Matrizen eines Teilraums symmetrischer Matrizen NND. Im allgemeinen kann der Fundamentaldefekt also nicht weggeredet werden; bezüglich der Varianz-Kovarianz-Komponenten müssen wir sogar gleich zwei Ausprägungen unterscheiden.

Ein lineares Modell $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ macht genau dann Sinn, wenn man unter Linearkombinationen der k zerlegenden Matrizen V_κ nur diejenigen zuläßt, die eine NND Summe $\sum t_\kappa V_\kappa$ ergeben. Folglich sollten auch die k Schätzwerte $\hat{t}_\kappa$ zusammen eine NND Schätzung $\sum \hat{t}_\kappa V_\kappa$ der Streuungsmatrix liefern. Dieses Problem behandeln wir im nächsten Kapitel dieses Teils.

Es ist aber auch der Fall von Bedeutung, bei dem die Nebenbedingung nicht den gesamten Vektorparameter t, sondern einzeln seine k Komponenten t_κ betrifft: Bei Varianzkomponenten-Modellen ist jedes der t_κ nichtnegativ, gleiches sollte dann auch für eine Schätzung $\hat{t}_\kappa$ gelten. Dieser nichtnegativen Schätzung von Varianzkomponenten ist der gesamte Teil IV gewidmet.

In dieser Arbeit konzentriert sich unser Interesse auf die Schätzung von Varianz-Kovarianz-Komponenten, doch zeigen wir im letzten Kapitel IV.5 (***Abschließender Ausblick***) Verbindungen auf, inwieweit die folgenden Untersuchungen zu einer allgemeinen Lösung des Fundamentaldefekts beitragen können.

Kapitel 2

EINE HINREICHENDE BEDINGUNG ZUR NICHTNEGATIV-DEFINITEN SCHÄTZUNG DER STREUUNGSMATRIX

In diesem Kapitel erörtern wir, wann eine NND geschätzte Streuungsmatrix vorliegt. Wir können diese Frage nur teilweise und im durch Invarianz reduzierten Modell beantworten: als hinreichend erweist sich, daß die zerlegenden Matrizen einen quadratischen Teilraum aufspannen.

1 Die Aussagen dieses Kapitels betreffen nicht ein ursprüngliches lineares Modell $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ sondern das M-reduzierte Modell $Z := MY \sim (0; \sum t_\kappa M V_\kappa M)$ (§I.2.6). Wir formulieren zunächst das Ergebnis, rechtfertigen anschließend die dabei verwendeten Ausdrücke und bereiten mit einem Satz über Maximum Likelihood Schätzung unter Normalverteilung den Beweis vor, der in §5 folgt.

2 SATZ. (*Eine hinreichende Bedingung*)

Gegeben seien a *unabhängige, identisch verteilte Zufallsvektoren* $Y_1, \ldots, Y_a$ *des* $\mathbb{R}^n$ *mit Mittelwert* $\sum b_\pi x_\pi$ *und Streuungsmatrix* $\sum t_\kappa V_\kappa$. *Die* k *zerlegenden Matrizen* $MV_\kappa M$ *seien linear unabhängig, und der von ihnen aufgespannte Teilraum* $\mathcal{B}$ *sei quadratisch und enthalte* M.

Für alle Realisierungen $y_1, \ldots, y_a$, *für die im* M*-reduzierten Modell die Stichprobenstreuungsmatrix* $S := a^{-1} \sum My_\alpha (My_\alpha)'$ *vom selben Rang* ν *ist wie* M, *ergibt die erwartungstreue invariante quadratische Schätzfunktion von kleinster Norm*

$$\hat{t} = (D_M' D_M)^{-1} D_M' \operatorname{vec} S$$

eine NND Schätzung $\sum \hat{t}_\kappa M V_\kappa M$ *mit Rang* ν *für die Streuungsmatrix des* M*-reduzierten Modells.*

3 Die a Zufallsvektoren MY_α können wir zu einem großen Modell zusammenbauen:

$$\tilde{Z} := [MY_1':\cdots:MY_a']'; \quad \tilde{V}_\kappa := I_a \otimes MV_\kappa M; \quad \tilde{Z} \sim (0; \sum t_\kappa \tilde{V}_\kappa).$$

Dafür lautet die Normalgleichung: $\tilde{D}'\tilde{D}\hat{t} = \tilde{D}'\cdot\tilde{Z}\otimes\tilde{Z}$, wobei $\tilde{D}=[\mathrm{vec}\ \tilde{V}_1:\cdots:\mathrm{vec}\ \tilde{V}_k]$; das ergibt: $D_M'D_M\hat{t} = D_M'\mathrm{vec}\ S$ wie in Satz 2.

Statt im M-reduzierten Modell kann man auch im ursprünglichen Modell argumentieren. Dann haben wir gerade bei a unabhängigen und identisch verteilten $Y_\alpha \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ erst durch Invarianz reduziert und danach das vergrößerte Modell von $\tilde{Z}$ zusammengebaut.

Es bieten sich weitere zwei Möglichkeiten an. Brown (1976 p.748) bildet erst ein vergrößertes Modell:

$$\tilde{Y} := [Y_1':\cdots:Y_a']'; \quad \tilde{X} := 1_a\otimes X; \quad \tilde{\tilde{V}}_\kappa := I_a\otimes V_\kappa; \quad \tilde{Y} \sim (\tilde{X}b; \sum t_\kappa \tilde{\tilde{V}}_\kappa);$$

anschließend reduziert er durch Invarianz: dies führt ***nicht*** zu der vorher angegebenen Normalgleichung.

Dewess (1975 p.501) arbeitet - wie in der multivariaten Analysis - nicht mit gleichen Mittelwerten $\mathcal{E}Y_\alpha = Xb$, sondern läßt a verschiedene zu: $\mathcal{E}Y_\alpha = Xb_\alpha$. Dies führt zum vergrößerten Modell mit $\tilde{Y}$ und $\tilde{\tilde{V}}_\kappa$ wie vor, aber

$$\tilde{\tilde{X}} := I_a\otimes X; \quad \tilde{\tilde{b}} := [b_1':\cdots:b_a']'; \quad \tilde{Y} \sim (\tilde{\tilde{X}}\tilde{\tilde{b}}; \sum t_\kappa \tilde{\tilde{V}}_\kappa).$$

Reduktion durch Invarianz ergibt jetzt wieder das Modell von $\tilde{Z}$. Es ist also nicht unberechtigt, $\hat{t}$ als erwartungstreue invariante quadratische Schätzfunktion mit kleinster Norm zu bezeichnen.

Der folgende Satz bringt eine Aussage über ein normales lineares Modell, als normalverteilungsfreies Korollar enthält er aber auch Satz 2.

4 SATZ. (*ML-Schätzer für* t *bei quadratischem Teilraum*)

Gegeben seien a *unabhängige, identisch normalverteilte Zufallsvektoren* $Z_1,\ldots,Z_a$ *des* $\mathbb{R}^\nu$ *mit Mittelwert Null und Streuungsmatrix* $\sum t_\kappa W_\kappa$. *Die* k *zerlegenden Matrizen* W_κ *seien linear unabhängig, und der von ihnen aufgespannte Teilraum* $\mathcal{B}$ *sei quadratisch und enthalte eine positiv-definite Matrix. Die Teilmenge* G *des* $\mathbb{R}^k$ *sei das Gebiet derjenigen Werte* t *des Streuungsparameters, für die* $\sum t_\kappa W_\kappa$ *positiv-definit ist.*

(a) Die erwartungstreue quadratische Schätzfunktion $\hat{t}$ *mit gleichmäßig kleinster Varianz löst die Likelihood-Gleichung.* $\mathcal{B}$ *enthält die Einheitsmatrix; am einfachsten erhält man* $\hat{t}$ *daher mit der Stichprobenstreuungsmatrix* $S := a^{-1}\sum Z_\alpha Z_\alpha'$ *aus der Normalgleichung* $D'D\hat{t} = D'\text{vec } S, \quad D := [\text{vec } W_1 \vdots \cdots \vdots \text{vec } W_k]$.

(b) Wenn die Stichprobenstreuungsmatrix S *positiv-definit ist – und für* $a \geq \nu$ *ist das Lebesgue-fast sicher der Fall –, dann ist* $\hat{t}$ *der Maximum-Likelihood-Schätzer für* $t \in G$ *und* $\sum \hat{t}_\kappa W_\kappa$ *positiv-definit.*

BEWEIS. (a) Die Likelihood-Gleichung lautet

$$(*) \quad D'[(\sum \hat{t}_\kappa W_\kappa) \otimes (\sum \hat{t}_\kappa W_\kappa)]^{-1} D\hat{t} = D'[(\sum \hat{t}_\kappa W_\kappa) \otimes (\sum \hat{t}_\kappa W_\kappa)]^{-1} \text{vec } S \ .$$

(vergleiche Anderson 1970 p.4). Nach II.1.7 und II.2.4 ist (*) gleichzeitig eine gewichtete Normalgleichung zur Aitken-Schätzung, die die erwartungstreue quadratische Schätzfunktion mit kleinster Varianz lokal unter $\sum \hat{t}_\kappa W_\kappa$ bestimmt.

Nach Voraussetzung existiert eine positiv-definite Matrix $B \in \mathcal{B}$. Für quadratische Teilräume gilt (Seely 1971 pp.711f) dann $B^{-1} \in \mathcal{B}$, und $I_\nu = \frac{1}{2}(BB^{-1}+B^{-1}B) \in \mathcal{B}$. Nach Satz 11.2.5 ist $\hat{t}$ von gleichmäßig kleinster Varianz; für $\hat{t} \in G$ also auch lokal unter $\sum \hat{t}_\kappa W_\kappa$. Somit löst $\hat{t}$ die Gleichung (*), und am einfachsten bestimmt man $\hat{t}$ als lokalen Minimum Varianz Schätzer unter I_ν.

(b) Die Stichprobenstreuungsmatrix S ist Lebesgue-fast sicher positiv-definit (C.R. Rao 1973 pp.559,598). - G wird mit Recht Gebiet genannt, denn G ist offen und zusammenhängend. Sein Rand ∂G besteht aus denjenigen $t \in \mathbb{R}^k$, für die $\sum t_\kappa W_\kappa$ NND und singulär ist.

Sei nun S positiv-definit. Wenn t gegen den Rand ∂G oder $\|t\|$ gegen ∞ konvergiert, dann strebt die Likelihood-Funktion $L(S,t)$ gegen Null (Anderson 1970 p.5). Da die Likelihood-Funktion für $t \in G$ positiv ist, existiert mindestens ein Maximum in G, kein Maximum liegt auf dem Rand ∂G, und der Maximum-Likelihood-Schätzer löst die Likelihood-Gleichung (*).

Als Lösung von (*) ist der ML-Schätzer erwartungstreu für t. Nach Seely (1971 p.716) existiert bei quadratischem Teilraum $\mathcal{B}$ eine vollständige und suffiziente Statistik für $t \in G$. Wegen des

Neyman-Kriteriums (Witting 1966 p.150) ist die Likelihood-Funktion und damit auch jede Lösung von (*) eine Funktion dieser Statistik. Nach dem Satz von Lehmann-Scheffé (Witting 1966 p.153) ist also der ML-Schätzer gleich der erwartungstreuen Schätzfunktion mit gleichmäßig kleinster Varianz. Diese ist Lebesgue-fast sicher eindeutig bestimmt, und eine ihrer Versionen ist $\hat{t}$, denn die von Seely (1971 p.716) als vollständig und suffizient nachgewiesene Statistik ist D'vec S. Also ist $\hat{t}$ gleich dem ML-Schätzer und bildet nach G ab, wenn auch zunächst nur Lebesgue-fast sicher.

Um diese Einschränkung aufzuheben, betrachten wir $\hat{t}$ nicht im Sinne von §3 als quadratische Funktion, sondern mit der Variablen S als Abbildung auf Sym(n). Als linearer Operator ist $\hat{t}$ aber offen, weil surjektiv. Da nun die positiv-definiten Matrizen offen in Sym(n) sind (IV.1.5.b), gilt dasselbe für ihr $\hat{t}$-Bild: Liegt für irgendeine positiv-definite Matrix S_0 der Wert $(D'D)^{-1}D'\mathrm{vec}\, S_0$ nicht in G, so gilt das gleiche auch für eine offene Umgebung von S_0 und somit für eine Menge positiven Lebesgue-Maßes; das kann nicht sein. ●

Wir ziehen aus diesem Satz drei Schlußfolgerungen. Erstens braucht für eine ausgezeichnete Klasse linearer Modelle die Likelihood-Gleichung nicht iterativ gelöst zu werden (Anderson 1970; Corbeil & Searle 1976a). Corbeil & Searle (1976a) betrachten Maximum-Likelihood-Schätzung im durch Invarianz reduzierten Modell und nennen dies ***Restricted maximum likelihood estimation (REML)***. Für die von ihnen (1976b) untersuchten vier Beispiele haben wir in §II.2.6 nachgewiesen, daß ein quadratischer Teilraum $\mathcal{B}$ vorliegt: Zweitens verschafft vorstehender Satz der von ihnen beobachteten Gleichheit der REML- und der herkömmlichen Schätzer theoretische Begründung. Drittens folgt Satz 2 als Korollar:

5 BEWEIS VON SATZ 2. Die maximalinvariante Statistik MY (§I.2.6) ist für Maximum-Likelihood Untersuchungen nicht geeignet, da sie zu Verteilungen führt, die auf den echten Teilraum $\mathcal{R}M$ des Stichprobenraumes $\mathbb{R}^n$ konzentriert sind. Wählt man für M eine volle-Rang-Faktorisierung (Ben-Israel & Greville 1974 pp.22,49):

$$M = QQ', \quad Q'Q = I_\nu, \quad Q \in \mathbb{R}^{n\times\nu},$$

dann ist die Statistik Q'Y ebenfalls maximalinvariant (Seely 1971 p.718; Kleffe 1976 p.707). Die a Zufallsvektoren $Z_\alpha := Q'Y_\alpha$ sind unabhängig und identisch verteilt mit Mittelwert Null und Streuungsmatrix $\sum t_\kappa W_\kappa$, wobei die k zerlegenden Matrizen $W_\kappa := Q'V_\kappa Q$ linear unabhängig sind und einen quadratischen Teilraum aufspannen, der I_ν enthält. Hat die M-reduzierte Stichprobenstreuungsmatrix $a^{-1}\sum My_\alpha(My_\alpha)'$ den Rang ν, so auch die Q'-reduzierte $a^{-1}\sum z_\alpha z_\alpha'$. Satz 4 besagt, daß

$$\hat{t} = (D'\cdot QQ'\otimes QQ'\cdot D)^{-1}D'\cdot QQ'\otimes QQ'\cdot \text{vec } a^{-1}\sum Y_\alpha Y_\alpha' = (D_M'D_M)^{-1}D_M'\text{vec } S$$

eine positiv-definite Matrix $\sum \hat{t}_\kappa Q'V_\kappa Q$ liefert. Dann ist $\sum \hat{t}_\kappa MV_\kappa M$ NND und vom Rang ν, wie behauptet. ●

6 Im nächsten Teil IV wenden wir uns der nichtnegativen Schätzung von Varianzkomponenten zu. Unter gewissen Bedingungen wird dadurch dieselbe Aufgabe gestellt wie bei der NND Schätzung der Streuungsmatrix. Sind nämlich die k zerlegenden Matrizen V_κ NND, und kann der von V_κ erfaßte Raum $\mathcal{R}V_\kappa$ nicht schon durch die übrigen V_λ, $\lambda \neq \kappa$, erklärt werden, genauer:

$$\mathcal{R}V_\kappa \not\subset \mathcal{R}\textstyle\sum_{\lambda\neq\kappa} V_\lambda \ , \quad \text{für alle } \kappa = 1,\dots,k \ ,$$

dann ist äquivalent, ob die geschätzte Streuungsmatrix $\sum \hat{t}_\kappa V_\kappa$ NND ist, oder die k Komponenten $\hat{t}_\kappa$ nichtnegativ sind; ein Beispiel dafür stellen Varianzkomponenten-Modelle dar, in denen jede Komponente t_κ erwartungstreu NND quadratisch schätzbar ist (Korollar IV.1.12). Beschränkt auf solche Modelle, bietet Teil IV eine zweite Teillösung für die NND Schätzung der Streuungsmatrix.

Teil IV

NICHTNEGATIVE SCHÄTZER IN VARIANZKOMPONENTEN-MODELLEN ALS LÖSUNGEN KONVEXER PROGRAMME

Kapitel 1

ERWARTUNGSTREUE NICHTNEGATIVE SCHÄTZBARKEIT

Zunächst illustrieren wir an einem Beispiel, daß die herkömmliche Theorie negative Schätzwerte für Varianzkomponenten nicht ausschließt. Nach einigen Hilfsaussagen über NND Matrizen folgt als Hauptergebnis dieses Kapitels eine Charakterisierung derjenigen Streuungszerlegungen, deren Varianzkomponenten erwartungstreu NND quadratisch schätzbar sind.

1 Im gesamten Teil IV betrachten wir Varianzkomponenten-Modelle $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$, für die nach Definition I.1.5 die k zerlegenden Matrizen V_κ NND und damit selbst Streuungsmatrizen sind, und der Streuungsparameter t im abgeschlossenen positiven Quadranten $\mathbb{R}^k_+$ variiert.

Für die in §I.3.6 beschriebenen Schätzverfahren bewirkt die Reduktion über die Streuungs-Mittelwert-Korrespondenz lediglich, daß man sich auf Schätzfunktionen für den Streuungsparameter beschränkt, die quadratisch sind. Das ist eine unvollkommene Qualifizierung, denn dadurch können die in solchem Sinne besten Schätzer immer noch zu negativen Werten für die nichtnegativen Varianzkomponenten führen. Dies beschreibt die zweite, für die Schätzung des Streuungsparameters t wesentliche Ausprägung des Fundamentaldefekts III.1.3; die Menge $\mathbb{R}^k_+$ der zulässigen Parameterwerte wird einer einfachen Theorie zuliebe, aber ansonsten sinnlos auf den linearen Raum $\mathbb{R}^k$ vergröbert.

Negative Schätzwerte für Varianzkomponenten stellen ein zentrales Problem der herkömmlichen Theorie dar; eine mathematische Behandlung existiert nicht und wird durch Ratschläge ersetzt (Searle 1971 pp.406-408). Den einzigen theoretischen Trost schöpft man aus Konsistenzaussagen, nach denen die defekten Schätzer schließlich fast sicher nichtnegativ werden (Brown 1976).

2 BEISPIEL. ***Heteroskedastische Varianzen.*** Für dieses Modell hatten wir in §II.1.9 den Mittelwert geschätzt, jetzt interessieren uns die Varianzkomponenten σ_κ^2. Sind diese k Komponenten positiv, dann ist nach Satz II.1.6(b) (im abgeleiteten Modell) der Aitken-Schätzer für $\sigma_1^2,\dots,\sigma_k^2$ von kleinster quasinormaler Varianz unter $\sigma_1^2,\dots,\sigma_k^2$; dieser Schätzer ist jedoch nicht praktikabel, denn um ihn zu berechnen, werden die unbekannten und zu schätzenden Parameter σ_κ^2 schon benötigt.

Als handhabbares Verfahren bleibt die Minimierung der Norm unter allen erwartungstreuen invarianten quadratischen Schätzfunktionen (C.R. Rao's MINQUE). Für den in diesem Sinn besten Schätzer $Y'\widehat{A}_\kappa Y$ einer einzelnen Komponente σ_κ^2 ergibt Satz II.2.2:

$$\widehat{A}_\kappa = \frac{n}{n_\kappa(n-2)} M_n \operatorname{Diag}[0:I_{n_\kappa}:0]\, M_n - \frac{1}{(n-1)(n-2)} M_n \ ;$$

$$Y'\widehat{A}_\kappa Y = \frac{n}{n_\kappa(n-2)} \sum_{\nu=1}^{n_\kappa}(Y_{\kappa\nu} - \overline{Y_{\cdot\cdot}})^2 - \frac{1}{n-2}\, s^2 \ ;$$

$$s^2 := \frac{1}{n-1} \sum (Y_{\kappa\nu} - \overline{Y_{\cdot\cdot}})^2 \ .$$

Die Schätzfunktion $Y'\widehat{A}_\kappa Y$ kann allerdings zu negativen Werten für den nichtnegativen Parameter σ_κ^2 führen. Beispielsweise erhält man den negativen Wert $-[(n-1)(n-2)]^{-1}\,\|y\|^2$ für jeden Punkt $y \neq 0$ des Stichprobenraumes $\mathbb{R}^n$, der in dem $(n-n_\kappa-1)$-dimensionalen Durchschnitt des Nullraumes von $M_n \operatorname{Diag}[0:I_{n_\kappa}:0]\, M_n$ und des Bildraumes von M_n liegt.

Auf einen total pathologischen Spezialfall weisen Horn & Horn (1975 p.876) hin. Für $k = 3$ und $n_1 = n_2 = n_3 = 1$ errechnet man:

$$Y'\widehat{A}_1 Y \cdot Y'\widehat{A}_2 Y \cdot Y'\widehat{A}_3 Y = -\,(Y_1-Y_2)^2(Y_1-Y_3)^2(Y_2-Y_3)^2 \ .$$

Bei zugrundeliegenden stetigen Verteilungen, etwa Normalverteilungen, wird also ***mit Wahrscheinlichkeit Eins*** eine der drei Varianzkomponenten negativ geschätzt! Es ist äußerst unbefriedigend, daß die Theorie einen solchen Unsinn nicht ausschließt.

3 Der Fundamentaldefekt kann entstehen, weil das formalisierte Problem nicht alle Informationen widerspiegelt, die über das Modell vorliegen: $Y'\widehat{A}Y$ ist genau dann erwartungstreue invariante quadratische Schätzfunktion von kleinster Norm (MN-UB-IQE; MINQUE) für $q't$ ($q \in \mathbb{R}^k$), wenn $\widehat{A}$ eine optimale Lösung des herkömmlichen Programms H ist:

$$H \quad \begin{cases} \|A\|^2 = \inf \\ A \in C(q) \\ A \in L := \{ MAM \mid A \in Sym(n) \} ; \end{cases}$$

dabei kennzeichnet $C(q) \subset Sym(n)$ die Menge aller symmetrischen $n \times n$-Matrizen A, die zu erwartungstreuen quadratischen Schätzfunktionen $Y'AY$ für $q't$ führen. Die optimalen Lösungen $\hat{A}$ von H können NND sein, müssen es aber nicht. Im folgenden nennen wir $Y'\hat{A}Y$ kurz den ***besten defekten Schätzer*** für $q't$.

Der Fundamentaldefekt wird vermieden, wenn die Norm unter allen erwartungstreuen NND invarianten quadratischen Schätzfunktionen minimiert wird; dazu definieren wir das Programm P:

$$P \quad \begin{cases} \|A\|^2 = \inf \\ A \in C(q) \\ A \in NND(n) \quad ; \end{cases}$$

dabei bezeichnet NND(n) die Menge aller nichtnegativ-definiten symmetrischen $n \times n$-Matrizen. Optimale Lösungen dieses Programms versehen wir mit einem Stern; $Y'A^*Y$ heißt kurz der ***beste nicht-negative Schätzer*** für $q't$.

Nach einigen grundlegenden Hilfsaussagen zeigen wir, daß P ein "Teilprogramm" von H ist, und wann für P zulässige Lösungen existieren.

4 HILFSSATZ.

(a) Für zwei Matrizen A *und* B *mit gleich vielen Zeilen gilt:*

$$\mathcal{R}A \perp \mathcal{R}B \quad \Leftrightarrow \quad B'A = 0 \; .$$

(b) Für eine symmetrische Matrizen A *und einen Projektor* Q *gilt:*

$$\mathcal{R}A \subset \mathcal{R}Q \quad \Leftrightarrow \quad A = QAQ \; .$$

(c) Für zwei NND Matrizen A *und* B *gilt:*

$$\mathcal{R}A + \mathcal{R}B = \mathcal{R}(A+B) \; ; \qquad \mathcal{N}A \cap \mathcal{N}B = \mathcal{N}(A+B) \; .$$

BEWEIS. (a) Die folgenden vier Aussagen sind äquivalent (Ben-Israel & Greville 1974 p.64): $\mathcal{R}A \perp \mathcal{R}B$; $\mathcal{R}A \subset \mathcal{R}^{\perp}B$; $\mathcal{R}A \subset \mathcal{N}B'$, $B'A = 0$. (b) Es gilt die folgende Gleichungskette (op.cit. p.55): $A=QA=(QA)'=AQ=QAQ$. (c) Die erste Gleichheit ist das orthogonale Komplement der zweiten; die zweite folgt mit der NND Quadratwurzel $A^{\frac{1}{2}}$ einer NND Matrix A (op.cit. p.121): $x'(A+B)x = \|A^{\frac{1}{2}}x\|^2 + \|B^{\frac{1}{2}}x\|^2 = 0 \Leftrightarrow Ax = Bx = 0$. ●

Der nächste Hilfssatz beschreibt die Geometrie der Menge der NND Matrizen: (a) Sie bilden einen abgeschlossenen konvexen Kegel, (b) dessen Inneres (*int*) im Raum Sym(n) der symmetrischen Matrizen aus den positiv-definiten Matrizen PD(n) besteht.

5 HILFSSATZ. (*Der Kegel* NND(n) *und Orthogonalität*)

(a) $\mathrm{NND}(n) = \{ B \in \mathrm{Sym}(n) \mid \forall A \in \mathrm{NND}(n): \langle A,B\rangle = 0 \}$.

(b) $\mathrm{int}\ \mathrm{NND}(n) = \mathrm{PD}(n)$.

(c) Für zwei NND Matrizen A *und* B *gilt:*

$$\mathcal{R}A \perp \mathcal{R}B \quad \Leftrightarrow \quad A \perp B .$$

BEWEIS. (a) Sei B symmetrisch und $\langle A,B\rangle \geq 0$ für alle NND A. Speziell mit $A = yy'$ folgt $0 \leq \mathrm{trace}\ Byy' = y'By$, also B NND. Seien A und B NND. Unter Benutzung der Quadratwurzel für NND Matrizen gilt: $\langle A,B\rangle = \mathrm{trace}\ A^{\frac{1}{2}}B^{\frac{1}{2}}B^{\frac{1}{2}}A^{\frac{1}{2}} = \|A^{\frac{1}{2}}B^{\frac{1}{2}}\|^2 \geq 0$. Teil (a) ist bewiesen (vergleiche Berman 1973 p.55).

(b) Die Behauptung folgt mit Theorem 13.1 in Rockafellar (1970 p.112). (c) Teil (a) impliziert: $A \perp B \Leftrightarrow B^{\frac{1}{2}}A^{\frac{1}{2}} = 0 \Leftrightarrow BA = 0$; Hilfssatz 4(a) beweist die Behauptung. ●

Mit Teil (c) wird manch eine der folgenden Fragen geometrisch gedeutet werden können. Außerdem wird er in Beweisen häufig den zentralen Schluß darstellen, so etwa im nächsten Hilfssatz.

6 HILFSSATZ. (P *als Teilprogramm von* H)

(a) Jede erwartungstreue NND quadratische Schätzfunktion $Y'AY$ *für* $q't$ *ist invariant.*

(b) Für optimale Lösungen $\hat{A}$ *von* H *und* A^* *von* P *gilt:*

$$\|\hat{A}\|^2 \leq \|A^*\|^2 .$$

BEWEIS. Für den Erwartungswert einer quadratischen Form gilt $\mathcal{E}Y'AY = \langle A, \mathcal{D}Y + \mathcal{E}Y(\mathcal{E}Y)'\rangle = \sum t_\kappa \langle A,V_\kappa\rangle + b'X'AXb$. Erwartungstreue für $q't$ impliziert $X'AX=0$, und $\langle A,XX'\rangle = 0$. Mit Hilfssatz 5(c) folgt: $\mathcal{R}A \subset \mathcal{R}^{\perp}X = \mathcal{R}M$, und mit Hilfssatz 4(b): $A = MAM$. Also ist $Y'AY$ invariant, und im Programm H wird über einer größeren Menge minimiert als in P. ●

Der folgende Satz charakterisiert erwartungstreue NND quadratische Schätzbarkeit an Hand der k zerlegenden Streuungsmatrizen und läßt sich geometrisch so deuten: solche Schätzbarkeit liegt genau dann vor, wenn man zur Erklärung des gesamten "Streuungsraumes" $\mathcal{R}\,M\sum V_{\kappa}M$ nicht auf V_{κ} verzichten kann; oder andersherum: wenn V_{κ} einen echten Beitrag zur Erklärung des Streuungsraumes $\mathcal{R}\,M\sum V_{\kappa}M$ liefert.

7 SATZ. (*Erwartungstreue nichtnegative Schätzbarkeit*)

Gegeben sei ein Varianzkomponenten-Modell $Y \sim (\sum b_{\pi}x_{\pi};\ \sum t_{\kappa}V_{\kappa})$. ***Setze*** $V := \sum V_{\kappa}$.

Eine einzelne Varianzkomponente t_{κ} ***ist erwartungstreu NND quadratisch schätzbar genau dann, wenn eine der folgenden vier Bedingungen gilt:***

$$\operatorname{rank} M\textstyle\sum_{\lambda\neq\kappa}V_{\lambda}M < \operatorname{rank} MVM \;; \qquad \mathcal{R}MV_{\kappa}M \not\subset \mathcal{R}\,M\sum_{\lambda\neq\kappa}V_{\lambda}M \;;$$

$$\mathcal{R}\,M\textstyle\sum_{\lambda\neq\kappa}V_{\lambda}M \neq \mathcal{R}MVM \;; \qquad \mathcal{N}\,M\sum_{\lambda\neq\kappa}V_{\lambda}M \not\subset \mathcal{N}\,MV_{\kappa}M \;.$$

BEWEIS. Die Äquivalenz der vier Bedingungen ist leicht einzusehen.- Nach Hilfssatz 6 und Satz I.3.10(d) ist t_{κ} genau dann erwartungstreu NND quadratisch schätzbar, wenn eine NND Matrix A existiert mit $\langle A,MV_{\lambda}M\rangle = \delta_{\kappa\lambda}$ für $\lambda=1,\dots,k$. Mit Satz 5(c) folgen $\mathcal{R}A \subset \mathcal{R}^{\perp}M\sum_{\lambda\neq\kappa}V_{\lambda}M = \mathcal{N}M\sum_{\lambda\neq\kappa}V_{\lambda}M$ und die Aussage, daß $\mathcal{R}A$ nicht orthogonal ist zu $\mathcal{R}MV_{\kappa}M$. Damit gilt einerseits:

$$\exists y\in\mathbb{R}^n: \quad M\textstyle\sum_{\lambda\neq\kappa}V_{\lambda}My = 0 \neq MV_{\kappa}My \;, \tag{*}$$

das heißt die letzte der vier angegebenen Bedingungen. Gilt andererseits (*), dann ist $A:=\|V_{\kappa}^{1/2}My\|^{-2}Myy'M$ die Matrix einer erwartungstreuen NND quadratischen Schätzfunktion $Y'AY$ für t_{κ}. ●

Dieser Satz erscheint zum ersten Mal in Pukelsheim (1977 Theorem 5.1), dort aber mit einem anderen Beweis. Er impliziert die Ergebnisse von LaMotte (1973):

8 KOROLLAR ZU SATZ 7 (L.R. LaMotte).

(a) Wenn $\sum_{\lambda\neq\kappa} V_\lambda$ *positiv-definit ist, dann ist* t_κ *nicht erwartungstreu NND quadratisch schätzbar.*

(b) Wenn V_κ *positiv-definit ist, dann ist keine der anderen Varianzkomponenten* t_λ, $\lambda \neq \kappa$, *erwartungstreu NND quadratisch schätzbar.*

Darüberhinaus beschreibt LaMotte (1973 p.728) diejenigen Linearformen q't, die erwartungstreu NND quadratisch schätzbar sind; mit unseren Bezeichnungen ergibt sich als notwendige und hinreichende Bedingung (Pukelsheim 1977 Lemma 5.1):

$$q \in \text{convex hull } \{ D_M{}' \cdot y \otimes y \mid y \in \mathbb{R}^n \} .$$

9 BEISPIEL. ***Einfache balancierte Zerlegung mit zufälligem Effekt.*** Für das Beispiel II.2.6(a) haben wir paarweise orthogonale Projektoren angegeben, die denselben Teilraum $\mathcal{B}$ aufspannen wie die zerlegenden Matrizen des Modells. Man rechnet leicht nach, daß

$$\text{convex hull } \{ D_M{}' \cdot y \otimes y \mid y \in \mathbb{R}^n \} = \left\{ \begin{pmatrix} ny'(M_{an} - I_a \otimes M_n) y \\ y' M_{an} y \end{pmatrix} \,\Big|\, y \in \mathbb{R}^n \right\} .$$

Also ist q't genau dann erwartungstreu NND quadratisch schätzbar, wenn gilt:

$$\begin{array}{lll} q_1 = q_2 = 0 & \text{falls } n = 1 & \text{und } a = 1 , \\ q_1 = q_2 \geq 0 & \text{falls } n = 1 & \text{und } a > 1 , \\ q_2 \geq q_1 = 0 & \text{falls } n > 1 & \text{und } a = 1 , \\ q_2 \geq q_1/n \geq 0 & \text{falls } n > 1 & \text{und } a > 1 . \end{array}$$

10 Nicht für alle Linearformen q't ist es angemessen, Nichtnegativität der Schätzung zu fordern. Da der Streuungsparameter t im positiven Quadranten $\mathbb{R}^k_+$ variiert, gilt

$$q \geq 0 \quad \Leftrightarrow \quad q \in \mathbb{R}^k_+ .$$

Für eine nichtnegative Schätzung kommen genau die ***nichtnegativen Linearformen*** q't, $q \in \mathbb{R}^k_+$, in Frage. Dabei sind alle nichtnegativen Linearformen genau dann erwartungstreu NND quadratisch schätzbar, wenn alle k Varianzkomponenten t_κ dies sind. Die dadurch bestimmte Teilklasse von Varianzkomponenten-Modellen kann man mit den Kriterien des Satzes 7 leicht an den k zerlegenden

Matrizen $MV_\kappa M$ erkennen. Die Varianzanalysen-Modelle zählen nicht dazu: Die gemeinsame Varianzkomponente $\sigma^2 I_n$ verhindert, daß die übrigen Komponenten erwartungstreu NND quadratisch schätzbar sind (Korollar 8.b). In dieser Hinsicht sind Modelle wie etwa das mit heteroskedastischen Varianzen denen der Varianzanalyse überlegen, und es erscheint der Überlegung wert, ob Varianzanalysen-Modelle nicht geeignet zu ersetzen sind.

11 BEISPIEL. ***Heteroskedastische Varianzen***. Satz 7 verhindert die in §2 beschriebene Pathologie von Horn & Horn: t_κ ist genau dann erwartungstreu NND quadratisch schätzbar, wenn $n-n_\kappa = \operatorname{rank} M_n \sum_{\lambda\neq\kappa} V_\lambda M_n < \operatorname{rank} M_n = n-1$, das heißt $n_\kappa > 1$. Im weiteren beschränken wir uns auf den interessanten Fall

$$k > 1 \quad \text{und für alle } \kappa = 1,\dots,k: \quad n_\kappa > 1 \ ,$$

vergleiche §I.3.12. Das Beispiel wird in §2.6 fortgesetzt.

Als letztes stellen wir eine Verbindung zu Teil III her, wie sie dort am Ende angekündigt wurde (§III.2.6).

12 KOROLLAR ZU SATZ 7.

Es seien alle k ***Varianzkomponenten*** t_κ ***erwartungstreu NND quadratisch schätzbar und*** $\hat{t} \in \mathbb{R}^k$ ***ein fester Wert.***

Es ist $\sum \hat{t}_\kappa V_\kappa$ ***NND genau dann, wenn*** $\hat{t} \in \mathbb{R}^k_+$.

In der Sprache mathematischer Programme haben wir im gegenwärtigen Kapitel untersucht, wann das Programm P aus §3 eine zulässige Lösung besitzt. Die Begriffsbildungen des nächsten Kapitels dienen dem Zweck, unter den zulässigen Lösungen die optimalen wirkungsvoll zu beschreiben.

Kapitel 2

NEGATIVITÄT ELIMINIERENDE PROJEKTOREN

In diesem Kapitel veranschaulichen wir einige Begriffe, um später das Primal- und Dualprogramm und deren Lösungen angeben und interpretieren zu können: Q-reduziertes Modell, Negativität eliminierende Projektoren, Positiv- und Negativteil einer symmetrischen Matrix.

1 Gegeben sei ein Varianzkomponenten-Modell $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ und eine erwartungstreu NND quadratisch schätzbare, nichtnegative Linearform $q't$, $q \in \mathbb{R}^k_+$. Bevor wir weiter nach "dem" besten nichtnegativen Schätzer suchen, zeigen wir, daß diese Sprechweise ihr Recht hat.

2 HILFSSATZ. (*Einzigkeit von* A^*)
Es existiert genau eine optimale Lösung A^* *des Minimierungsprogramms* P *aus* §1.3, *oder äquivalent: es gibt genau einen besten nichtnegativen Schätzer für* $q't$.

BEWEIS. Die Existenz von A^* folgt mit Rockafellar's Theorem 27.3 (1970 p.267): Die Menge $C := C(q) \cap NND(n)$ ist abgeschlossen, konvex und nach Voraussetzung nichtleer. Die Funktion $h(A) := \|A\|^2$ ist closed, proper, convex und hat keine direction of recession (op.cit. pp.52,24,265). - Da h strikt konvex ist, ist A^* eindeutig. ●

Nach Satz II.2.2 ist $\sum \lambda_\kappa M V_\kappa M$ die Matrix des besten defekten Schätzers für $q't$, wann immer $D_M' D_M \lambda = q$. Liegt q im Bild des positiven Quadranten $\mathbb{R}^k_+$ unter $D_M' D_M$, dann existiert ein nichtnegativer Koeffizientenvektor λ, die Matrix $\sum \lambda_\kappa M V_\kappa M$ ist NND und bestimmt wegen Hilfssatz 1.6 auch den besten nichtnegativen Schätzer für $q't$. Dies ist ein Spezialfall des nächsten Hilfssatzes.

3 HILFSSATZ.

Setze $V := \sum V_\kappa$; $J := \{\ \iota\in\{1,\dots,k\} \mid q_\iota > 0\ \}$; $V_J := \sum_{\iota\in J} V_\iota$. ***Sei*** Q ***der Projektor auf den Nullraum von*** $XX' + (V-V_J)$, ***definiere*** $D_Q := Q\otimes Q\cdot D$.

Wenn $q \in D_Q'D_Q(\mathbb{R}^k_+)$, ***dann ist der beste defekte Schätzer im Modell*** $QY \sim (0;\sum t_\kappa QV_\kappa Q)$ ***gleich dem besten nichtnegativen Schätzer im ursprünglichen Modell:***

$$A^* = \sum \lambda_\kappa QV_\kappa Q\ , \quad \textit{wann immer} \quad D_Q'D_Q\lambda = q\ .$$

BEWEIS. Die Behauptung folgt, wenn für jede erwartungstreue NND quadratische Schätzfunktion $Y'AY$ für $q't$ gilt: $Y'AY = (QY)'A(QY)$, oder wegen Hilfssatz 1.4(b) äquivalent: $\mathcal{R}A \subset \mathcal{R}Q$. Aus der Erwartungstreue erhält man $0 = q_\kappa = \langle A,V_\kappa\rangle$ für alle $\kappa\notin J$ und mit Hilfssatz 1.5(c): $\mathcal{R}A \subset \bigcap_{\kappa\notin J} \mathcal{R}^\perp V_\kappa = \mathcal{N}(V-V_J)$. Nach Hilfssatz 1.6 gilt: $\mathcal{R}A \subset \mathcal{R}M = \mathcal{N}XX'$, insgesamt also $\mathcal{R}A \subset \mathcal{R}Q$. ●

Dieser Hilfssatz motiviert die folgende Definition.

4 DEFINITION. Gegeben sei ein Varianzkomponenten-Modell $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ und ein Projektor Q des $\mathbb{R}^n$, dessen Bildraum kleinergleich dem von M ist: $\mathcal{R}Q \subset \mathcal{R}M$.

(a) Das von QY erzeugte lineare Modell $QY \sim (0;\ \sum t_\kappa QV_\kappa Q)$ heißt das Q-***reduzierte Modell*** des ursprünglichen Modells.

(b) Q heißt ein ***Negativität eliminierender Projektor*** (***zur Schätzung von*** $q't$, $q \in \mathbb{R}^k_+$), falls der beste defekte Schätzer für $q't$ im Q-reduzierten Modell gleich dem besten nichtnegativen Schätzer im ursprünglichen Modell ist, das heißt

$$A^* = \sum \lambda_\kappa QV_\kappa Q\ ; \quad D_Q'D_Q\lambda = q\ ; \quad D_Q := Q\otimes Q\cdot D.$$

5 Negativität eliminierende Projektoren reduzieren auf ein Teilmodell, das wohl klein genug ist, um die auftretende Negativität zu eliminieren, doch auch nicht zu klein, um Erwartungstreue und Gleichheit mit A^* zu erhalten.

Reduktionen vereinfachen eine Frage auf zweierlei Weise: sie beschränken den Blick auf einen Teil des Ganzen, oder sie lenken ihn zum Kern der Sache. Zur ersten Gruppe zählen die Reduktionen durch Linearität, Erwartungstreue und Invarianz, zur zweiten die gerade angeführte Reduktion durch Negativität eliminierende Projektoren oder die der Streuungs-Mittelwert-Korrespondenz zugrundeliegende Reduktion multilinearer Abbildungen auf lineare.

6 BEISPIEL. ***Heteroskedastische Varianzen***. Unter den Voraussetzungen des §I.11 schätzen wir zunächst eine einzelne Varianzkomponente σ_κ^2. Für den Projektor Q des Hilfssatzes 3 gilt

$$Q = \mathrm{Diag}[0{:}M_{n_\kappa}{:}0] \in \mathrm{Proj}(n) \ ,$$

denn wegen $(J_n + I_n - V_\kappa)Q = 0$ ist $\mathcal{R}Q \subset \mathcal{N}(J_n + I_n - V_\kappa)$, und zudem haben beide Räume die Dimension $n_\kappa -1$. Weitere Rechnung nach Hilfssatz 3 führt dann zum besten nichtnegativen Schätzer für σ_κ^2:

$$A_\kappa{*} = \frac{1}{n_\kappa-1}\ \mathrm{Diag}[0{:}M_{n_\kappa}{:}0] \ ;$$

$$Y'A_\kappa{*}Y = \frac{1}{n_\kappa-1}\ \sum_{\nu=1}^{n_\kappa}(Y_{\kappa\nu} - \overline{Y_{\kappa\cdot}})^2 =: s_\kappa^2 \ .$$

Wir haben gezeigt: der ***customary unbiased estimator*** (J.N.K. Rao 1973 p.13), das heißt die κ-te Stichprobenstreuung s_κ^2, ist der beste nichtnegative Schätzer für σ_κ^2: er minimiert die Norm unter allen erwartungstreuen NND quadratischen Schätzfunktionen für σ_κ^2.

Es kann aber auch der beste defekte Schätzer - C.R. Rao's MINQUE - NND sein; mit der Inversen von $D_M'D_M$ aus §I.3.12 folgt

$$q \in D_M'D_M(\mathbb{R}^k_+) \quad \Leftrightarrow \quad \min_{\kappa=1}^{k} q_\kappa/n_\kappa \geq \frac{q.}{n(n-1)} \ .$$

In diesem Fall gilt $Q = M_n$ und

$$A^* = \frac{n}{n-2}\ M_n\sum\Big(\frac{q_\kappa}{n_\kappa} - \frac{q.}{n(n-1)}\Big)V_\kappa M_n \ .$$

Sei allgemein $q \in \mathbb{R}^k_+$, J und V wie in Hilfssatz 3 und $n_J := \sum_{\iota\in J} n_\iota$; $a_J := \sum_{\iota\in J} (n_\iota/n_J)e_\iota \in \mathbb{R}^k$. Analog zur Schatzung von σ_κ^2 setzen sich die nicht verschwindenden Untermatrizen der Blockmatrix Q zu M_{n_J} zusammen, genauer gelten:

$$Q = V_J - n_J^{-1}\ V_J 1_n 1_n' V_J \in \mathrm{Proj}(n) \ ;$$

$$A^* = \frac{n_J}{n_J-2}\ Q\sum\Big(\frac{q_\iota}{n_\iota} - \frac{q.}{n_J(n_J-1)}\Big)V_\kappa Q.$$

$$q \in D_Q'D_Q(\mathbb{R}^k_+) \quad \Leftrightarrow \quad \min_{\iota\in J} q_\iota/n_\iota \geq \frac{q.}{n_J(n_J-1)} \ .$$

Unter der letztgenannten Bedingung ist Y'A*Y der beste nichtnegative Schätzer für q't. Von Fall zu Fall reicht also Hilfssatz 3 aus, um den besten nichtnegativen Schätzer zu berechnen.

Das Programm H aus §1.3 unterscheidet sich von P dadurch, daß der lineare Raum L in H durch den konvexen Kegel NND(n) in P ersetzt wird; das vorstehende Beispiel weist auf eine weitere wichtige Stelle, wo Linearität verlorengeht: Für den besten defekten Schätzer $Y'\hat{A}Y$ beziehungsweise den besten nichtnegativen Schätzer $Y'A^*Y$ einer Linearform $q't$ gilt beziehungsweise gilt nicht:

$$\hat{A} = \sum q_\kappa \hat{A}_\kappa \ ; \qquad A^* \neq \sum q_\kappa A_\kappa^* \ ;$$

dabei sind $Y'\hat{A}Y$ ($Y'A_\kappa^*Y$) der beste defekte (nichtnegative) Schätzer der einzelnen Komponenten t_κ.

Das Beispiel wird in §4.6 fortgesetzt.

Der nächste Hilfssatz wird für die weiteren Ausführungen nicht benötigt, vermittelt aber ein Gefühl dafür, in welcher Verbindung die optimalen Lösungen $\hat{A}$ von H und A^* von P stehen.

7 HILFSSATZ.

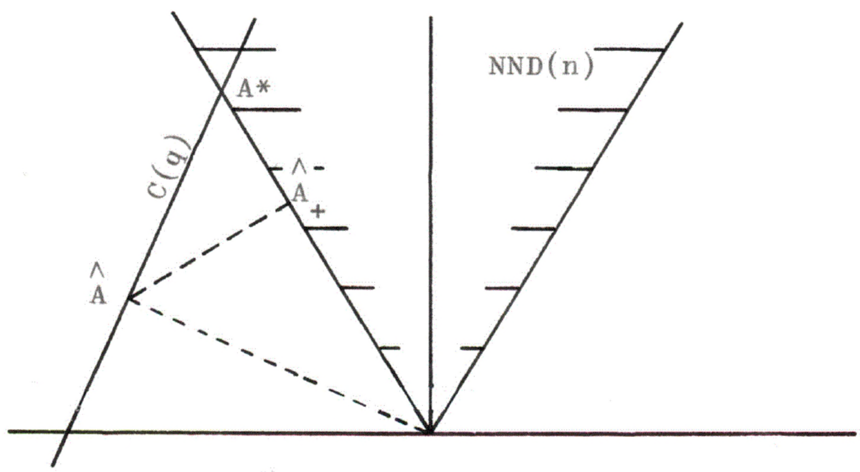
$$\|\hat{A}-A^*\|^2 = \inf \{ \|\hat{A}-A\|^2 \mid A \in C(q) \cap NND(n) \} .$$

BEWEIS. Die Bezeichnungen sind in §1.3 erklärt. Sei $A \in C(q) \cap L$. Da $\hat{A}$ die Norm minimiert, ist es orthogonal zu A^*-A; daher

$$\|\hat{A}-A\|^2 = \|\hat{A}-A^*\|^2 + 2\langle\hat{A}-A^*, A^*-A\rangle + \|A^*\|^2 + \|A\|^2 - 2\langle A^*, A\rangle =$$

$$= \|\hat{A}-A^*\|^2 + \|A\|^2 - \|A^*\|^2 .$$

Für $A \in C(q) \cap NND(n)$ folgt daraus $\|\hat{A}-A^*\|^2 \leq \|\hat{A}-A\|^2$. ●



Figur zu Hilfsatz 7

Den Zusammenhang zwischen $\hat{A}$ und A^* werden wir in Kapitel 4 (***Beste nichtnegative Schätzer als Lösungen konvexer Programme***) näher beschreiben, indem wir den Negativteil von $\hat{A}$ (vergleiche Riesz & Nagy 1955 p.277) untersuchen.

8 DEFINITION. Für eine symmetrische Matrix A bezeichne $E(\lambda)$ den Projektor auf den Eigenraum zum Eigenwert λ von A. Für eine reelle Zahl λ sei $\lambda_+ := \max\{\lambda, 0\}$ der Positivteil und $\lambda_- := \max\{-\lambda, 0\}$ der Negativteil.

Für eine symmetrische Matrix A wird definiert:

$$A_+ := \sum_{\lambda \in \mathrm{spectrum}(A)} \lambda_+ E(\lambda) \ ;$$

$$A_- := \sum_{\lambda \in \mathrm{spectrum}(A)} \lambda_- E(\lambda) \ ;$$

A_+ (A_-) heißt ***Positivteil*** (***Negativteil***) von A.

Positiv- wie Negativteil sind nichtnegative Linearkombinationen von paarweise orthogonalen Projektoren; es gilt daher offensichtlich:

$$A_+, A_- \in \mathrm{NND}(n) \ ; \quad A = A_+ - A_- \ ;$$

$$A_+ A_- = 0 \ ; \quad A_+ \perp A_- \ .$$

Beispielsweise ist ein Schätzer $Y'AY$ genau dann NND, das heißt "defektfrei", wenn $A_- = 0$. Die zweite Folgerung, die wir wiederholt benutzen werden, betrifft invariante Schätzer: $A = MAM$. Dafür gilt

$$\mathcal{R}A_- \subset \mathcal{R}A \subset \mathcal{R}M \ ,$$

so daß nach Hilfssatz 1.4(b) auch der Negativteil über M faktorisiert: $A_- = MA_-M$. Für die Projektoren N und N_M aus §I.3.11 gilt dann insbesondere

$$N \operatorname{vec} A_- = N_M \operatorname{vec} A_- \ .$$

Im nächsten Kapitel formulieren wir einen kurzen Einstieg in die Theorie mathematischer Programmierung, um danach mit solchen Methoden unser Problem allgemein zu lösen.

Kapitel 3

MINIMIERUNGSPROGRAMME

Wir führen Minimierungsprogramme ohne Spezialisierung zu linearen oder konvexen Programmen ein und zeigen die Nützlichkeit dieses Ansatzes an einem Beispiel aus der Testtheorie.

1 Ein Minimierungsprogramm wird in suggestiver Schreibweise so notiert:

$$P \qquad \begin{cases} \zeta(x) = \inf \\ \;x \in C \\ \rho(x) \succcurlyeq 0\,; \end{cases}$$

dabei unterstellen wir gerade genug, um dieser Notation Sinn zu geben: C sei eine nichtleere Teilmenge einer Menge $\mathcal{C}$; die Restriktionsfunktion ρ bilde C in einen geordneten Vektorraum $(\mathcal{Z},\succcurlyeq)$ ab; und die Zielfunktion ζ sei eine $\mathbb{R}$-wertige Funktion auf der Menge $\{\, x \in C \mid \rho(x) \succcurlyeq 0 \,\}$. Wir übernehmen die üblichen Bezeichnungen (Rockafellar 1970 p.274) eines ***optimalen Wertes***, einer ***zulässigen Lösung*** und einer ***optimalen Lösung*** von P.

In Vektorräumen identifiziert man Ordnungen $\succcurlyeq$ mit konvexen Kegeln K über die Äquivalenz

$$z \succcurlyeq 0 \quad \Leftrightarrow \quad z \in K\ .$$

Eine Teilmenge K von $\mathcal{Z}$ heißt ***Kegel***, falls $\alpha K \subset K$ für alle $\alpha \geq 0$; konvexe Kegel nennen wir auch ***Ordnungskegel***. Ist $\mathcal{Z}^d$ ein Teilraum des algebraischen Dualraumes aller Linearformen auf $\mathcal{Z}$, so induziert der Ordnungskegel K im Raum $\mathcal{Z}^d$ den dualen Ordnungskegel K^d aller ordnungstreuen Linearformen z_d:

$$K^d := \{\, z^d \in \mathcal{Z}^d \mid \forall z{\in}K\colon\, z^d(z) \geq 0 \,\},$$

und die duale Ordnung $z^d \succcurlyeq 0 \Leftrightarrow z^d \in K^d$.

Der folgende Hilfssatz reicht aus, um die Problematik mathematischer Programmierungstheorie im wesentlichen zu umreißen.

2 HILFSSATZ. (*Programmerzeugende Ungleichungen*)

Für das Minimierungsprogramm P *gilt:*

$$\inf_{x\in C;\ \rho(x)\geqslant 0} \zeta(x) \geq$$

$$\geq \inf_{x\in C} \sup_{z^d\geqslant 0} \zeta(x) - z^d(\rho(x)) \geq$$

$$\geq \sup_{z^d\geqslant 0} \inf_{x\in C} \zeta(x) - z^d(\rho(x)) .$$

BEWEIS. In der ersten Ungleichung gilt Gleichheit, falls auf der rechten Seite das Infimum über $\{ x\in C \mid \rho(x)\geqslant 0 \}$ gesucht wird. Vergrößert man diese Menge auf alle $x\in C$, dann kann das Infimum nur kleiner werden. Die zweite Ungleichung ist trivial. ●

3 In der letzten Zeile der programmerzeugenden Ungleichungen wird ein Maximierungsprogramm auf dem dualen Raum $\mathcal{Z}^d$ definiert; setzen wir

$$\psi(z^d) := \inf \{ \zeta(x)-z^d(\rho(x)) \mid x\in C \} ,$$

so lautet es in Kurzschreibweise:

$$\mathrm{D} \qquad \begin{cases} \psi(z^d) = \sup \\ \quad z^d \geqslant 0 \end{cases} .$$

Wir nennen D einen ***Dualkandidaten*** für P; sind die optimalen Werte von D und P gleich, so heißt D allgemein ***Dualprogramm*** von P, und zur Unterscheidung wird P ***Primalprogramm*** genannt. Da sich die optimalen Werte gegenseitig beschränken, folgt das

4 KOROLLAR ZU HILFSSATZ 2. (*Blitznachweis von Optimalität*)

Es seien x ***eine zulässige Lösung von*** P ***und*** z^d ***eine zulässige Lösung von*** D.

Es gilt $\zeta(x) = \psi(z^d)$ ***genau dann, wenn*** x ***optimale Lösung von*** P, z^d ***optimale Lösung von*** D ***und*** D ***ein Dualprogramm von*** P ***ist.***

5 Ohne daß Linearität, Konvexität, Stetigkeit oder ähnliches gefordert wäre, berührt Hilfssatz 2 die Hauptpunkte der Theorie mathematischer Programme.

(a) Er führt auf einfachem Weg zum Dualkandidaten D.

(b) Ist D ein Dualprogramm von P?

(c) Für lokalkonvexes $\mathcal{Z}$ und einen abgeschlossenen Ordnungskegel K gilt in der ersten Ungleichung immer Gleichheit: übrig bleibt ein Minimax-Problem.

(d) Existiert eine optimale Lösung von P, das heißt wird das Infimum angenommen?

(e) Existiert eine optimale Lösung von D, das heißt wird das Supremum angenommen?

(f) Der Dualkandidat D ist ein Optimierungsprogramm in einer anderen Variablen als P; es mag einfacher zu lösen sein.

(g) Eine besondere Rolle spielt die Funktion ψ; im Fall $\mathcal{C} = \mathcal{Z} = \mathbb{R}^n$ und $\rho = \mathrm{id}$ ist ψ bis auf das Vorzeichen die konjugierte Funktion einer konvexen Zielfunktion $\zeta = f$ (Rockafellar 1970 p.104).

Golstein (1973 pp.7-11) gab den Anstoß für den hier formulierten "theorielosen" Einstieg in Minimierungsprogramme. Ein Beispiel aus der Testtheorie soll seine Nützlichkeit belegen und anregen, in der Schätztheorie ähnlich vorzugehen.

6 BEISPIEL. ***Beste Tests als Lösungen linearer Programme.*** Wir folgen Witting (1966 pp.69-73): Gegeben sei eine von μ dominierte Klasse $\{ p_\vartheta d\mu \mid \vartheta \in H+\{\vartheta_1\} \}$ von Wahrscheinlichkeitsverteilungen über einem Stichprobenraum $(\mathcal{X},\mathcal{B})$. Gesucht ist ein bester Test zum Niveau α für H gegen $\{\vartheta_1\}$. Bezeichne Φ die Menge aller Testfunktionen, dann gilt es zu lösen

$$P \quad \begin{cases} -\mathcal{E}_{\vartheta_1}\varphi = \inf \\ \varphi \in \Phi \\ (\mathcal{E}_\vartheta\varphi)_{\theta\in H} \leq \alpha \end{cases} .$$

Sei $\mathcal{H}$ ein σ-Körper über H und $p_\vartheta(x)$ $\mathcal{H}\times\mathcal{B}$-meßbar, dann ist $\rho(\varphi) := (\alpha - \mathcal{E}_\vartheta\varphi)_{\vartheta\in H}$ eine $\mathcal{H}$-meßbare Funktion auf H: Als $\mathcal{Z}$ wählen wir den linearen Raum aller beschränkten $\mathcal{H}$-meßbaren Funktionen auf H, als dualen Raum $\mathcal{Z}^d$ den linearen Raum aller signierten Maße auf $(H, \mathcal{H})$. Dann ist jedes nichtnegative endliche Maß λ auf $(H, \mathcal{H})$ eine zulässige Lösung des Dualkandidaten. Mit den Abkürzungen

$$d(x) := p_{\vartheta_1}(x) - \int_H p_\vartheta(x)\,d\lambda$$

und d_+ (d_-) für den Positivteil (Negativteil) von d gilt:

$$\zeta(\varphi) - \psi(\lambda) = -\mathcal{E}_{\vartheta_1}\varphi - \inf_{\varphi\in\Phi}[-\mathcal{E}_{\vartheta_1}\varphi - \int_H(\alpha - \mathcal{E}_\vartheta\varphi)\,d\lambda] =$$
$$= -\int_{\mathcal{X}}\varphi(x)[d(x)+\int_H p_\vartheta(x)\,d\lambda]\,d\mu + \int_H\alpha\,d\lambda + \sup_{\varphi\in\Phi}\int_{\mathcal{X}}\varphi(x)d(x)\,d\mu =$$
$$= -\int_{\mathcal{X}}\varphi(x)[d_+(x)-d_-(x)]\,d\mu - \int_H \mathcal{E}_\vartheta\varphi\,d\lambda + \int_H\alpha\,d\lambda + \int_{\mathcal{X}} d_+(x)\,d\mu =$$
$$= \int_H[\alpha - \mathcal{E}_\vartheta\varphi]\,d\lambda + \int_{d>0}[1-\varphi(x)]d_+(x)\,d\mu + \int_{d<0}\varphi(x)d_-(x)\,d\mu \;.$$

Für zulässige Tests φ und zulässige Maße λ gilt daher $\zeta(\varphi) = \psi(\lambda)$ genau dann, wenn bis auf eine $\mathcal{H}$-meßbare λ-Nullmenge

$$\mathcal{E}_\vartheta\varphi = \alpha$$

und bis auf eine $\mathcal{B}$-meßbare μ-Nullmenge

$$\varphi(x) = \begin{cases} 1 & \text{falls } d(x)>0 \\ 0 & \text{falls } d(x)<0 \end{cases} \;.$$

Weitere statistische Anwendungen mathematischer Programme beschreibt Krafft (1970). Bei linearen Modellen wurden Programmierungsmethoden bisher nur genutzt, um Algorithmen zur Berechnung sinnvoller Schätz***werte*** zu bestimmen, vergleiche Thompson (1962), Judge & Takayama (1966), Hudson (1969), Liew (1976). Statt dessen folgen wir dem vorstehenden Beispiel und behandeln beste Schätz***funktionen*** als optimale Lösungen mathematischer Programme.

Kapitel 4

BESTE NICHTNEGATIVE SCHÄTZER ALS LÖSUNGEN KONVEXER PROGRAMME

Mit Methoden der mathematischen Programmierung geben wir für den besten nichtnegativen Schätzer zwei Darstellungen an: Zum einen wird er berechnet aus dem besten defekten Schätzer, indem dessen Negativteil passend korrigiert wird; zum andern ergibt er sich als bester defekter Schätzer in einem reduzierten Modell, wobei der reduzierende Negativität eliminierende Projektor über das Dualprogramm bestimmt wird.

1 Gegeben sei ein Varianzkomponenten-Modell $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ und eine erwartungstreu NND quadratisch schätzbare Linearform $q't$, $q \in \mathbb{R}^k_+$, des Streuungsparameters t. Der beste nichtnegative Schätzer $Y'A^*Y$ für $q't$ (§1.3) wird bestimmt durch die optimale Lösung A^* des Programms

$$P \qquad \begin{cases} \|A\|^2 = \inf \\ A \in C(q) \\ A \succcurlyeq 0 \end{cases} .$$

Es bezeichnet $C(q) \subset \mathrm{Sym}(n)$ diejenigen symmetrischen $n{\times}n$-Matrizen A, für die $Y'AY$ erwartungstreu für $q't$ ist. $A \succcurlyeq 0$ steht für $A \in \mathrm{NND}(n) \subset \mathrm{Sym}(n)$.

Um einen Dualkandidaten D für P zu konstruieren, benutzen AA wir die Matrix $\hat{A}$ des besten defekten Schätzers $Y'\hat{A}Y$ für $q't$ (MN-UB-IQE (Satz II.2.2); MINQUE) und den Projektor N_M aus §I.3.11:

2 HILFSSATZ. (*Dualkandidat für* P)

Setze $s(B) := \|N_M \operatorname{vec} B\|^2$ ***und*** $\psi(B) := \|\hat{A}\|^2 - \langle\hat{A},B\rangle - s(\tfrac{1}{2} B)$.

Dann ist $\quad D \quad \begin{cases} \psi(B) = \sup \\ B \succcurlyeq 0 \end{cases}$ ***ein Dualkandidat für*** P.

BEWEIS. Da Sym(n) mit dem euklidischen Matrixprodukt ein Hilbertraum ist, wählen wir als dualen Raum wiederum Sym(n).

Hilfssatz 1.5(a) besagt, daß der Kegel NND(n) selbstdual ist: $NND(n)^d = NND(n)$. Wegen $A \succcurlyeq 0$ und Hilfssatz 1.6(a) liegen die zulässigen Lösungen für P nicht nur in C(q), sondern - mit L aus §1. 3 - sogar in $C := C(q) \cap L$ (UB-IQE für q't). Den affinen Raum C beschreiben wir mit $\hat{A}$ (spezielle UB-IQE für q't) und dem Projektor N_M (alle UB-IQE für 0): Das Programm P ist gleichwertig zu

$$\begin{cases} \|A\|^2 = \inf \\ A \in \{ A \in Sym(n) \mid \exists Z \in Sym(n): \text{vec } A = \text{vec } \hat{A} + N_M \text{vec } Z \} \\ A \succcurlyeq 0 \end{cases} .$$

Für diese Formulierung konstruieren wir D gemäß §3.3. Als Minimum Norm-Lösung steht vec $\hat{A}$ orthogonal auf $\mathscr{R}N$ und $\mathscr{R}N_M$, und es gilt:

$$\|A\|^2 - \langle A,B\rangle = \|\hat{A}\|^2 + s(Z) - \langle\hat{A},B\rangle - \text{vec}'Z \cdot N_M \cdot \text{vec}\, B =$$
$$= \|\hat{A}\|^2 - \langle\hat{A},B\rangle - s(\tfrac{1}{2}B) + s(Z-\tfrac{1}{2}B) .$$

Dies wird für $Z = \frac{1}{2}B$ minimiert, und es folgt die Behauptung. ●

Mit dem in Definition 2.8 eingeführten Negativteil einer symmetrischen Matrix berechnen wir jetzt die optimale Lösung des Dualkandidaten D: ohne den optimalen Wert zu verletzen, finden wir in NND(n) eine Teilmenge $\mathcal{M}$, über der ψ direkt abgeschätzt werden kann.

SATZ. (*Optimale Lösung des Dualkandidaten*)

Gegeben sei ein Varianzkomponenten-Modell $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ *und eine erwartungstreue NND quadratisch schätzbare Linearform* q't, $q \in \mathbb{R}^k_+$, *von* t. *Für die vorstehend definierten Programme* P *und* D *gilt:*

(a) Sei $\hat{A}$ *NND. Dann ist* 0 *optimale Lösung von* D.

(b) Sei $\hat{A}$ *nicht NND, das heisst* $\hat{A}_- \neq 0$. *Dann ist* $\|N \cdot \text{vec } \hat{A}_-\|^2 > 0$ *und* $B^* := 2 s^* \hat{A}_-$ *optimale Lösung von* D, *wobei* $s^* := \|\hat{A}_-\|^2 / \|N \cdot \text{vec } \hat{A}_-\|^2$. *Ausserdem ist die durch*

$$\text{vec } A^* := \text{vec } \hat{A} + s^* N \cdot \text{vec } \hat{A}_-$$

definierte symmetrische Matrix A^* *NND.*

BEWEIS. (a) Dieser Teil folgt sofort aus Hilfssatz 1.5(a): $\psi(B) \le \|\hat{A}_-\|^2 - 0 - 0 = \psi(0)$. (b) Für Teil (b) leiten wir (i) eine notwendige Bedingung für optimale B* her, maximieren (ii) entlang von Halbstrahlen, konstruieren (iii) "ohne Verlust" eine nicht-leere Menge $\mathcal{M}$, über der wir (iv) ψ direkt abschätzen können.

(i) *Für optimale* B^* *gilt*:

$$\forall B \in NND(n): \langle B, \hat{A} \rangle + \tfrac{1}{2}\, vec' B \cdot N_M \cdot vec\, B^* \geq 0 \ .$$

Denn: Wir verifizieren die Voraussetzungen zu Theorem 27.4 in Rockafellar (1970 pp.270f). Es ist $-\psi$ proper, convex und dom$(-\psi)$ = Sym(n) (op.cit. pp.23f). Satz 1.5(b) ergibt:

$$ri\ dom(-\psi) \cap ri\ NND(n) = PD(n) \neq \emptyset \ .$$

Besagtes Theorem liefert als notwendige und hinreichende Bedingung für die Optimalität von B^*:

$$-grad_{B^*}(-\psi) \text{ ist normal an } NND(n) \text{ in } B^* \ .$$

Das bedeutet (op.cit. p.15), daß für alle $B \in NND(n)$ gilt:

$$\langle B-B^*, grad_{B^*}\psi \rangle \leq 0 \ ,$$

$$\langle vec(B-B^*), vec\, \hat{A} + \tfrac{1}{2} N_M\, vec\, B^* \rangle \geq 0,$$

$$\langle B, \hat{A} \rangle + \tfrac{1}{2} vec' B \cdot N_M \cdot vec\, B^* \geq \langle B^*, \hat{A} \rangle + \tfrac{1}{2} s(B^*) \ .$$

Für optimale B^* ist also das auf der linken Seite in der Variablen B definierte Funktional auf dem Kegel NND(n) nach unten beschränkt; die einzig mögliche endliche Schranke ist aber 0.

(ii) *Für alle* $B \in NND(n)$ *mit* $vec'\hat{A}_- \cdot N \cdot vec\, B \neq 0$ *gilt* $s(B) > 0$ *und*

$$\psi(B) \leq \|\hat{A}\|^2 + \frac{\langle \hat{A}, B \rangle^2}{s(B)} =: f(B) \ .$$

Denn: Die Cauchy-Schwarz'sche Ungleichung liefert:

$$0 < \|\hat{A}_-\|^{-2} \langle vec\, \hat{A}_-, N\, vec\, B \rangle^2 \leq s(B) \ .$$

Im Falle $\langle \hat{A}, B \rangle = 0$ gilt $\psi(B) \leq \psi(0) \leq f(B)$; im Falle $\langle \hat{A}, B \rangle < 0$ ist f(B) das Maximum entlang des Halbstrahls $\{ \lambda B \mid \lambda > 0 \}$.

(iii) *Es gilt*

$$\mathcal{M} := \{ B \in NND(n) \mid \langle \hat{A}, B \rangle < 0 \neq vec'\hat{A}_- \cdot N \cdot vec\, B \} \neq \emptyset$$

und

$$\sup_{B \in NND(n)} \psi(B) = \sup_{B \in \mathcal{M}} f(B) \ .$$

Denn: $\hat{A}_-$ ist NND und $\langle A, \hat{A}_- \rangle = -\|\hat{A}_-\|^2 < 0$. Die Eigenschaft $s(\hat{A}_-) > 0$ folgt aus Hilfssatz 3.2: $s(\hat{A}_-) = 0$ impliziert wegen $\psi(\lambda \hat{A}_-) =$

$\|\hat{A}\|^2 + \lambda\|\hat{A}_-\|^2$, daß sup $\psi = +\infty$; das kann aber nicht sein, da die vorausgesetzte zulässige Lösung von P ψ nach oben beschränkt.

Untersucht man in (i) speziell $B = \hat{A}_-$, dann erhält man

$$\mathrm{vec}'\hat{A}_-\cdot N\cdot \mathrm{vec}\, B^* \geq 2\|\hat{A}_-\|^2 > 0 \ .$$

Der Fall $\langle\hat{A}B^*\rangle \geq 0$ kann für optimale B^* nicht eintreten, da sonst $\psi(B^*) \leq \psi(0) < \psi(2s^*\hat{A}_-) = f(\hat{A}_-)$. Somit liegen optimale B^* notwendig in $\mathcal{M}$, und es gilt die behauptete Gleichheit der Suprema.

(iv) Für $B\in\mathcal{M}$ gilt mit Hilfssatz 1.5(a):

$$0 < -\langle\hat{A},B\rangle = \langle\hat{A}_-,B\rangle - \langle\hat{A}_+,B\rangle \leq \langle\hat{A}_-,B\rangle \leq \|\hat{A}_-\|\cdot\|B\|$$

Benutzt man außerdem die Abschätzung aus (11), um $s^{-1}(B)$ nach oben abzuschätzen, so folgt für $B\in\mathcal{M}$:

$$f(B) \leq \|\hat{A}\|^2 + \langle \mathrm{vec}\,\hat{A}_-, N\,\mathrm{vec}\, B\rangle^{-2}\|B\|^2\|\hat{A}_-\|^4 \ .$$

Die rechte Seite wird maximiert, indem bei gegebenem $y = N\,\mathrm{vec}\,\hat{A}_-$ das innere Produkt $x'y$ minimiert wird für $\|x\|=1$ und $\langle x,y\rangle > 0$. In das innere Produkt $x'y$ geht x aber nur über seine Projektion auf y ein, so daß äquivalent $\lambda y'y$ minimiert wird mit $\|\lambda y\| = 1$: das Minimum ist $\|y\| = \|N\,\mathrm{vec}\,\hat{A}_-\|$, und für $B\in\mathcal{M}$:

$$f(B) \leq \|\hat{A}\|^2 + \|\hat{A}_-\|^4/s(\hat{A}_-) = \|\hat{A}\|^2 + s^*\|\hat{A}_-\|^2 = \psi(2s^*\hat{A}_-)\ .$$

Dies bedeutet, daß B^* optimale Lösung von D ist. Die Aussage über A^* erhalten wir, wenn wir zurück nach (i) gehen:

$$\forall B\in NND(n)\colon \langle B,A^*\rangle \geq 0$$

sichert wegen der Selbstdualität von NND(n) (1.5.a), daß A^* NND ist. ●

Jetzt fehlt nur noch ein kleiner Schritt, um auch das Primalprogramm zu lösen.

4 SATZ. (*Bester nichtnegativer Schätzer für* $q't$)

Gegeben bei ein Varianzkomponenten-Model $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ *und eine erwartungstreu NND quadratisch schätzbare Linearform* $q't$, $q\in\mathbb{R}^k_+$, *von* t. *Der nach Satz* I.2.2 *gegebene beste defekte Schätzer* (MN-UB-IQE; MINQUE) $Y'\hat{A}Y$ *für* $q't$ *sei nicht NND:* $\hat{A}_- \neq 0$.

Dann bestimmt die durch

$$\mathrm{vec}\,A^* := \mathrm{vec}\,\hat{A} + s^*N\,\mathrm{vec}\,\hat{A}_- \ ; \quad s^* := \|\hat{A}_-\|^2/\|N\,\mathrm{vec}\,\hat{A}_-\|^2 \ ;$$

definierte symmetrische Matrix A^* *den besten nichtnegativen Schätzer* $Y'A^*Y$ *für* $q't$, *das heißt die erwartungstreue NND quadratische Schätzfunktion von kleinster Norm* (MN-UB-NND·QE). *Ihre Norm ist gleich* $\|\hat{A}\|^2 + s^*\|\hat{A}_-\|^2$.

BEWEIS. A^* ist NND nach Satz 3(b). Als Summe eines erwartungstreuen Schätzers für $q't$ und eines erwartungstreuen Schätzers für Null ist $Y'A^*Y$ erwartungstreu für $q't$. Also ist A^* eine zulässige Lösung des Programms P aus §1. Die Zielfunktion hat an dieser Stelle den Wert

$$\zeta(A^*) = \|\hat{A}\|^2 + (s^*)^2\, s(\hat{A}_-) = \|\hat{A}\|^2 + s^*\|\hat{A}_-\|^2 ,$$

ist also gleich $\psi(B^*)$: Der Blitznachweis von Optimalität 3.4 ergibt die Behauptung. ●

Die angegebene Darstellung für den besten nichtnegativen Schätzer $Y'A^*Y$ läßt sich so deuten, daß der Negativteil $\hat{A}_-$ des besten defekten Schätzers $Y'\hat{A}Y$ - also gerade der defekte Teil - geeignet korrigiert wird. Es ist bemerkenswert, daß die Berechnung von A^*, wie sie hier vorgeführt wurde, maßgeblich auf der Lösung des Dualprogramms D beruht; es würde gewiß die Einsicht fördern, wenn die Nichtnegativ-Definitheit von A^* "direkt" und ohne Rückgriff auf D gezeigt werden könnte.

Im folgenden Korollar wiederholen wir die Sätze 3 und 4 in der Sprache der Programmierungstheorie.

5 KOROLLAR ZU DEN SÄTZEN 3 UND 4. (*Die Programme* P *und* D)

(a) D *ist ein Dualprogramm von* P.

(b) Die folgenden drei Aussagen sind äquivalent:

$\hat{A}_- = 0$.

$\hat{A}$ *ist optimale Lösung von* P .

0 *ist optimale Lösung von* D .

(c) Die folgenden vier Aussagen sind äquivalent:

$\hat{A}_- \neq 0$.

$N \operatorname{vec} \hat{A}_- \neq 0$.

$N \operatorname{vec} \hat{A}_- \neq 0$ *und* A^* *ist optimale Lösung von* P .

$N \operatorname{vec} \hat{A}_- \neq 0$ *und* $2s^*\hat{A}_-$ *ist optimale Lösung von* D .

6 BEISPIEL. ***Heteroskedastische Varianzen.*** In §1.2 ist der beste defekte Schätzer für eine einzelne Komponente σ_κ^2 berechnet worden; dabei gehören zu $\widehat{A}_\kappa$ die folgenden vier Eigenwerte λ_i mit den Eigenraum-Projektoren $E(\lambda_i)$

$$\begin{aligned}
&\lambda_1 = -[(n-1)(n-2)]^{-1} < 0; \qquad \lambda_2 = 0 > \lambda_1 ;\\
&\lambda_3 = (n^2 - n - n_\kappa n)[n_\kappa(n-1)(n-2)]^{-1} > \lambda_2 ;\\
&\lambda_4 = (n^2 - n - n_\kappa)[n_\kappa(n-1)(n-2)]^{-1} > \lambda_3 ;\\
&E(\lambda_1) = (I_n-V_\kappa) - [n-n_\kappa]^{-1}(I_n-V_\kappa)J_n(I_n-V_\kappa) ;\\
&E(\lambda_2) = \overline{J_n} ;\\
&E(\lambda_3) = n[n_\kappa(n-n_\kappa)]^{-1} M_nV_\kappa J_nV_\kappa M_n ;\\
&E(\lambda_4) = V_\kappa - n_\kappa^{-1}V_\kappa J_nV_\kappa = \mathrm{Diag}[0:M_{n_\kappa}:0] ;
\end{aligned}$$

mit beziehungsweisen Rängen $n-n_\kappa-1$; 1; 1; $n_\kappa-1$. Aus dieser Spektralzerlegung folgt $\widehat{A}_{\kappa-} = -\lambda_1 E(\lambda_1)$. Man berechne zunächst

$$\langle E(\lambda_1),MV_iM\rangle = \langle E(\lambda_1),V_i\rangle = n_i(1-\delta_{i\kappa})(n-n_\kappa-1)[n-n_\kappa]^{-1}$$

und dann mit der in §I.3.12 angegebenen Inversen von $D_M'D_M$:

$$\begin{aligned}
s(\widehat{A}_{\kappa-}) &= \|\widehat{A}_{\kappa-}\|^2 - \lambda_1^2 \textstyle\sum_{i,j=1} \langle E(\lambda_1),V_i\rangle\ (D_M'D_M)^{-1}{}_{ij}\ \langle E(\lambda_1),V_j\rangle\\
&= (n-n_\kappa -1)n_\kappa(n_\kappa-1)[(n-n_\kappa)(n-1)^3(n-2)^3]^{-1} ;
\end{aligned}$$

$$s^* = (n-1)(n-2)(n-n_\kappa)[n_\kappa(n_\kappa-1)]^{-1} .$$

Für $y'A^*y = (y\otimes y)'\mathrm{vec}\, A^*$ lautet der gemäß Satz 4 zweite Term:

$$\begin{aligned}
&s^*\ (y\otimes y)'N\ \mathrm{vec}\ \widehat{A}_{\kappa-} =\\
&= \frac{n-n_\kappa}{n_\kappa(n_\kappa-1)}y'E(\lambda_1)y - \frac{(n-n_\kappa-1)(n^2-2n+n)}{n_\kappa(n_\kappa-1)(n-1)(n-2)}y'M_nY + \frac{(n-n_\kappa-1)n}{n_\kappa(n_\kappa-1)(n-2)}y'M_nV_\kappa M_ny\\
&= -\lambda_1\ y'E(\lambda_1)y - \lambda_3\ y'E(\lambda_3)y + \left(\frac{1}{n_\kappa-1} - \lambda_4\right)\ y'E(\lambda_4)y\ .
\end{aligned}$$

Satz 4 führt wiederum zum besten nichtnegativen Schätzer $Y'A_\kappa{}^*Y$ für σ_κ^2 wie er schon in §2.6 angegeben wurde. In §10 werden wir $A_\kappa{}^*$ ein drittes Mal bestimmen, dann mit Negativität eliminierenden Projektoren.

Wir eröffnen uns jetzt weitere Wege, um A^* zu berechnen, indem wir statt des "theorielosen" Einstiegs in mathematische Programme, wie wir ihn in Kapitel 3 beschrieben haben, W. Fenchel's ausgefeilte Dualitätsaussagen für konvexe Programme verwenden. Dabei verzichten wir zunächst auf die Kenntnis der Sätze 3 und 4.

7 SATZ. (*Fenchel'sche Dualität für* P *und* D)

Gegeben bei ein Varianzkomponenten-Model $Y \sim (\sum b_\pi x_\pi; \sum t_\kappa V_\kappa)$ *und eine erwartungstreu NND quadratisch schätzbare Linearform* $q't$, $q \in \mathbb{R}^k_+$, *von* t. *Für die in* §1 *und* §2 *definierten Programme* P *und* D *gilt:*

(a) D *ist ein Dualprogramm von* P.

(b) $A^* \in \mathrm{Sym}(n)$ *ist optimale Lösung von* P *und* $B^* \in \mathrm{Sym}(n)$ *ist optimale Lösung von* D *genau dann, wenn gelten*

$$(1)\ A^* \in C(q)\ ; \quad (2)\ A^* \geqslant 0\ ; \quad (5)\ B^* \geqslant 0\ ;$$
$$(3)\ \langle A^*,B^*\rangle = 0\ ; \quad (4)\ \exists \lambda \in \mathbb{R}^k: \quad 2A^*-MB^*M = \sum \lambda_\kappa M V_\kappa M\ .$$

BEWEIS. Die Behauptung folgt als Anwendung des Theorem 31.4 in Rockafellar (1970 p.335).

Wir setzen $C = C(q) \cap L$ wie im Beweis zu Hilfssatz 2. Es sei $\delta(A|C)$ die konvexe Indikatorfunktion für C (op.cit. p.28), sie hat den Wert 0, falls $A \in C$, und den Wert $+\infty$, falls $A \notin C$. Auf Sym(n) definieren wir die Funktion $f(A) := \|A\|^2 + \delta(A|C)$. Mit $K :=$ NND(n) ist das Programm P äquivalent mit der Bestimmung von $\inf\{ f(A) | A \in K \}$. K ist ein abgeschlossener konvexer Kegel mit $K^* = K^d = K$ (§1.5.a); da mindestens eine zulässige Lösung für P existiert, ist f proper, convex und closed (op.cit. pp.24,52). Es gilt $-f^*(B) = \inf_{A \in C} f(A) - \langle A,B\rangle = \psi(B)$ (op.cit. p.104); also ist der Dualkandidat D äquivalent mit der Bestimmung von $-\inf\{ f^*(B) | B \in K^*\}$. Schließlich besagt Hilfssatz 1.5(b):

$$\mathrm{ri\ dom}(f^*) \cap \mathrm{ri}\ K^* = PD(n) \neq \emptyset.$$

Das erwähnte Theorem liefert jetzt Behauptung (a) und beschreibt die Optimalität von A* und B* äquivalent durch

$$(6)\ B^* \in \partial f(A^*); \quad (2)\ A^* \in K; \quad (5)\ B^* \in K^*; \quad (3)\ \langle A^*,B^*\rangle = 0\ .$$

Eigenschaft (6) bedeutet, daß B* Subgradient von f in A* ist, das heißt (op.cit. p.214):

$$(6)\ \forall Z \in \mathrm{Sym}(n):\ \|Z\|^2 + \delta(Z|C) \geq \|A^*\|^2 + \delta(A^*|C) + \langle B^*, Z-A^*\rangle.$$

In dieser Formulierung ist (6) äquivalent dazu, daß (1) gilt und

$$(6a)\ \forall Z \in C:\ \|Z - A^* + A^*\|^2 \geq \|A^*\|^2 + \langle B^*, Z-A^*\rangle.$$

Substituieren wir $W := Z-A^*$, dann variiert W über den zu C gehörenden linearen Raum C-C, oder gleichwertig: vec W $\in \mathcal{R}N_M$:

$$\text{(6a)}\ \forall \text{vec}\ W \in \mathcal{R}N_M: \quad \|W\|^2 \geq \langle B^*-2A^*, W\rangle \ .$$

Untersucht man für positive und negative Vielfache αW die Grenzwerte für $\alpha\to 0$, so erhält man vec$(B^*-2A^*) \perp \mathcal{R}N_M$, oder äquivalent vec$(2A^*-MB^*M) \subset \mathcal{R}^\perp N = \mathcal{R}D_M$, das heißt (4). ●

Durchschlagende Kraft erhält dieser Satz, wenn wir die optimale Lösung B^* aus Satz 3 einsetzen:

8 KOROLLAR DER SÄTZE 3 UND 7 (*Konstruktionsanleitung für* A^*)

Sei $\hat{A}$ *nicht* NND. *Es ist* $Y'A^*Y$ *bester nichtnegativer Schätzer für* $q't$ *genau dann, wenn gelten*

$$(1)\ A^* \in C(q)\ ; \quad (2)\ A^* \succcurlyeq 0\ ; \quad (3)\ \langle A^*, \hat{A}_-\rangle = 0\ ;$$
$$(4)\ \exists\lambda\in\mathbb{R}^k: \quad A^* - s^*\hat{A}_- = \textstyle\sum \lambda_\kappa\, MV_\kappa M\ .$$

Damit ist die Infimum-Bedingung des Programms P aus §1 durch die A^* enthüllenden Eigenschaften (3) und (4) ersetzt worden. Neben Satz 4 haben wir jetzt die Möglichkeit, A^* mittels Negativität eliminierender Projektoren zu berechnen.

9 SATZ. (*Negativität eliminierende Projektoren*)

Gegeben bei ein Varianzkomponenten-Model $Y \sim (\sum b_\pi x_\pi;\ \sum t_\kappa V_\kappa)$ *und eine erwartungstreu NND quadratisch schätzbare Linearform* $q't$, $q \in \mathbb{R}^k_+$, *von* t. *Der nach Satz* I.2.2 *gegebene beste defekte Schätzer* (MN-UB-IQE; MINQUE) $Y'\hat{A}Y$ *für* $q't$ *sei nicht NND:* $\hat{A}_- \neq 0$. *Es sei* Q^* *der Projektor auf den Nullraum von* $XX' + \hat{A}_-$.

(a) Q^* *ist ein Negativität eliminierender Projektor, das heißt: Für den besten nichtnegativen Schätzer* (MN-UB-NND·QE) $Y'A^*Y$ *für* $q't$ *gilt:*

$$A^* = \textstyle\sum \lambda_\kappa\, Q^*V_\kappa Q^* \ \textit{mit beliebiger Lösung}\ \lambda\ \textit{von}\ D_{Q^*}'D_{Q^*}\,\lambda = q\ .$$

(b) Sei Q_* *der Projektor auf den Bildraum von* A^*. *Dann gilt* $Q^* \succcurlyeq Q_*$ *und jeder Projektor* Q *im Intervall* $[Q_*, Q^*]$ *ist ein Negativität eliminierender Projektor zur Schätzung von* $q't$.

BEWEIS. Für Projektoren sind die Ordnungen $\mathcal{R}Q_* \subset \mathcal{R}Q^*$ und $Q^* - Q_* \succcurlyeq 0$ gleich (Ben-Israel & Greville 1974 p.71). Wir benutzen die Hilfssätze 1.4;1.5;1.6 und die Eigenschaften (1)-(4) aus Korollar 8. (1) und (2) implizieren $\mathcal{R}A^* \subset \mathcal{R}MA^*M \subset \mathcal{R}M = \mathcal{N}XX'$, aus (3) folgt $\mathcal{R}A^* \subset \mathcal{R}^{\perp}\hat{A}_- = \mathcal{N}\hat{A}_-$, zusammen gilt $\mathcal{R}Q_* = \mathcal{R}A^* \subset \mathcal{N}\ XX'+\hat{A}_- = \mathcal{R}Q^*$ und $Q^* \geqslant Q_*$. Sei $Q \in [Q_*,Q^*]$ ein Projektor. Einerseits bewirkt $Q \geqslant Q_*$, daß $\mathcal{R}A^* \subset \mathcal{R}Q$ und $A^* = QA^*Q$. Andererseits impliziert $Q^* \geqslant Q$, daß $\mathcal{R}Q \subset \mathcal{R}Q^* = \mathcal{N}\hat{A}_- \cap \mathcal{R}M$ und daher zum ersten $\mathcal{R}Q \subset \mathcal{R}M$ und $QM = Q$ und zum zweiten $\mathcal{R}Q \subset \mathcal{N}\hat{A}_-$ und $\hat{A}_-Q = 0$. Dies verwenden wir nun in Eigenschaft (4):

$$A^* = s^*\hat{A}_- + \sum \lambda_\kappa\ MV_\kappa M = QA^*Q = 0 + \sum \lambda_\kappa\ QV_\kappa Q.$$

Erwartungstreue (1) bestimmt λ als beliebige Lösung von $D_Q'D_Q\lambda = q$: Q ist ein Negativität eliminierender Projektor (§2.4.b). ●

Der Negativität eliminierende Projektor Q^* wird durch den Negativteil des besten defekten Schätzers $Y'AY$ festgelegt: der beste defekte Schätzer mag zwar nicht defektfrei sein, er enthält aber doch alle wesentliche Information zur besten nichtnegativen Schätzung.

10 BEISPIEL. Heteroskedastische Varianzen. Wie in §6 angekündigt, berechnen wir A_κ^* jetzt mit dem Negativität eliminierenden Projektor Q^* aus Satz 9. Man findet leicht, auch ohne die Spektralzerlegung von $\hat{A}_\kappa$ in §6, daß

$$Q^* = E(\lambda_3) + E(\lambda_4) = E(\lambda_3) + Q_* \underset{\neq}{\geqslant} Q_* .$$

Die Q^*-reduzierten zerlegenden Matrizen $Q^*V_iQ^*$ sind:

$$Q^*V_iQ^* = n_i[n-n_\kappa]^{-2}\ (V_\kappa - \tfrac{n_\kappa}{n}I_n)J_n(V_\kappa - \tfrac{n_\kappa}{n}I_n) \qquad \text{für } i \neq \kappa$$
$$Q^*V_\kappa Q^* = M_nV_\kappa M_n\ .$$

Mit dem Vektor $z := n^{-1}(n-n_\kappa)\ e_\kappa + n_\kappa[n-n_\kappa]^{-1} \sum_{i\neq\kappa} n^{-1}n_i\ e_i$ gilt:

$$D_{Q^*}'D_{Q^*} = zz' + (n_\kappa-1)\ e_\kappa e_\kappa'\ ;$$
$$\lambda = n[n_\kappa(n_\kappa-1)]^{-1}\ e_\kappa - (n-n_\kappa)[n_\kappa(n_\kappa-1)]^{-1}\ 1_k\ .$$

Schließlich erhält man $\sum \lambda_i Q^*V_iQ^* = [n_\kappa-1]^{-1}\ \mathrm{Diag}[0:M_{n_\kappa}:0] = A_\kappa^*$.

11 Zum Schluß kehren wir zu der allgemeinen Diskussion von §1.1 zurück, welcher Stellenwert dem Problem negativer Varianzschätzung zusteht.

Zwei Annahmen stehen am Anfang der Theorie: die *Modellannahme* beschränkt die Klasse der Verteilungen von Y auf solche, die den

Momentenbedingungen $\mathcal{E}Y = \sum b_\pi x_\pi$ und $\mathcal{D}Y = \sum t_\kappa V_\kappa$ genügen; die ***Entscheidungsannahme*** beschränkt die Klasse der Entscheidungsfunktionen auf Schätzfunktionen von - sagen wir - σ_κ^2 und darunter auf die quadratischen, NND und erwartungstreuen.

Von einem "Problem" zu sprechen ist müßig, wenn die beiden Annahmen miteinander verträglich sind: Im vorstehenden Beispiel heteroskedastischer Varianzen existieren erwartungstreue NND quadratische Schätzfunktionen, falls nur $n_\kappa > 1$; der beste defekte Schätzer (MINQUE) hat zwar durchaus interessante Eigenschaften, aber seine mangelnde Nichtnegativ-Definitheit disqualifiziert ihn trotzdem.

Anders im Fall der Varianzanalyse, wo $\mathcal{D}Y = \sigma^2 I_n + \sum \sigma_\kappa^2 V_\kappa$ und die gemeinsame Var1anzkomponente σ^2 verhindert, daß die anderen Komponenten σ_κ^2 erwartungstreu NND schätzbar sind. Den zwei Annahmen entsprechend gibt es eine doppelte Möglichkeit zu reagieren.

Zum ersten mag man die Modellannahme in Frage stellen: Nelder (1954) interpretiert negative Varianzkomponenten, indem er einen zufälligen Versuchsplan (randomisierte Blöcke) unterstellt; McHugh & Mielke (1968) gelingt eine Deutung dadurch, daß sie die Grundgesamtheit als endlich ansehen. Mit Satz 1.7 (***Erwartungstreue nichtnegative Schätzbarkeit***) haben wir eine notwendige und hinreichende Bedingung bewiesen, wann obige Modell- und Entscheidungsannahme nicht zu einem Problem führen; will man die Klasse linearer Modelle nicht verlassen, so sollte man trachten, daß reale Experiment durch ein Modell solcher Art zu beschreiben.

Zum zweiten kann man die Entscheidungsannahme bemängeln: Thompson (1962) maximiert die Likelihood-Funktion unter den Nebenbedingung $\hat{\sigma}_\kappa^2 \geq 0$; Federer (1968) schlägt eine zwar nichtnegative Schätzfunktion vor, die aber kaum weitere Güteeigenschaften besitzt. Es gibt andere Möglichkeiten, die dem bisherigen Vorgehen weniger fremd sind: In der Klasse aller NND quadratischen Schätzfunktionen Y'AY für q't minimiere man die Verzerrung (Bias) $\|D_M' \mathrm{vec}\, A - q\|^2$, oder aber man bestimme Minimax Mean Square Error Schätzer (§1.3.5).

Wie wir im Abschließenden Ausblick gleich begründen werden, sind diese Minimierungsaufgaben von der gleichen Bauart wie die gerade gelöste nichtnegative Schätzung von Varianzkomponenten.

Kapitel 5

ABSCHLIESSENDER AUSBLICK

1 Im Mittelpunkt des expositorischen Teils stand die ***Streuungs-Mittelwert-Korrespondenz***, die die quadratischen Streuungsschätzer als lineare Schätzfunktionen im abgeleiteten Modell darstellt. Der verwendete Tensorkalkül der Algebra reduziert aber nicht nur quadratische Funktionen auf lineare, sondern ganz allgemein multilineare. Das gibt den Anstoß, auch die Schätzung von Schiefe und Exzeß mittels transformierter Modelle abzuhandeln. Im Ergebnis konnten die heuristischen Schätzer von Anscombe (1961) in die allgemeine Theorie eingeordnet werden.

Damit erweist sich die Schätztheorie linearer Modelle für alle Parameter: Mittelwert, Streuung, Schiefe und Exzeß, als eine von Grund auf lineare Theorie. Zwar rechtfertigt sich eine solche Linearisierung durch die handlichen Verfahren, die sie liefert, doch stellt sie auch Anforderungen an die Skrupellosigkeit des Benutzers: Häufig ist der Übergang von der Menge der möglichen Parameterwerte zu ihrer linearen Hülle modellwidrig, und dieser ***Fundamentaldefekt linearer Modelle*** wird von der rein linearen Theorie einfach mißachtet.

Um diesen Defekt zu untersuchen, forderten wir in Teil III unabhängige Wiederholungen eines normalverteilten Beobachtungsvektors. Die Ergebnisse bleiben bruchstückhaft, und auch das Vorgehen bricht mit dem vorher entwickelten Stil: es widerspricht der normalverteilungsfreien Schätztheorie, auf die in der ersten Hälfte Wert gelegt wurde, und führt zu einer Art asymptotischer Aussagen für Verfahren, die im übrigen durch ihre finiten Eigenschaften charakterisiert sind.

2 Der Fundamentaldefekt umfaßt als wichtigen Sonderfall die ***nichtnegative Schätzung von Varianzkomponenten***. Dafür haben wir in Teil IV eine vollständige Lösung angeboten, die von der eingesetzten Theorie her von größerer Tragweite zu sein verspricht.

Zu dieser Aussicht gelangen wir, wenn wir mit der durch die Streuungs-Mittelwert-Korrespondenz begründeten Anschauung die nichtnegative Schätzung von Varianzkomponenten auf die Mittelwertschätzung übertragen. Betrachten wir ein lineares Modell $\mathcal{E}Y = Xb$, in dem zum ersten b beschränkt ist auf einen Kegel K_p des $\mathbb{R}^p$, und zum zweiten - und dies ist der wesentliche Punkt - Y nur Werte in einem Kegel K_n des $\mathbb{R}^n$ annimmt. Eine zu schätzende lineare Funktion Lb, $L \in \mathbb{R}^{\ell \times p}$ erzeugt den Bildkegel $L(K_p)$; es sei

$$\Pi[K_n, L(K_p)] := \{ \hat{L} \in \mathbb{R}^{\ell \times n} \mid \hat{L}(K_n) \subset L(K_p) \}$$

der Kegel aller kegelerhaltenden linearen Abbildung von $\mathbb{R}^n$ in $\mathbb{R}^\ell$. Gesucht sei nun ein ***bester kegelerhaltender Schätzer für*** Lb, der per Definition unter allen erwartungstreuen kegelerhaltenden linearen Schätzfunktionen die Norm minimiert; für einen solchen Schätzer $\hat{L}Y$ ist $\hat{L}$ optimale Lösung des Programms

$$P \qquad \begin{cases} \|\hat{L}\|^2 = \inf & \\ \hat{L} \in C(L) & (\equiv \text{Erwartungstreue}) \\ \hat{L} \in \Pi[K_n, L(K_p)]. & \end{cases}$$

(Einige Eigenschaften des Kegels Π nennt Berman 1970 pp.50-57).

Mit dieser Formulierung werden eine Reihe wichtiger Spezialfälle abgedeckt. (a) Bei der nichtnegativen Schätzung von Varianzkomponenten erhält man genau das Programm aus §1.3: Für $q \in \mathbb{R}^k_+$ gilt $q'\,\mathbb{R}^k_+ = \mathbb{R}^k_+$, und es ist $\Pi[NND(n), \mathbb{R}^k_+] = NND(n)$.

(b) Für die NND Schätzung der Streuungsmatrix sei $\overline{G}$ der Kegel derjenigen Parameterwerte $t \in \mathbb{R}^k$, für die $\sum t_\kappa V_\kappa$ NND ist; untersucht werden muß dann $\Pi[NND(n), q'\overline{G}$.

(c) Als Beispiel aus der Mittelwertschätzung betrachte die Schätzung eines nichtnegativen Mittelwertparameters $b \in \mathbb{R}^p_+$; allerdings muß auch ein Kegel K_n festgelegt werden.

(d) Verzichtet man auf Erwartungstreue und wechselt die Zielfunktion, dann erfaßt der hier diskutierte Programmtyp auch die in §4.11 erwähnten Minimum Bias und Minimax Mean Square Error NND quadratischen Schätzfunktionen für q't.

Teil IV zeichnet vor, wie ein Versuch aussehen könnte, den ***konvexen Fundamentaldefekt*** in voller Allgemeinheit zu beheben.

●

Anhang

Literatur- und Autorenverzeichnis

Albert, A. (1973). [Zitiert auf S. 321]
The Gauss-Markov theorem for regression models with possibly singular covariances.
SIAM J. Appl. Math. 24 182-187.

Anderson, T.W. (1970). [Zitiert auf S. 7, 57f]
Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices.
pp. 1-24 in:
Bose, R.C., I.M. Chakravarti, P.C. Mahalanobis, C,R. Rao & K.J.C. Smith (Eds.).
Essays in Probability and Statistics.
Chapel Hill: The University of North Carolina Press.

Anscombe, F.J. (1961). [Zitiert auf S. 7, 40, 45, 47, 49, 89]
Examination of residuals.
Proc. Fourth Berkeley Symp. Math. Statist. Probab. 1 1-36.

Ben-Israel, A. & T.N.E. Greville (1974). [Zitiert auf S. 19, 26, 33, 58, 65f, 86]
Generalized Inverses: Theory and Applications.
New York: John Wiley & Sons.

Berman, A. (1973). [Zitiert auf S. 66, 90]
Cones, Matrices, and Mathematical Programming.
Lecture Notes in Economics and Mathematical Systems Vol. 79.
Berlin: Springer-Verlag.

Brown, K.G. (1976). [Zitiert auf S. 56, 63]
Asymptotic behavior of MINQUE-type estimators of variance components.
Ann. Statist. 4 746-754.

Corbeil, R.R, & S.R. Searle (1976a). [Zitiert auf S. 7, 58]
Restricted maximum likelihood (REML) estimation of variance components in the mixed model.
Technometrics 18 31-38.

Corbeil, R.R. & S.R. Searle (1976b). [Zitiert auf S. 7, 36, 38, 58]
A comparison of variance component estimators.
Biometrics 32 779-791.

Dewess, G. (1973). [Zitiert auf S. 56]
Zur Anwendung der Schätzmethode MINQUE auf Probleme der Prozeßbilanzierung.
Math. Operationsforsch. Statist. 4 299-313.

Diplomarbeit. Siehe Pukelsheim (1974).

Drygas, H. (1970). [Zitiert auf S. 31]
The Coordinate-Free Approach to Gauss-Markov Estimation.
Lecture Notes in Operations Research and Mathematical Systems Vol. 40.
Berlin: Springer-Verlag.

Federer, W.T. (1968). [Zitiert auf S. 88]
Non-negative estimators for components of variance.
J. Roy. Statist. Soc. Ser. C Appl. Statist. 17 171-174.

Gnot, S., W. Klonecki & R Zmyślony (1976) [Zitiert auf S. 25]
Uniformly minimum variance unbiased estimation in euclidean vector spaces.
Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. 24 281-286.

Golstein, E.G. (1973). [Zitiert auf S. 77]
Konvexe Optimierung.
Berlin: Akademie-Verlag.

Greub, W.H. (1967). [Zitiert auf S. 18, 42]
Multilinear Algebra.
Berlin: Springer-Verlag.

Horn, S.D. & R.A. Horn (1975). [Zitiert auf S. 64, 69]
Comparison of estimators of heteroscedastic variances in linear models.
Amer. Statist. Assoc. 70 872-879.

Hudson, D.J. (1969). [Zitiert auf S. 78]
Least-squares fitting of a polynomial constrained to be either non-negative, non-decreasing or convex.
J. Roy. Statist. Soc. Ser. B 31 113-118.

Judge, G.G. & T. Takayama (1966). [Zitiert auf S. 78]
Inequality restrictions in regression analysis.
J. Amer. Statist. Assoc. 67 156-181.

K1effe, J. (1976). [Zitiert auf S. 20, 59]
A note on MINQUE for normal models.
Math. Operationsforsch. Statist. 7 707-714.

Kleffe, J. & R. Pincus (1974). [Zitiert auf S. 25]
Bayes and best quadratic unbiased estimators for parameters of the covar1ance matrix in a normal linear model.
Math. Operationsforsch. Statist. 5 43-67.

Krafft, 0. (1970). [Zitiert auf S. 78]
Programming methods in statistics and probability theory.
pp. 425-446 in:
Rosen, J.B., O.L. Mangasarian & K. Ritter (Eds.).
Nonlinear Programming.
New York: Academic Press.

LaMotte, L.R. (1973). [Zitiert auf S. 67f]
On non-negative quadratic unbiased estimation of variance components.
J. Amer. Statist. Assoc. 68 728-730.

Liew, C.K. (1976). [Zitiert auf S. 78]
Inequality constrained least-squares estimation.
J. Amer. Statist. Assoc. 71 746-751.

McHugh, R.B. & P.W. Mielke Jr. (1968).[Zitiert auf S. 88]
Negative variance estimates and statistical dependence in nested sampling.
J. Amer. Statist. Assoc. 65 1000-1003.

Mitra, S.K. (1971). [Zitiert auf S. 5]
Another look at Rao's MINQUE of variance components.
Bull. Inst. Internat. Statist. 44(2) 279-283.

Nelder, J.A. (1954). [Zitiert auf S. 88]
The interpretation of negative components of variance.
Biometrika 41 549-548.

Plackett, R.L. (1972), [Zitiert auf S. 25]
Studies in the history of probability and statistics XXIX.
The discovery of the methods of least squares.
Biometrika 59 239-251.

Pukelsheim, F. (1974). [Zitiert auf S. 5-7, 18]
Schätzen von Mittelwert und Streuungsmatrix in Gauss-Markoff-Modellen.
Freiburg im Breisgau: Diplomarbeit.

Pukelsheim, F. (1976). [Zitiert auf S. 18, 35]
Estimating variance components in linear models.
J. Multivariate Anal. 6 626-629.

Pukelsheim, F. (1977). [Zitiert auf S. 18, 42, 48, 67f]
On Hsu's model in regression analysis.
Math. Operationsforsch. Statist. Ser. Statist. 8 323-331.

Rao, C.R. (1972). [Zitiert auf S. 9]
Estimation of variance and covariance components in linear models.
J. Amer. Statist. Assoc. 67 112-115.

Rao, C.R. (1973). [Zitiert auf S. 4, 9, 17, 25, 31-33, 50, 57]
Linear Statistical Inference and Its Applications. Second Edition.
New York: John Wiley & Sons.

Rao, C.R. (1976). [Zitiert auf S. 25, 33]
Estimation of parameters in linear models.
Ann. Statist. 4 1023-1037.

Rao, J.N.K. (1973). [Zitiert auf S. 72]
On the estimation of heteroscedastic variances.
Biometrics 29 11-24.

Riesz, F. & B. Sz.-Nagy (1955). [Zitiert auf S. 74]
Functional Analysis.
New York: Frederick Ungar Publ. Comp.

Rockafellar, R.T. (1970). [Zitiert auf S. 9, 66, 70, 75, 77, 81, 85]
Convex Analysis.
Princeton, N.J.: Princeton University Press.

Scheffé, H. (1959). [Zitiert auf S. 4, 16f, 19]
The Analysis of Variance.
New York: John Wiley & Sons.

Searle, S.R. (1968). [Zitiert auf S. 39]
Another look at Henderson's methods of estimating variance components.
Biometrics 24 749-778 (Discussion pp. 779-787).

Searle, S.R. (1971). [Zitiert auf S. 4, 9, 16, 39, 50, 63]
Linear Models.
New York: John Wiley & Sons.

Seely, J. (1970). [Zitiert auf S. 5, 39]
Linear spaces and unbiased estimation - application to the mixed model.
Ann. Math. Statist. 41 1735-1748.

Seely, J. (1971). [Zitiert auf S. 7, 20, 35, 38, 57f]
Quadratic subspaces and completeness.
Ann. Math. Statist. 42 710-721.

Styan, G.P.H. (1973). [Zitiert auf S. 41]
Hadamard products and multivariate statistical analysis.
Linear Algebra and Appl. 6 217-240.

Thompson, W.A. Jr. (1962). [Zitiert auf S. 78, 88]
The problem of negative estimates of variance components.
Ann. Math. Statist. 33 273-289.

Witting, H. (1966). [Zitiert auf S. 5, 9, 15, 22, 58, 77]
Mathematische Statistik.
Stuttgart: B.G. Teubner.

Zyskind, G. (1975). [Zitiert auf S. 31f]
Error structures, projections and conditional inverses in linear model theory.
pp. 647-663 in:
Srivastava, J.N. (Ed.).
A Survey of Statistical Design and Linear Models.
Amsterdam: North Holland Publ. Comp.

Symbolliste

Sachregister

Abbildung, quadratische A. 18
-, symmetrische A. 42
Aitken-Modell 33
Aitken-Schätzer 21, 33, 44, 47f, 64
Asymptotik 4, 63, 89

Basismatrix, euklidische 43
Basisvektor, euklidischer 41
Bias, Minimierung des B. 25, 88, 90
BLE 24
BLIMBE 25
BLUE 24
BQUE 24

design-condition I, II 45
design-condition III 49
Dualkandidat 76, 80
Dualprogramm 76, 83, 85

Entscheidungsannahme 22-24, 88
Erwartungstreue 23, 25f, 67, 90
Exzeß 40

Fenchel 84f
Fundamentaldefekt 53f, 63f, 89f
-, konvexer F. 90

Gauss 22, 25
Gauss-Markoff-Schätzer 24
Gauss'sche Risikofunktion 22
Grundannahme der Mathematischen Statistik 15

Henderson-Schätzer 39
Heteroskedastische Varianzen 13, 26, 33, 64, 69, 72, 84, 87

Kegel 66, 75, 90
Kleinste-Quadrate-Schätzer 24f, 44, 48
konjugierte Funktion 77
Konstruktionsanleitung für A* 86
Konsistenz 63
Kovarianzkomponente 16
Kroneckerpotenz, symmetrische K. 42
Kroneckerprodukt 18

Least squares 24
Legendre 25
Lehmann-Scheffé 31f, 43, 58
Linearität, Verlust der L. 73
Linearform 22, 31, 34
-, nichtnegative L. 68
Lösung, optimale L. 75, 80, 82, 85f
-, Blitznachweis für optimale L. 76
-, zulässige L. 65, 67-69, 75

Matrixinverse, generalisierte M. 31
Maximum-Likelihood Schätzer 56-58
Mean Square Error 22
Methode der kleinsten Quadrate 25
Minimax Mean Square Error Schätzer 24, 33, 88, 90
Minimax-Probleme 77
Minimierungsprogramm 75, 78, 90
MINQUE 24, 64, 72, 79, 82, 86
Mittelwert-Translation 20
Mittelwertzerlegung, einfache M. 26
Modell, abgeleitetes M. 21
-, klassisches M. 40
-, lineares M. 14, 27
-, M-reduziertes M. 20, 40, 55
-, M. mit festen Effekten 16
-, M. mit gemischten Effekten 16
-, M. mit zufälligen Effekten 16, 26, 68
-, Q-reduziertes M. 71
-, quasinormales lineares M. 15
-, vergrößertes M. 56
Modellannahme 14, 87
Modelle, transformierte M. 21
Momente 14f, 21, 22f, 35, 41, 43, 46f, 88
Moore-Penrose Inverse 32
Multivariate Analysis 56

Negativteil 74
Negativität eliminierender Projektor 71, 86
NND, NND(n) 9, 65f, 73
-, Inneres von N. 58, 66
Norm, Minimierung der N. 23, 31, 34, 36, 38, 44, 47f, 82f, 86
Normalgleichung 31, 34, 56
-, gewichtete N. 33, 57
Normalverteilungsannahme 15, 42, 55f
Nullschätzer 26

OLSE 24
Orthogonalität 32, 66
Ordnungskegel 75

PD, PD(n) 66
Polynom dritten (vierten) Grades 40
Positivteil 74
Primalprogramm 76, 83
Programm H 65
Programm P 65, 75, 77, 79, 83, 85
Programm D 76, 79f, 83, 85
Programmerzeugende Ungleichungen 76
Produkt, äußeres P. 18
-, inneres P. 19
Projektor 14
-, Negativität eliminierender P. 71, 86

Sachregister

Quadrant, positiver 63
quadratische Funktion 18
quadratischer Teilraum symmetrischer Matrizen 351, 55-59
Quadratwurzel einer NND Matrix 33
quasinormal 15, 35, 43f, 47f

Reduktion 71
-, R. durch Erwartungstreue 23
-, R. durch Invarianz 20
-, B. durch Linearität 23
-, B. durch Negativität eliminierende Projektoren 72
-, R. multilinearer Abbildungen auf lineare 18
REML 38, 58
Ridge estimator 24
Risikofunktion 22

Schätzbarkeit, erwartungstreue S. 25f, 36
-, erwartungstreue NND quadratische S. 59
Schätzer, Aitken-S. 33, 44, 47f, 64
-, bester defekter S. 65
-, bester kegelerhaltender S. 90
-, bester nichtnegativer S. 65, 82, 86
Schätzfunktion 22
-, erwartungstreue S. 23
-, erwartungstreue NND quadratische S. 66
-, invariante quadratische S. 20, 24
-, lineare S. 24
Schiefe 40
SLSE 24
Statistik, maximalinvariante S. 20, 58
-, vollständig suffiziente S. für t 38, 57f
Stichprobenstreuung 49, 64, 72
Streuungs-Mittelwert-Korrespondenz 21, 34, 89
Summationskonvention 9
Symmetrisator 42
symmetrisch 42

Tensorprodukt 18
Test, bester T. 77

Varianz, lokale Minimierung der V. 23, 31, 35
-, gleichmäßige Minimierung der V. 23, 36
Varianzanalysen-Modell 15f, 36-39, 68f, 88
Varianzanalysen-Schätzer 38f
Varianzkomponente 16
Varianzkomponenten-Modell 14, 17, 63
Varianzschätzung, negative V. 63f, 87f
volle-Rang-Faktorisierung 58

Wert, optimaler W. 75