

Verallgemeinerte Maximum-Likelihood-Methoden und der
selbstinformativ Grenzwert

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum naturalium
(dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von
Herrn Dipl.-Math. Jan Johannes
geboren am 23.09.1973 in Berlin

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:

1. Prof. Dr. Olaf Bunke
2. Prof. Dr. Henning Läufer
3. Prof. Dr. Michael Neumann

eingereicht am: 06. September 2002
Tag der mündlichen Prüfung: 16. Dezember 2002

Zusammenfassung

Es sei \mathbf{X} eine das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta | \theta \in \Theta)$ beschreibende Zufallsvariable. Ein zentrales Problem der Mathematischen Statistik ist, mit Hilfe einer Beobachtung $\mathbf{X} = x$ den Parameter $\theta \in \Theta$ bzw. einen abgeleiteten Parameter $\varphi(\theta) \in \Gamma$ zu schätzen. Im Fall einer dominierten Verteilungsfamilie $\{P_\theta | \theta \in \Theta\}$ ist es möglich, das Maximum-Likelihood-Prinzip (MLP) anzuwenden. Eine Alternative zu der Konstruktion einer Maximum-Likelihood-Schätzung (MLS) stellt ein Bayesscher Ansatz dar. Unter Regularitätsbedingungen erweist sich nun, dass die MLS mit dem Grenzwert einer Folge von Bayesschen Schätzungen (Bunke et al. (1976)) übereinstimmt. Damit ist der Grenzwert unabhängig von der a priori Gewichtung. Ferner kann eine Bayessche Schätzung auch in einer nicht dominierten Verteilungsfamilie definiert werden, was eine Möglichkeit zur Erweiterung des MLPs auf nicht dominierte Verteilungsfamilien liefert. Außerdem werden die bereits existierenden Ansätze einer verallgemeinerten MLS (vMLS) von Kiefer und Wolfowitz (1956) sowie Gill (1989) vorgestellt. Basierend auf diesen bekannten Ergebnissen definieren wir einen selbstinformativen Grenzwert und a posteriori Träger. Insbesondere kann der selbstinformativ Grenzwert als Beschreibung einer nichtinformativen Schätzung auf der Grundlage eines Bayesschen Ansatzes und als eine mögliche Verallgemeinerung des MLPs betrachtet werden. Im Spezialfall einer dominierten Verteilungsfamilie geben wir hinreichende Bedingungen an, unter denen die Menge der MLSen einem selbstinformativen a posteriori Träger oder, falls die MLS eindeutig ist, einem selbstinformativen Grenzwert entspricht. Das Ergebnis für den selbstinformativen a posteriori Träger wird dann auf ein allgemeineres Modell ohne dominierte Verteilungsfamilie erweitert. Insbesondere wird gezeigt, dass die Menge der vMLSen nach Kiefer und Wolfowitz ein selbstinformativer a posteriori Träger ist. Weiterhin wird der selbstinformativ Grenzwert bzw. a posteriori Träger in einem Modell mit nicht identifizierbarem Parameter bestimmt. In dem Spezialfall eines normalen linearen Modells entspricht dann der selbstinformativ Grenzwert einer MLS. Im Mittelpunkt dieser Arbeit steht ein multivariates semiparametrisches lineares Modell. Dazu betrachten wir zuerst ein nichtparametrisches Modell, wobei die bekannte vMLS nach Kiefer und Wolfowitz sowie nach Gill die empirische Verteilungsfunktion ist. Wir weisen nach, dass unter der a priori Annahme eines Dirichlet Prozesses der selbstinformativ Grenzwert existiert und der empirischen Verteilungsfunktion entspricht. Anschließend untersuchen wir das multivariate semiparametrische lineare Modell und bestimmen die vMLSen nach Kiefer und Wolfowitz oder Gill sowie unter der a priori Annahme eines Dirichlet Prozesses und einer Normal-Wishart-Verteilung den selbstinformativen Grenzwert. Im Allgemeinen sind die so erhaltenen Schätzungen verschieden. Sowohl die vMLSen nach Kiefer und Wolfowitz als auch die vMLSen nach Gill sind nur in Spezialfällen, der selbstinformativ Grenzwert dagegen immer eindeutig. Weiterhin ist der selbstinformativ Grenzwert immer eine vMLS nach Gill, aber nur in Spezialfällen eine vMLS nach Kiefer und Wolfowitz. Abschließend gehen wir dann auf den Spezialfall eines semiparametrische Lokationsmodell ein, in dem die vMLSen nach Kiefer und Wolfowitz sowie Gill und der selbstinformativ Grenzwert wieder identisch sind.

Inhaltsverzeichnis

Einleitung	1
Notationen	4
1 Schätzproblem	6
1.1 Statistisches Experiment	6
1.2 Schätzung eines abgeleiteten Parameters	6
1.3 Bayessches Modell	7
2 Dominierte Verteilungsfamilie	10
2.1 Modellannahmen	10
2.2 Maximum-Likelihood-Methode	10
2.3 Grenzwert einer Folge Bayesscher Schätzungen	11
2.3.1 Modellannahmen	11
2.3.2 Bayessche Schätzung	12
2.3.3 Grenzwert einer Folge Bayesscher Schätzungen	12
3 Verallgemeinerungen des Maximum-Likelihood-Prinzips	15
3.1 Ansatz von Kiefer und Wolfowitz	15
3.2 Ansatz von Gill	16
3.3 Vergleich der Ansätze	17
4 Selbstinformativer Grenzwert	18
4.1 Modellannahmen	18
4.2 Iterative Prozedur	18
4.3 Definition des selbstinformativen Grenzwertes	24
4.4 Dominierte Verteilungsfamilie	25
4.5 Nicht dominierte Verteilungsfamilie	29
4.6 Modell mit nicht identifizierbarem Parameter	35
4.6.1 Bayessches Modell	35
4.6.2 Selbstinformativer Grenzwert	36
4.6.3 Multivariates normales lineares Modell	37
4.6.4 Multivariates normales zeitabhängiges lineares Modell	43
5 Nichtparametrisches Modell	46
5.1 Modellannahmen	46
5.2 Ansatz von Kiefer und Wolfowitz	47
5.3 Ansatz von Gill	50

5.4	Selbstinformativer Grenzwert	52
5.4.1	Zufälliges Wahrscheinlichkeitsmaß	52
5.4.2	Dirichlet Prozess	55
5.4.3	Bayessche Schätzung	57
5.4.4	Selbstinformativer Grenzwert	57
6	Semiparametrisches lineares Modell	59
6.1	Modellannahmen	59
6.2	Ansatz von Kiefer und Wolfowitz	60
6.3	Ansatz von Gill	62
6.4	Selbstinformativer Grenzwert	64
6.4.1	A priori Annahmen	64
6.4.2	Bayessche Schätzungen	65
6.4.3	Selbstinformativer Grenzwert	67
6.5	Semiparametrisches Lokationsmodell	82
7	Ausblick	84
	Literaturverzeichnis	85

Einleitung

Beobachtungen sind die Grundlage eines jeden Experiments. Basierend auf diesen Daten ist das Ziel einer statistischen Analyse, Schlussfolgerungen über das beobachtete Objekt zu erzielen. Dazu werden die zu beobachtenden Daten als „zufällig“ aufgefasst. Mit „zufällig“ beschreiben wir eine Situation, in der wir nicht die zugrunde liegenden Mechanismen kennen. Diese Zufallsabhängigkeit modellieren wir mathematisch mit Hilfe der Begriffsbildung der Wahrscheinlichkeitstheorie (z.B. [Bauer \(1991\)](#)). In diesem Sinn fassen wir die Beobachtung x als Realisierung einer Zufallsvariable \mathbf{X} auf und unterstellen somit, dass sich die zugrunde liegenden Mechanismen mit Hilfe einer Wahrscheinlichkeitsverteilung P beschreiben lassen. Das Ziel der Mathematischen Statistik ist nun, mit Hilfe einer Beobachtung x Schlussfolgerungen über die unbekannte Verteilung P zu erzielen. Ein erster Schritt einer statistischen Analyse ist die Beschreibung und somit die Wahl einer geeigneten Familie von Verteilungen \mathcal{F} , in der die unbekannte Verteilung P enthalten ist oder deren Elemente die Verteilung P zumindest hinreichend gut approximieren. Wählen wir eine „kleine“ Familie \mathcal{F} , ist also nur zwischen „wenigen“ Verteilungen zu entscheiden, so lässt sich im Allgemeinen mit der in einer Beobachtung enthaltenen „Information“ eine genauere Aussage treffen. Jedoch ist der Wert der Aussage gering, wenn die unbekannte Verteilung P nicht annähernd durch ein Element der Verteilungsfamilie \mathcal{F} approximiert wird. Somit kommt der Verteilungsannahme, der Wahl der Verteilungsfamilie \mathcal{F} , eine grundlegende Bedeutung zu.

Im Allgemeinen sind wir nicht an der Kenntnis der unbekanntes Verteilung, sondern nur an einem abgeleiteten Parameter $\gamma(P) \in \Gamma$, wie z.B. Erwartungswert, Median oder Varianz der unbekanntes Verteilung P , interessiert. An Hand einer Beobachtung der unbekanntes Verteilung P möchten wir nun den Parameter $\gamma(P)$ spezifizieren (schätzen), also einen Wert $\hat{\gamma} \in \Gamma$ wählen. Auf Grund der Zufälligkeit unserer Beobachtung wird mit Ausnahme von wenigen Spezialfällen der geschätzte Wert $\hat{\gamma}$ nicht dem gesuchten Parameter $\gamma(P)$ entsprechen, so dass Kriterien benötigt werden, die eine „optimale“ Schätzung des Parameters $\gamma(P)$ charakterisieren. Die Untersuchung von Optimalitätskriterien und diesbezüglich konstruierter optimaler Schätzungen $\hat{\gamma} \in \Gamma$ an Hand einer Beobachtung sind die wesentlichen Bestandteile der Schätztheorie in der Mathematischen Statistik (vgl. [Schmetterer \(1966\)](#), [Witting \(1985\)](#), [Strasser \(1985\)](#)). Eine häufig angewandte und ebenso häufig auch kritisierte Konstruktionsmethode einer Schätzung ist das Maximum-Likelihood-Prinzip (MLP). Eine zentrale Voraussetzung für die Anwendbarkeit des MLP ist die Wahl einer dominierten Verteilungsfamilie \mathcal{F} , d.h., für alle Elemente der Verteilungsfamilie \mathcal{F} existiert ein dominierendes σ -endliches Maß. Jedoch ist diese Einschränkung in einer Vielzahl von Situationen nicht sinnvoll. Ein klassisches Beispiel ist das nichtparametrische Modell, d.h. es liegt keine Information über die Struktur der unbekanntes Verteilung P vor, so dass wir die Familie \mathcal{P} aller Verteilungen wählen. Somit stellt sich die Frage, ob das MLP auch auf Situationen in denen die Verteilungsfamilie nicht dominiert ist, übertragen werden kann. Alternativ zu der frequentistischen Konstruktion von Schätzungen, wie dem MLP, kann auch ein Bayesscher Ansatz betrachtet werden. Auf der grundlegenden Annahme einer a priori Gewichtung der Werte $\gamma \in \Gamma$ werden mit Hilfe einer Beobachtung die Gewichte der Werte $\gamma \in \Gamma$ angepasst und an Hand dieser a posteriori Gewichtung eine Bayessche Schätzung konstruiert. Damit können auf eine elegante Weise neben der

Beobachtung auch vorhandene Informationen über den Wert $\gamma(P)$ in die Schätzung einfließen. Insbesondere ist die Konstruktion einer Bayesschen Schätzung auch möglich, wenn die Verteilungsfamilie nicht dominiert ist. Andererseits ist die Annahme einer a priori Gewichtung ein zentraler Kritikpunkt des Bayesschen Ansatzes, da diese eine Vorinformation über die Werte $\gamma \in \Gamma$ voraussetzt. Somit besteht das Interesse an der Beschreibung einer Bayesschen Schätzung in einer Situation ohne Vorinformation.

Überblick

Zum besseren Verständnis der Zusammenhänge der in den folgenden Kapiteln vorgestellten Resultate werden wir jetzt die grundsätzlichen Ideen im Überblick darstellen.

Im Kapitel 1 werden grundlegende Begriffe und Definitionen der Mathematischen Statistik vorgestellt. Insbesondere wird der Bayessche Ansatz detailliert dargelegt. Der Spezialfall einer dominierten Verteilungsfamilie wird im Kapitel 2 betrachtet. In dieser Situation existiert dann unter Regularitätsbedingungen eine Maximum-Likelihood-Schätzung (MLS). Weiterhin werden wir den Bayesschen Ansatz untersuchen und eine Möglichkeit der Konstruktion einer Bayesschen Schätzung in einer Situation ohne Vorinformation angeben. Dazu wählen wir eine spezielle Folge von Bayesschen Schätzungen, wobei der Grenzwert dieser Folge, falls er existiert, eine Situation ohne Vorinformation charakterisiert. Insbesondere entspricht dieser Grenzwert unter Regularitätsbedingungen der MLS in dieser Verteilungsfamilie (Bunke et al. (1976)) und ist somit offensichtlich unabhängig von der a priori Gewichtung. Dieses Ergebnis ist der zentrale Ausgangspunkt für die weiteren Ausführungen. Zum einen liefert die Betrachtung des Grenzwertes der Folge von Bayesschen Schätzungen eine mögliche Charakterisierung einer Situation ohne Vorinformation. Zum anderen hängt die Definition der Bayesschen Schätzung nicht von der einschränkenden Annahme einer dominierten Verteilungsfamilie ab, so dass ein möglicher Ansatz zur Übertragung des MLP's auf nicht dominierte Verteilungsfamilien besteht. In den vergangenen Jahren wurde eine Vielzahl an alternativen Ansätzen einer Verallgemeinerung des MLP's in der Literatur diskutiert, wie z.B. die „Methode der Sieves“ (Grenander (1981)), „Penalized Likelihood“ (Geman und Hwang (1982)) oder „Empirical Likelihood“ (Owen (2001)). Diese Arbeit hat nicht das Ziel, einen vollständigen Überblick über die verschiedenen Methoden und Ansätze zu geben. Wir möchten vielmehr das alternative Verfahren eines Grenzwertes einer Folge von Bayesschen Schätzungen zwei bekannten Konzepten der Verallgemeinerung des MLP's gegenüberstellen. So werden wir zuerst im Kapitel 3 den Ansatz von Kiefer und Wolfowitz (1956) sowie von Gill (1989) einer Verallgemeinerung des MLP's vorstellen. Für eine ausführliche Diskussion dieser Konzepte verweisen wir auf die zugrunde liegenden Arbeiten. Beide Konzepte definieren für eine fixierte Beobachtung eine verallgemeinerte Maximum-Likelihood-Schätzung (vMLS) und unterscheiden sich somit von den benannten Alternativen einer Verallgemeinerung des MLP's, die auf der Basis einer mit dem Stichprobenumfang wachsenden Verteilungsfamilie Folgen von Schätzungen konstruieren. Insbesondere entspricht eine vMLS nach Kiefer und Wolfowitz und unter Regularitätsbedingungen eine vMLS nach Gill der MLS, falls diese existiert. Ferner weisen wir nach, dass unter Regularitätsbedingungen eine vMLS nach Kiefer und Wolfowitz auch eine vMLS nach Gill ist. Anschließend charakterisieren wir im Kapitel 4 die Folge der Bayesschen Schätzungen mittels einer iterativen Prozedur. Konvergiert nun die Folge der zugehörigen Bayesschen Schätzungen, so wird der erhaltene Grenzwert als selbstinformativer Grenzwert bezeichnet (Bunke (2002)). Wir werden die in der Arbeit von Bunke und Johannes (2002) weiter spezifizierten Ideen hier detailliert darlegen. So wird unter anderem der Begriff eines schwachen selbstinformativen a posteriori Trägers eingeführt. Ferner untersuchen wir drei Spezialfälle. Wir werden hinreichende Bedingungen für die Existenz eines selbstinformativen Grenzwertes und eines schwachen selbstinformativen a posteriori Trägers unter der Annahme einer dominierten Verteilungsfamilie angeben. Insbesondere ist die Menge der MLS unter Regularitätsbedingungen ein schwacher selbstinformativer a posteriori

Träger. Weiterhin geben wir für eine nicht dominierte Verteilungsfamilie hinreichende Bedingungen für die Existenz eines schwachen selbstinformativen a posteriori Trägers an, wobei dann unter Regularitätsbedingungen die Menge der vMLS nach Kiefer und Wolfowitz einen schwachen selbstinformativen a posteriori Träger darstellt. Abschließend betrachten wir den Spezialfall eines nicht identifizierbaren Parameters, der einen weiteren Einblick in die Struktur des selbstinformativen Grenzwertes sowie des schwachen selbstinformativen a posteriori Trägers gewährt. Wir verzichten hier auf die Darstellung der Ergebnisse für den Fall der Schätzung einer Hazardfunktion auf der Basis zensierter Daten (vgl. [Bunke und Johannes \(2002\)](#)) und bemerken nur, dass unter Regularitätsbedingungen der selbstinformativ Grenzwert der vMLS nach Gill entspricht. Dagegen werden wir detailliert auf ein multivariates semiparametrisches lineares Modell eingehen. Dazu betrachten wir zuerst im Kapitel 5 den klassischen Fall eines nichtparametrischen Modells. Wir nehmen an, dass die unbekannte Verteilung P in der Familie aller Verteilungen \mathcal{P} variiert. Im Allgemeinen ist diese Verteilungsfamilie \mathcal{P} nicht dominiert, so dass die MLS nicht definiert ist. Insbesondere ist die vMLS, die empirische Verteilungsfunktion der Beobachtungen, nach [Kiefer und Wolfowitz \(1956\)](#) sowie [Gill \(1989\)](#) bekannt. Wir werden nachweisen, dass unter der Annahme eines Dirichlet Prozesses der selbstinformativ Grenzwert ebenfalls der empirischen Verteilungsfunktion entspricht und somit eine von der a priori Annahme unabhängige Schätzung darstellt. Abschließend untersuchen wir im Kapitel 6 ein multivariates semiparametrisches lineares Modell und bestimmen die vMLS nach Kiefer und Wolfowitz oder Gill sowie den selbstinformativ Grenzwert. Im Allgemeinen sind die so erhaltenen Schätzungen verschieden. So ist die vMLS nach Kiefer und Wolfowitz sowie nach Gill nur in Spezialfällen, der selbstinformativ Grenzwert dagegen immer eindeutig. Weiterhin ist der selbstinformativ Grenzwert immer eine vMLS nach Gill, aber nur in Spezialfällen nach Kiefer und Wolfowitz. Am Ende dieses Kapitels gehen wir dann auf den Spezialfall eines semiparametrischen Lokationsmodells ein, in dem die vMLS nach Kiefer und Wolfowitz sowie Gill und der selbstinformativ Grenzwert wieder identisch sind. Abschließend geben wir einen Ausblick auf offene Fragen und Variationen der Modellannahmen (Kapitel 7).

Regelmäßig benutzte Symbole und Abkürzungen werden im Anschluss an die Einleitung zusammengefasst.

Mein besonderer Dank gilt meinem Betreuer, Herrn Professor Dr. O. Bunke, für die vielen wertvollen Hinweise und anregenden Diskussionen, die den Fortgang meiner Arbeit wesentlich beeinflussten, sowie meiner Familie und meinen Freunden für ihre intensive Unterstützung und ihr Verständnis in den vergangenen Jahren. Vor allem danke ich meiner Schwester für den Rückhalt und die Kraft, die sie mir in vielen Gesprächen gab. Für ihre Ruhe und Gelassenheit auch nach etlichen Stunden intensiver Befragung möchte ich Dr. R. Thrum und Dr. B. Droge danken.

Notationen

Die Bezeichnungen folgen den gewöhnlichen Konventionen. Die benutzten mathematischen sowie spezielle im Text auftretende Symbole werden in der folgenden Tabelle zusammengefasst.

Mit der Ausnahme von Annahmen, sie werden getrennt nummeriert und es steht ein A vor der Kapitelnummer, wird innerhalb eines Kapitels fortlaufend nummeriert. Ferner zeigt das Symbol \square das Ende eines Beweises, einer Bemerkung bzw. eines Beispiels an.

Symbolverzeichnis

$A := B$	A ist definiert durch B
$A \subset B$	A ist enthalten in B
A^c	Komplement einer Menge A
$[a, b], (a, b)$	abgeschlossenes, offenes Intervall von a nach b
\mathbb{N}, \mathbb{N}_0	$\{1, 2, \dots\}, \{0, 1, 2, \dots\}$
\mathbb{R}, \mathbb{R}^+	$(-\infty, \infty), [0, \infty)$
\mathbb{R}^k	k -dimensionaler euklidischer Raum
$\mathbb{R}^{n \times p}$	Menge der $n \times p$ Matrizen
$K_\varepsilon(x)$	offene Kugel mit Radius $\varepsilon > 0$ um x
$\overline{K}_\varepsilon(x)$	abgeschlossene Kugel mit Radius $\varepsilon > 0$ um x
$\Theta, \mathcal{X}, \Gamma$	Parameterraum, Menge der Beobachtungen, Menge der abgeleiteten Parameter
$\mathbf{X}, \vartheta, \gamma$	Zufallsvariable mit Werten in $\mathcal{X}, \Theta, \Gamma$
$\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}$	σ -Algebra über $\Omega, \mathcal{X}, \Theta, \Gamma$
$\mathfrak{L}, \mathfrak{L}^p, \mathfrak{L}^{n \times p}$	Borel- σ -Algebra über $\mathbb{R}, \mathbb{R}^p, \mathbb{R}^{n \times p}$
$\mathcal{X} \times \Theta$	kartesisches Produkt von \mathcal{X} und Θ
$\mathfrak{B} \otimes \mathfrak{C}$	Produkt- σ -Algebra über $\mathcal{X} \times \Theta$
\mathcal{X}^n	n -faches kartesisches Produkt $\mathcal{X} \times \dots \times \mathcal{X}$
x^n	Vektor (x_1, \dots, x_n) in \mathcal{X}^n
$\mathbb{1}_n^t \otimes x$	Vektor (x, \dots, x) in \mathcal{X}^n
μ, ν	σ -endliche Maße
$\mu \times \nu$	Produktmaß
$\mathcal{P}, \mathcal{P}_+, \mathcal{P}_+^0$	Menge aller Verteilungen, mit endlichem zweiten Moment, mit Erwartungswert Null und endlichem zweiten Moment
$P, P_\theta, P^{\mathbf{X}}$	Wahrscheinlichkeitsverteilungen
$P_n \xrightarrow{w} P$	schwache Konvergenz der Maße P_n gegen P
$P^{\mathbf{X}} \mathbf{Y}=y$	bedingte Wahrscheinlichkeitsverteilung
δ_x	Punktmaß an der Stelle x
\hat{F}_{x^n}	empirische Verteilung der Werte x_1, \dots, x_n

$N_{np}(A, \Sigma)$	multivariate Normalverteilung auf $\mathbb{R}^{n \times p}$
$NW(l, k, p, M_0, A_0, S_0)$	Normal-Wishart-Verteilung
$GMt(l, k, p, M_0, A_0, S_0)$	verallgemeinerte t-Verteilung
$\mathcal{D}_{\alpha\mu}$	Dirichlet Prozess mit Parameter α und μ
$\mathbb{E} \mathbf{X}, \text{Cov} \mathbf{X}$	Erwartungswert, Kovarianzmatrix von \mathbf{X}
$\mathbb{E}_{\boldsymbol{\vartheta} \boldsymbol{\gamma}}[g(\boldsymbol{\vartheta})], \mathbb{E}[g(\boldsymbol{\vartheta}) \boldsymbol{\gamma}]$	bedingte Erwartung unter $P^{\boldsymbol{\vartheta} \boldsymbol{\gamma}}$
$\hat{\mu}, \hat{\sigma}^2, \hat{B}, \hat{\Sigma}, \hat{\varphi}$	Schätzung für $\mu, \sigma^2, B, \Sigma, \varphi$
$\mathbb{1}_k, J_k, I_k$	k -dimensionaler Einsvektor, $k \times k$ Einsmatrix, $k \times k$ Einheitsmatrix
$A^t, sp[A], \text{Rang}[A], A $	Transponierte, Spur, Rang, Determinante von A
A^{-1}, A^+	Inverse, Moore-Penrose Inverse von A
$\mathcal{N}(M), \mathcal{R}(M)$	Kern, Bild einer linearen Abbildung M
$P_{\mathcal{N}(M)}, P_{\mathcal{R}(M)}$	Projektionsmatrizen
$A \otimes B$	Kroneckerprodukt
$\text{Diag}[\lambda_1, \dots, \lambda_k]$	Diagonalmatrix

Kapitel 1

Schätzproblem

1.1 Statistisches Experiment

Der Ausgangspunkt unserer Untersuchungen ist eine Beobachtung x , die wir immer als eine Realisierung einer Zufallsvariable \mathbf{X} auffassen. Analog sind die Beobachtungen x_1, \dots, x_n Realisierungen von Zufallsvariablen $\mathbf{X}_1, \dots, \mathbf{X}_n$. Ferner werden wir häufig die Beobachtungen x_1, \dots, x_n zu einem Vektor $x^n = (x_1, \dots, x_n)$ zusammenfassen und x^n als eine Realisierung der Zufallsvariable $\mathbf{X}^n = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ ansehen. Wir schreiben dann abkürzend $\mathbf{X} = x$ oder $\mathbf{X}^n = x^n$. Die Menge der Beobachtungen bezeichnen wir generell mit \mathcal{X} oder \mathcal{X}^n dem n -fachen kartesischen Produkt $\mathcal{X} \times \dots \times \mathcal{X}$. Wir werden im Folgenden annehmen, dass die Menge \mathcal{X} mit einer Metrik $d_{\mathcal{X}}$ versehen sei, und dass der metrische Raum $(\mathcal{X}, d_{\mathcal{X}})$ vollständig und separabel (polnisch) sei. Die induzierte Borel- σ -Algebra auf \mathcal{X} bezeichnen wir mit \mathfrak{B} . Die induzierte Borel- σ -Algebra \mathfrak{B}^n auf dem polnischen Raum \mathcal{X}^n entspricht dann dem n -fachen Kartesischen Produkt $\mathfrak{B} \times \dots \times \mathfrak{B}$ der Borel- σ -Algebra \mathfrak{B} (vgl. Billingsley (1968), S. 224). Weiterhin fassen wir o.B.d.A. die Zufallsvariable \mathbf{X} als eine messbare Abbildung von dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ in den messbaren Raum $(\mathcal{X}, \mathfrak{B})$ auf. Die Zufallsvariable \mathbf{X} induziert dann eine Wahrscheinlichkeitsverteilung $P^{\mathbf{X}}$ auf \mathfrak{B} mit $P^{\mathbf{X}}(B) = P(\mathbf{X} \in B)$ für $B \in \mathfrak{B}$. Die Verteilung $P^{\mathbf{X}}$ ist uns unbekannt und mit Hilfe einer Beobachtung x der Zufallsvariable \mathbf{X} möchten wir nun Rückschlüsse auf die zugrunde liegende Verteilung $P^{\mathbf{X}}$ ziehen. Dazu nehmen wir weiterhin an, dass die unbekannt Verteilung $P^{\mathbf{X}}$ in einer Familie \mathcal{F} von Verteilungen auf \mathcal{X} enthalten ist. Wir sprechen dann von einer adäquaten Verteilungsannahme. Äquivalent dazu können wir auf dem messbaren (Ω, \mathfrak{A}) eine Familie von Wahrscheinlichkeitsverteilungen annehmen, so dass \mathcal{F} dann die Menge der von \mathbf{X} induzierten Verteilungen auf \mathcal{X} darstellt. Die Familie der Verteilungen \mathcal{F} werden wir o.B.d.A. immer als parametrisiert $\mathcal{F} = \{P_{\theta} \mid \theta \in \Theta\}$ annehmen. Die Menge Θ heißt Parameterraum und könnte zum Beispiel die Menge \mathcal{F} selbst sein. Wir sprechen von einer parametrischen Verteilungsfamilie, falls der Parameterraum Θ als eine Teilmenge des k -dimensionalen euklidischen Raumes \mathbb{R}^k gewählt werden kann. Ist dies nicht der Fall, so bezeichnen wir die Verteilungsfamilie als nichtparametrisch. Zusammenfassend nennen wir $(\mathcal{X}, \mathfrak{B}, P_{\theta} \mid \theta \in \Theta)$ ein statistisches Experiment und schreiben abkürzend, dass die Zufallsvariable \mathbf{X} das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_{\theta} \mid \theta \in \Theta)$ beschreibt.

1.2 Schätzung eines abgeleiteten Parameters

Häufig möchten wir mit Hilfe einer Beobachtung $\mathbf{X} = x$ nur auf spezielle Eigenschaften der zugrunde liegenden unbekannt Verteilung $\mathbf{X} \sim P_{\theta}$, $\theta \in \Theta$ schließen. Dazu nehmen wir an, dass über der

Menge Θ eine σ -Algebra \mathfrak{C} gegeben ist und betrachten eine messbare Funktion φ von (Θ, \mathfrak{C}) in den messbaren Raum (Γ, \mathfrak{D}) . Für ein $\theta \in \Theta$ heißt $\varphi(\theta) \in \Gamma$ abgeleiteter Parameter. Unser Interesse ist nun, mit Hilfe einer Beobachtung x der Zufallsvariable $\mathbf{X} \sim P_\theta$ ($\theta \in \Theta$) den abgeleiteten Parameter $\varphi(\theta) \in \Gamma$ zu schätzen. Eine mögliche Schätzung für $\varphi(\theta)$ ist dann jeder Wert $\gamma \in \Gamma$. Damit erhalten wir auf natürliche Weise das Problem der Wahl eines Wertes $\gamma_x \in \Gamma$ für eine Beobachtung $\mathbf{X} = x$.

Bemerkung 1.1 Ein mögliches Maß für die Qualität einer Schätzung ist die Risikofunktion. Dazu betrachten wir eine messbare Funktion $\hat{\varphi}$ von $(\mathcal{X}, \mathfrak{B})$ nach (Γ, \mathfrak{D}) . $\hat{\varphi}$ bezeichnen wir als Schätzfunktion. Eine Schätzung $\hat{\varphi}(x) \in \Gamma$ ist somit der Funktionswert von $\hat{\varphi}$ an der Stelle x . Eine nichtnegative Funktion $L(\gamma_1, \gamma_2)$ auf der Menge $\Gamma \times \Gamma$ bezeichnen wir als Verlustfunktion. Für eine Realisierung x der Zufallsvariable \mathbf{X} mit der Verteilung P_θ ($\theta \in \Theta$) spiegelt $L(\hat{\varphi}(x), \varphi(\theta))$ dann den Verlust für die Wahl von $\hat{\varphi}(x) \in \Gamma$ wieder. Das bezüglich der Verteilung P_θ von \mathbf{X} gewichtete Mittel der Werte $L(\hat{\varphi}(x), \varphi(\theta))$ über alle möglichen Beobachtungen $x \in \mathcal{X}$,

$$R_L(\hat{\varphi}, \varphi(\theta)) := \int_{\mathcal{X}} L(\hat{\varphi}(x), \varphi(\theta)) P_\theta(dx),$$

heißt nun Risikofunktion. Insbesondere nennen wir eine Schätzfunktion $\hat{\varphi}_*$ gleichmäßig beste Schätzfunktion in einer Klasse \mathcal{D} von Schätzfunktionen $\hat{\varphi}$ für den abgeleiteten Parameter φ , falls für alle $\theta \in \Theta$ und $\hat{\varphi} \in \mathcal{D}$ gilt $R_L(\hat{\varphi}_*, \varphi(\theta)) \leq R_L(\hat{\varphi}, \varphi(\theta))$. \square

1.3 Bayessches Modell

Die zentrale Idee des Bayesschen Modells ist es, den Parameter θ als eine Realisierung einer Zufallsvariable ϑ anzusehen. Unter der Bedingung $\vartheta = \theta$ beschreibt nun die Zufallsvariable \mathbf{X} das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta | \theta \in \Theta)$. Insbesondere nehmen wir an, dass der Parameterraum Θ mit einer Metrik d_Θ versehen und der metrische Raum (Θ, d_Θ) polnisch sei. \mathfrak{C} bezeichnet dann die induzierte Borel- σ -Algebra über Θ . Der Bayessche Ansatz bedeutet nun, auf dem messbaren Raum (Θ, \mathfrak{C}) eine a priori Verteilung anzunehmen und die statistischen Aussagen über den Parameter mit Hilfe der a posteriori Verteilung, die aus der a priori Verteilung und einer Realisierung der Zufallsvariable \mathbf{X} bestimmt wird, zu gewinnen.

Wir gehen wie bisher von einem messbaren Raum (Ω, \mathfrak{A}) aus und fassen o.B.d.A. \mathbf{X} und ϑ als messbare Abbildung von Ω nach \mathcal{X} bzw. Θ auf. Die a priori Verteilung der Zufallsvariable ϑ auf \mathfrak{C} bezeichnen wir mit Π . Insbesondere sehen wir nun eine Verteilung P_{θ_0} aus der Familie der Verteilungen $\{P_\theta | \theta \in \Theta\}$ als bedingte Verteilung $P^{\mathbf{X} | \vartheta = \theta}$ von \mathbf{X} unter der Bedingung $\vartheta = \theta_0$ an. Mittels der Familie der bedingten Verteilungen $\{P_\theta | \theta \in \Theta\}$ und der a priori Verteilung Π definieren wir eine Verteilung P auf dem messbaren Raum (Ω, \mathfrak{A}) durch

$$P(\mathbf{X}^{-1}(B) \cap \vartheta^{-1}(C)) = \int_C P_\theta(B) \Pi(d\theta) \quad \text{für alle } B \in \mathfrak{B}, C \in \mathfrak{C}.$$

Insbesondere ist die Borel- σ -Algebra auf dem polnischen Raum $(\mathcal{X} \times \Theta)$ dann das Kartesische Produkt $\mathfrak{B} \times \mathfrak{C}$ (vgl. Billingsley (1968), S. 224) und der zufällige Vektor (\mathbf{X}, ϑ) induziert nun auf dem messbaren Raum $(\mathcal{X} \times \Theta, \mathfrak{B} \times \mathfrak{C})$ eine Verteilung $P^{\mathbf{X}, \vartheta}$. Die a priori Verteilung Π von ϑ entspricht der Randverteilung bzgl. der gemeinsamen Verteilung $P^{\mathbf{X}, \vartheta}$ von (\mathbf{X}, ϑ) , so dass wir sie auch mit P^ϑ

bezeichnen. Die Randverteilung $P^{\mathbf{X}}$ von \mathbf{X} ergibt sich aus

$$P^{\mathbf{X}}(B) = P(\mathbf{X}^{-1}(B) \cap \vartheta^{-1}(\Theta)) = \int_{\Theta} P_{\theta}(B) P^{\vartheta}(d\theta) \quad \text{für alle } B \in \mathfrak{B}.$$

Eine reguläre bedingte Verteilung $P^{\vartheta | \mathbf{X}=x}$ (vgl. Bellach et al. (1978), S. 112, Definition 1) von ϑ gegeben eine Beobachtung $\mathbf{X} = x$ heißt a posteriori Verteilung. Ferner existiert eine reguläre bedingte Verteilung, wenn Θ , wie hier angenommen, ein polnischer Raum ist (vgl. Ash (1972), S. 265, Satz 6.6.5 und 6.6.6, Bellach et al. (1978), S. 113, Satz 2). Des weiteren ist die a posteriori Verteilung nur bis auf $P^{\mathbf{X}}$ -Nullmengen eindeutig bestimmt, so dass die folgenden Aussagen immer für eine Festlegung der a posteriori Verteilung gelten.

Es sei eine Beobachtung x von \mathbf{X} unter der Bedingung $\vartheta = \theta$ und eine Festlegung $P^{\vartheta | \mathbf{X}=x}$ der a posteriori Verteilung gegeben. Eine mit Hilfe von $P^{\vartheta | \mathbf{X}=x}$ konstruierte Schätzung $\hat{\varphi}(x)$ für den abgeleiteten Parameter $\varphi(\theta)$ bezeichnen wir als Bayessche Schätzung. Ist es zum Beispiel möglich den Erwartungswert von $\varphi(\vartheta)$ bzgl. der a posteriori Verteilung $P^{\vartheta | \mathbf{X}=x}$ zu definieren, so wäre

$$\hat{\varphi}(x) := \int_{\Theta} \varphi(\theta) P^{\vartheta | \mathbf{X}=x}(d\theta) \quad (1.1)$$

eine mögliche sensible Schätzung für $\varphi(\theta)$.

Bemerkung 1.2 Für eine gegebene Verlustfunktion L und a priori Verteilung Π untersuchen wir basierend auf der vorgestellten Risikofunktion (vgl. Bemerkung 1.1) das Bayessche Risiko

$$R_L[\hat{\varphi}, \Pi] := \int_{\Theta} R_L(\hat{\varphi}, \varphi(\theta)) \Pi(d\theta) \quad (1.2)$$

als ein Maß der Qualität einer Schätzfunktionen $\hat{\varphi}$ für den abgeleiteten Parameter φ im Bayesschen Kontext. Wir bezeichnen eine Schätzfunktion $\hat{\varphi}_*$ als Bayessche Schätzfunktion in einer Klasse \mathcal{D} von Schätzfunktionen $\hat{\varphi}$ für den abgeleiteten Parameter φ , falls $R_L[\hat{\varphi}_*, \Pi] \leq R_L[\hat{\varphi}, \Pi]$ für alle $\hat{\varphi} \in \mathcal{D}$ gilt. Offensichtlich beeinflusst die Verlustfunktion und die a priori Verteilung direkt die Struktur einer Bayesschen Schätzfunktion. Insbesondere lässt sich das Bayessche Risiko in der Form

$$\begin{aligned} R_L[\hat{\varphi}, \Pi] &= \int_{\Theta} R_L(\hat{\varphi}, \varphi(\theta)) \Pi(d\theta) \\ &= \int_{\Theta} \int_{\mathfrak{X}} L(\hat{\varphi}(x), \varphi(\theta)) P_{\theta}(dx) \Pi(d\theta) \\ &= \int_{\mathfrak{X}} \int_{\Theta} L(\hat{\varphi}(x), \varphi(\theta)) P^{\vartheta | \mathbf{X}=x}(d\theta) P^{\mathbf{X}}(dx) \end{aligned}$$

darstellen. Der Ausdruck

$$R_L[\hat{\varphi}(x), P^{\vartheta | \mathbf{X}=x}] := \int_{\Theta} L(\hat{\varphi}(x), \varphi(\theta)) P^{\vartheta | \mathbf{X}=x}(d\theta) \quad (1.3)$$

wird als a posteriori Risiko bezeichnet. Die Minimierung des Bayesschen Risikos kann somit auf die Minimierung des a posteriori Risikos für festes $x \in \mathfrak{X}$ zurückgeführt werden. Insbesondere wäre die Minimierung des a posteriori Risikos $R_L[\gamma, P^{\vartheta | \mathbf{X}=x}]$ für alle $\gamma \in \Gamma$ ein Kriterium für die Wahl einer

Bayesschen Schätzung $\hat{\varphi}_*(x) \in \Gamma$ für eine Beobachtung $x \in \mathcal{X}$. Wählen wir nach diesem Kriterium für jedes $x \in \mathcal{X}$ eine Bayessche Schätzung $\hat{\varphi}_*(x)$ und ist die so definierte Abbildung $\hat{\varphi}_* : \mathcal{X} \rightarrow \Gamma$ messbar, so ist $\hat{\varphi}_*$ eine Bayessche Schätzfunktion bzgl. des Bayesschen Risikos $R_L[\hat{\varphi}, \Pi]$ in der Klasse aller Schätzfunktionen $\hat{\varphi}$ für den abgeleiteten Parameter φ . Natürlicher Weise hat die a priori Verteilung im Allgemeinen einen nicht unwesentlichen Einfluss auf die Bayessche Schätzfunktion. Die a priori Verteilung kann als eine Gewichtung der Parameter bzgl. einer Vorinformation über dem Parameterraum aufgefasst werden. Diese Abhängigkeit ist auch ein zentraler Kritikpunkt am Bayesschen Ansatz, da in vielen Fällen keine Vorinformation vorhanden ist. Trotz dieser Einschränkung kann in einer Vielzahl von Situationen unter Regularitätsbedingungen nachgewiesen werden, dass die resultierenden Bayesschen Schätzfunktionen eine gute Alternative zu Schätzfunktionen des frequentistischen Ansatzes darstellen. Für eine ausführliche Diskussion sei auf die zahlreiche Literatur zu diesem Thema verwiesen (z.B. DeGroot (1970), Ibragimov und Khas'minskij (1981), Hartigan (1983), Box und Tiao (1992) oder Bernardo und Smith (1994)). \square

Beispiel 1.3 Es sei die Menge Γ der abgeleiteten Parameter eine Teilmenge des \mathbb{R}^k und mit der Borel- σ -Algebra \mathcal{L}^k versehen. Π sei eine a priori Verteilung auf Γ mit $\int \|\varphi(\theta)\|^2 \Pi(d\theta) < \infty$ und für die Verlustfunktion gelte

$$L_Q(\gamma_1, \gamma_2) = \|\gamma_1 - \gamma_2\|^2,$$

wobei $\|\cdot\|$ die euklidische Norm im \mathbb{R}^k ist. Dann minimiert der a posteriori Erwartungswert

$$\hat{\varphi}(x) := \int_{\Theta} \varphi(\theta) \Pi_x(d\theta)$$

das Bayessche Risiko $R_{L_Q}[\gamma, \Pi]$ für alle $\gamma \in \Gamma$ (vgl. Anderson (1984), S.83). Weiterhin ist der a posteriori Erwartungswert eine messbare Funktion von \mathcal{X} nach Γ (vgl. Bellach et al. (1978), S.101), so dass $\hat{\varphi}$ eine Bayessche Schätzfunktion für φ in der Klasse aller Schätzfunktionen ist. \square

Kapitel 2

Dominierte Verteilungsfamilie

2.1 Modellannahmen

Es sei \mathbf{X} eine das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$ beschreibende Zufallsvariable. Wir nehmen weiterhin an, dass ein σ -endliches Maß μ auf \mathfrak{B} existiere, so dass für alle $\theta \in \Theta$ das Maß P_θ absolut stetig bezüglich des Maßes μ sei, d.h. $\mu(B) = 0$ impliziert $P_\theta(B) = 0$. Die Verteilungsfamilie $\{P_\theta \mid \theta \in \Theta\}$ heißt dann dominiert bzgl. μ . In diesem Fall liefert der Satz von Radon-Nikodym die Existenz einer Dichte $p(x, \theta)$ bzgl. μ , d.h. $p(x, \theta) = (dP_\theta/d\mu)(x)$ und für alle $B \in \mathfrak{B}$ gilt

$$P_\theta(B) = \int_B p(x, \theta) \mu(dx).$$

Des weiteren nehmen wir an, dass $p(x, \theta)$ eine messbare Funktion des Variablenpaares (x, θ) ist, d.h. p ist eine messbare Abbildung von $(\mathcal{X} \times \Theta, \mathfrak{B} \otimes \mathfrak{C})$ auf die reelle Achse $(\mathbb{R}, \mathfrak{L})$.

Bemerkung 2.1 Falls die σ -Algebra \mathfrak{B} abzählbar erzeugt ist, kann die Dichte $p(x, \theta)$ immer messbar in (x, θ) gewählt werden (vgl. [Strasser \(1985\)](#), S. 17, Lemma 4.1 und S. 18, Lemma 4.6). Weiterhin ist die Wahl des dominierenden Maßes für die statistische Auswertung unerheblich. Sind $p_1(x, \theta)$ und $p_2(x, \theta)$ Dichten bzgl. zwei verschiedener dominierender Maße μ_1 und μ_2 , dann unterscheiden sich p_1 und p_2 nur durch einen nicht von θ abhängenden multiplikativen Faktor. Es gilt für $[\mu_1 + \mu_2]$ -f.a. $x \in \mathcal{X}$

$$\frac{dP_\theta}{d(\mu_1 + \mu_2)}(x) = p_1(x, \theta) \frac{d\mu_1}{d(\mu_1 + \mu_2)}(x) = p_2(x, \theta) \frac{d\mu_2}{d(\mu_1 + \mu_2)}(x)$$

(vgl. [Ibragimov und Khas'minskij \(1981\)](#), S. 11 und [Strasser \(1985\)](#), S. 2, Lemma 1.7). Da aber sowohl p_1 als auch p_2 nur bis auf μ_1 - bzw. μ_2 -Nullmengen eindeutig sind, ist im Allgemeinen die statistische Auswertung von der Festlegung der Radon-Nikodym-Dichten abhängig. \square

Im Folgenden werden wir zunächst ein statistisches Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$ betrachten, dessen Verteilungsfamilie $\{P_\theta \mid \theta \in \Theta\}$ durch ein σ -endliches Maß μ dominiert ist. Ziel dieser Arbeit ist es, die aus der Literatur bekannten Ergebnisse der nächsten beiden Abschnitte auf statistische Experimente, deren Verteilungsfamilie nicht dominiert sind, zu übertragen.

2.2 Maximum-Likelihood-Methode

Es sei \mathbf{X} eine das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$ beschreibende Zufallsvariable. Des weiteren gelten die Annahmen des Abschnittes 2.1. Es sei $x \in \mathcal{X}$ fixiert. Dann bezeichnen wir die Funktion

$L_x : \Theta \rightarrow [0, \infty]$ gegeben durch

$$L_x(\theta) := p(x, \theta) \quad (2.1)$$

als Likelihood-Funktion. Eine Schätzung $\hat{\theta}(x)$ für den Parameter θ wird Maximum-Likelihood-Schätzung genannt, falls gilt

$$L_x(\hat{\theta}(x)) \geq L_x(\tilde{\theta}), \quad \text{für alle } \tilde{\theta} \in \Theta. \quad (2.2)$$

Ist die Lösung der Ungleichung 2.2 nicht eindeutig, so bezeichnen wir jede Lösung als Maximum-Likelihood-Schätzung. Weiterhin heißt eine Schätzung $\hat{\varphi}(x)$ für den abgeleiteten Parameter φ Maximum-Likelihood-Schätzung, falls gilt

$$\hat{\varphi}(x) = \varphi(\hat{\theta}(x)), \quad (2.3)$$

wobei $\hat{\theta}(x)$ eine Maximum-Likelihood-Schätzung für den Parameter θ ist. Betrachten wir nun die Abbildung $\hat{\theta}(x)$, die jedem x eine Maximum-Likelihood-Schätzung $\hat{\theta}(x)$ zuordnet und ist die Abbildung $\hat{\theta} : \mathcal{X} \rightarrow \Theta$ messbar, so bezeichnen wir sie als Maximum-Likelihood-Schätzfunktion.

Die Wahl der Festlegung der Radon-Nikodym Dichte $p(x, \theta)$ hat einen entscheidenden Einfluss auf die Existenz einer Maximum-Likelihood-Schätzung (vgl. Bemerkung 2.1). Das folgende Beispiel soll dies veranschaulichen (vgl. Pfanzagl (1969)).

Beispiel 2.2 Es sei $\mathcal{X} = [0, 2)$, $\mathfrak{L}_{\mathcal{X}}$ die Borel- σ -Algebra und $\lambda_{\mathcal{X}}$ die Einschränkung des Lebesgue-Maßes auf \mathcal{X} . Die Familie der Wahrscheinlichkeitsverteilungen $\{P_{\theta} \mid \theta \in [1, 2)\}$ ist durch ihre Dichten bzgl. des Lebesgue-Maßes $\lambda_{\mathcal{X}}$

$$p_{\theta}(x) := \frac{1}{\theta} \mathbb{1}_{[0, \theta]}(x), \quad \text{oder} \quad p'_{\theta}(x) := \frac{1}{\theta} \mathbb{1}_{[0, \theta)}(x), \quad x \in \mathcal{X}, \quad \theta \in [1, 2)$$

gegeben. $\mathbb{1}_B$ ist die Indikatorfunktion auf der Menge $B \in \mathfrak{B}$. Offensichtlich sind p_{θ} und p'_{θ} Radon-Nikodym Dichten von P_{θ} bzgl. des Lebesgue-Maßes $\lambda_{\mathcal{X}}$ für alle $\theta \in [1, 2)$. Betrachten wir die Dichten p_{θ} , $\theta \in [1, 2)$, so ist $\hat{\theta}(x) = \max(1, x)$ eine Maximum-Likelihood-Schätzung für eine Beobachtung $\mathbf{X} = x$. Dagegen existiert für die Dichten p'_{θ} , $\theta \in [1, 2)$ keine Maximum-Likelihood-Schätzung. \square

Damit ist die Definition einer Maximum-Likelihood-Schätzung bzgl. einer beliebigen Version der Radon-Nikodym-Dichte nicht sinnvoll, sondern es werden Regularitätsbedingungen für die Dichten angenommen. Nehmen wir zum Beispiel an, dass Θ eine kompakte Teilmenge des euklidischen Raumes \mathbb{R}^k ist und die Abbildung $\theta \rightarrow p_{\theta}(x)$ stetig bzgl. der gewöhnlichen Topologie auf dem euklidischen Raum Θ für alle $x \in \mathcal{X}$, so ist die Existenz einer Maximum-Likelihood-Schätzfunktion gesichert (vgl. Schmetterer (1966), S.375, Lemma 3.3). Allgemeinere Regularitätsbedingungen für die Existenz einer Maximum-Likelihood-Schätzfunktion werden zum Beispiel in Pfanzagl (1969) angegeben.

2.3 Grenzwert einer Folge Bayesscher Schätzungen

2.3.1 Modellannahmen

Der zufällige Parameter besitze die a priori Verteilung Π und unter der Bedingung $\vartheta = \theta$ beschreibe die Zufallsvariable \mathbf{X} das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_{\theta} \mid \theta \in \Theta)$ (vgl. Abschnitt 1.3). Weiterhin sei die Verteilungsfamilie $\{P_{\theta} \mid \theta \in \Theta\}$ durch ein σ -endliches Maß μ dominiert (vgl. Abschnitt 2.1) und besitze die Verteilung Π eine Dichte π bzgl. des σ -endlichen Maßes ν . Insbesondere nehmen wir an, dass der Parameterraum Θ eine Teilmenge des \mathbb{R}^k ist.

2.3.2 Bayessche Schätzung

Wir definieren eine messbare Funktion $p : \Theta \rightarrow [0, \infty]$ durch

$$p(x) := \int_{\Theta} p(x, \theta) \pi(\theta) \nu(d\theta). \quad (2.4)$$

Nehmen wir weiterhin an, dass $p(x) < \infty$ μ -f.ü. gilt. Dann ist p eine Dichte der Randverteilung $P^{\mathbf{X}}$ von \mathbf{X} bzgl. μ . Definieren wir nun für alle $x \in \mathcal{X}$ eine messbare Funktion $\pi_x : \Theta \rightarrow [0, \infty]$ durch

$$\pi_x(\theta) := \begin{cases} \frac{p(x, \theta) \pi(\theta)}{p(x)} & : 0 < p(x) < \infty \\ 0 & : \text{sonst} \end{cases}, \quad (2.5)$$

so ist π_x für $P^{\mathbf{X}}$ -f.a. $x \in \mathcal{X}$ die Dichtefunktion bzgl. ν eines Wahrscheinlichkeitsmaßes Π_x auf Θ . Insbesondere ist die Abbildung $g : \mathfrak{C} \times \mathcal{X} \rightarrow [0, 1]$ mit $g(x, C) = \Pi_x(C)$ eine reguläre bedingte Verteilung (vgl. Bellach et al. (1978), S.112). Damit entspricht Π_x einer Festlegung der a posteriori Verteilung von ϑ unter der Bedingung $\mathbf{X} = x$. Für eine Beobachtung $\mathbf{X} = x$ wählen wir als eine Bayessche Schätzung für den Parameter θ den a posteriori Erwartungswert (vgl. Beispiel 1.3)

$$\hat{\theta}(x) := \int_{\Theta} \theta \pi_x(\theta) \nu(d\theta). \quad (2.6)$$

2.3.3 Grenzwert einer Folge Bayesscher Schätzungen

Das folgende Ergebnis beruht auf einer Arbeit von Bunke et al. (1976), dieses Ergebnis bildet den Ausgangspunkt für die in dieser Arbeit vorgestellten Untersuchungen.

Wir definieren eine Folge von messbaren Funktionen $p_r : \Theta \rightarrow [0, \infty]$ für $r \in \mathbb{N}$ durch

$$p_r(x) := \int_{\Theta} [p(x, \theta)]^r \pi(\theta) \nu(d\theta). \quad (2.7)$$

Es sei $x \in \mathcal{X}$ mit $p_r(x) < \infty$ für alle $r \in \mathbb{N}$ fixiert.

Analog zu (2.5) definieren wir eine messbare Funktion $\pi_x^r : \Theta \rightarrow [0, \infty]$ durch

$$\pi_x^r(\theta) := \begin{cases} \frac{[p(x, \theta)]^r \pi(\theta)}{p_r(x)} & : 0 < p_r(x) < \infty \\ 0 & : \text{sonst} \end{cases} \quad (2.8)$$

und bezeichnen mit Π_x^r das Wahrscheinlichkeitsmaß mit der Dichtefunktion π_x^r bzgl. ν .

Annahme A.2.1

1. Es existiert eine Maximum-Likelihood-Schätzung $\hat{\theta}_M(x)$ für θ .
2. Es existiert ein $\delta_0 > 0$, so dass für alle $0 < \delta \leq \delta_0$ ein $\varepsilon > 0$ existiert mit

$$\sup_{\theta \in \overline{K}_\delta^c(\hat{\theta}_M(x))} p(x, \theta) < \inf_{\theta \in \overline{K}_\varepsilon(\hat{\theta}_M(x))} p(x, \theta) \quad \text{mit } \overline{K}_\delta^c(\hat{\theta}_M(x)) := \Theta \setminus \overline{K}_\delta(\hat{\theta}_M(x)).$$

3. Für alle $\varepsilon > 0$ ist $\Pi(\overline{K}_\varepsilon(\hat{\theta}_M(x)))$ positiv.

$\overline{K}_\varepsilon(\theta)$ ist die abgeschlossene Kugel mit Radius $\varepsilon > 0$ um $\theta \in \Theta$.

Bemerkung 2.3 Mit 1. und 2. fordern wir somit, dass die Maximum-Likelihood-Schätzung eindeutig ist. Die Annahme 2. ist erfüllt, falls die Dichtefunktionen $p(x, \theta)$ in θ stetig sind und Θ kompakt ist. Die Annahme der Kompaktheit kann dabei durch lokale Kompaktheit und die Eigenschaft

$$\lim_{\theta \rightarrow \infty} p_\theta(x) = 0$$

ersetzt werden, falls wir die Dichtefunktionen $p(x, \theta)$ zu stetigen Funktionen auf dem kompaktifizierten Raum Θ^∞ erweitern können. $\hat{\theta}_M(x)$ ist dann die eindeutige Maximum-Likelihood-Schätzung in Θ^∞ , falls o.B.d.A. eine $\theta \in \Theta$ mit $p(x, \theta) > 0$ existiert. \square

Satz 2.4 (Bunke et al. (1976))

Es gelte die Annahme A.2.1. Dann konvergiert für $r \rightarrow \infty$ die Folge der Verteilungen Π_x^r schwach gegen das Punktmaß $\delta_{\hat{\theta}_M(x)}$:

$$\Pi_x^r \xrightarrow{w} \delta_{\hat{\theta}_M(x)} \quad \text{für } r \rightarrow \infty.$$

Bemerkung 2.5 Eine Folge P_1, P_2, \dots von Wahrscheinlichkeitsmaßen auf dem messbaren Raum (Θ, \mathcal{C}) konvergiert schwach gegen das Wahrscheinlichkeitsmaß P für $n \rightarrow \infty$, wenn für alle beschränkten stetigen Funktionen $f : \Theta \rightarrow \mathbb{R}$ (vgl. Billingsley (1968), S. 11, Satz 2.1)

$$\lim_{r \rightarrow \infty} \int_{\Theta} f(\theta) P_n(d\theta) = \int_{\Theta} f(\theta) P(d\theta)$$

gilt. Wir schreiben dann abkürzend $P_n \xrightarrow{w} P$ für $n \rightarrow \infty$. \square

Weiterhin wählen wir den „a posteriori“ Erwartungswert

$$\hat{\theta}_r(x) := \int_{\Theta} \theta \Pi_x^r(d\theta)$$

als Schätzung für den Parameter θ . Dann minimiert $\hat{\theta}_r(x)$ das „a posteriori“ Risiko

$$R_{L_Q}[\hat{\theta}(x), \Pi_x^r] = \int L_Q(\hat{\theta}(x), \theta) \Pi_x^r(d\theta)$$

bzgl. der quadratischen Verlustfunktion L_Q (vgl. Beispiel 1.3), falls gilt $\int \|\theta\|^2 \Pi(d\theta) < \infty$. Im Kapitel 4 werden wir eine alternative Beschreibung der Konstruktion der Folge der Schätzungen $(\hat{\theta}_r(x))_{r \in \mathbb{N}}$ angeben und zeigen, dass die Schätzung $\hat{\theta}_r(x)$ eine Bayessche Schätzung bzgl. einer speziellen Wahl der a priori Verteilung ist (vgl. Beispiel 4.1).

Annahme A.2.2 Es existiert ein $T < \infty$, so dass gilt

1. $\lim_{r \rightarrow \infty} \sup_{\theta \in \overline{K}_T(\hat{\theta}_M(x))} \|\theta - \hat{\theta}_M(x)\| \pi_x^r(\theta) = 0.$
2. $\liminf_{r \rightarrow \infty} \int_{\overline{K}_T(\hat{\theta}_M(x))} \pi_x^r(\theta) \pi(\theta) \nu(d\theta) > 0.$

Satz 2.6 (Bunke et al. (1976))

Es gelten die Annahmen A.2.1 und A.2.2. Dann konvergiert die Folge der Bayesschen Schätzungen $\hat{\theta}_r(x)$ gegen die Maximum-Likelihood-Schätzung $\hat{\theta}_M(x)$:

$$\lim_{r \rightarrow \infty} \|\hat{\theta}_r(x) - \hat{\theta}_M(x)\| = 0.$$

Bemerkung 2.7 Unter den Regularitätsbedingungen A.2.1 und A.2.2 konvergiert die betrachtete Folge von Bayesschen Schätzungen gegen die Maximum-Likelihood-Schätzung. Der Grenzwert ist dann offensichtlich unabhängig von der Wahl der a priori Verteilung und dieser Grenzübergang somit eine Möglichkeit der Konstruktion sensibler Schätzungen in „nicht informativen“ Situationen. Außerdem kann eine Bayessche Schätzung auch definiert werden, wenn die Verteilungsfamilie $\{P_\theta \mid \theta \in \Theta\}$ nicht dominiert ist. Es stellt sich somit die Frage, ob dieser Grenzübergang eine mögliche Verallgemeinerung der Maximum-Likelihood-Methode darstellt. Im Kapitel 4 werden wir auf der Grundlage dieses Grenzüberganges den selbstinformativen Grenzwert definieren und nachweisen, dass dieses Konzept auch in Situationen in denen die Maximum-Likelihood-Schätzung nicht eindeutig ist, ein wesentlicher Bestandteil der Annahme A.2.1, sinnvolle Ergebnisse liefert. Des weiteren werden wir in den folgenden Kapiteln den selbstinformativen Grenzwert mit einer Auswahl verallgemeinerter Maximum-Likelihood-Schätzungen vergleichen. \square

Kapitel 3

Verallgemeinerungen des Maximum-Likelihood-Prinzips

Aus einer Vielzahl alternativer Ansätze einer Verallgemeinerung des Maximum-Likelihood-Prinzips auf Situationen, in denen die Verteilungsfamilie nicht dominiert ist, stellen wir zwei Konzepte vor. Der erste Ansatz geht auf eine Arbeit von [Kiefer und Wolfowitz \(1956\)](#) zurück und der zweite stammt von [Gill \(1989\)](#). Beide Konzepte bieten im Gegensatz zur „Methode der Sieves“ ([Grenander \(1981\)](#)), der „Penalized Likelihood“ ([Geman und Hwang \(1982\)](#)) oder der „Empirical Likelihood“ ([Owen \(2001\)](#)) für einen festen Stichprobenumfang und ohne Einschränkung der Verteilungsfamilie eine Verallgemeinerung des Maximum-Likelihood-Prinzips an.

Wir nehmen an, dass eine Zufallsvariable \mathbf{X} das statistische Experiment

$$(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$$

beschreibt. Die Verteilungsfamilie $\{P_\theta \mid \theta \in \Theta\}$ ist nicht als dominiert vorausgesetzt.

3.1 Ansatz von Kiefer und Wolfowitz

Existiert kein dominierendes Maß, so ist die Likelihood-Funktion und damit auch die MLS nicht definiert. Andererseits ist für zwei Wahrscheinlichkeitsverteilungen P und Q auf $(\mathcal{X}, \mathfrak{B})$ die Verteilungsfamilie $\{P, Q\}$ immer dominiert durch das Maß $[P + Q]$. Nehmen wir nun an, dass p bzw. q Versionen der Radon-Nikodym Dichte von P bzw. Q bzgl. $[P + Q]$ seien, so können wir an Hand einer Beobachtung $\mathbf{X} = x$ die Maße P und Q mit Hilfe der Werte $p(x)$ bzw. $q(x)$ ihrer Dichten bzgl. $P + Q$ an der Stelle x vergleichen. Die Wahl des dominierenden Maßes beeinflusst den Vergleich nicht (vgl. Bemerkung 2.1). Dagegen sind die Versionen der Radon-Nikodym Dichten bzgl. des dominierenden Maßes $[P + Q]$ nur bis auf $[P + Q]$ -Nullmengen eindeutig bestimmt, so dass im Allgemeinen die Wahl der Festlegung der Radon-Nikodym-Dichte den Vergleich beeinflusst. Ausführlich wird das Problem der Wahl einer Version der Radon-Nikodym Dichte in [Scholz \(1980\)](#) diskutiert.

Ferner bezeichnen wir für $\theta_1, \theta_2 \in \Theta$ eine Version der Radon-Nikodym Dichte von P_{θ_1} bzgl. $[P_{\theta_1} + P_{\theta_2}]$ mit $\frac{dP_{\theta_1}}{d[P_{\theta_1} + P_{\theta_2}]}$.

Definition 3.1 (Kiefer und Wolfowitz (1956))

Es sei eine Beobachtung $X = x$ gegeben. Dann heißt $\hat{\theta}(x) \in \Theta$ verallgemeinerte Maximum-Likelihood-Schätzung (vMLS), falls für alle $\theta \in \Theta$ gilt

$$\frac{dP_{\hat{\theta}}}{d[P_{\hat{\theta}} + P_{\theta}]}(x) \geq \frac{dP_{\theta}}{d[P_{\theta} + P_{\hat{\theta}}]}(x).$$

Bemerkung 3.2 Wir nehmen zusätzlich an, dass die Verteilungsfamilie $\{P_{\theta} \mid \theta \in \Theta\}$ durch ein σ -endliches Maß μ dominiert sei. Für alle $\theta \in \Theta$ bezeichnen wir eine Version der Radon-Nikodym-Dichten von P_{θ} bzgl. μ mit

$$p_{\theta}(x) := \frac{dP_{\theta}}{d\mu}(x), \quad \text{für } x \in \mathcal{X}. \quad (3.1)$$

Existiert weiterhin eine MLS $\hat{\theta}_M(x)$ für eine Beobachtung $\mathbf{X} = x$ und wählen wir insbesondere für alle $\theta_1, \theta_2 \in \Theta$ und alle $x \in \mathcal{X}$ die Festlegung

$$\frac{dP_{\theta_1}}{d[P_{\theta_1} + P_{\theta_2}]}(x) = \begin{cases} \frac{p_{\theta_1}(x)}{p_{\theta_1}(x) + p_{\theta_2}(x)} & : (p_{\theta_1} + p_{\theta_2})(x) > 0 \\ 0 & : (p_{\theta_1} + p_{\theta_2})(x) = 0 \end{cases}$$

der Radon-Nikodym-Dichte von P_{θ_1} bzgl. $[P_{\theta_1} + P_{\theta_2}]$, so ist $\hat{\theta}_M(x)$ auch eine verallgemeinerte Maximum-Likelihood-Schätzung nach Kiefer und Wolfowitz für diese Festlegung der Radon-Nikodym-Dichten. In diesem Sinn ist dann das Konzept von Kiefer und Wolfowitz eine Verallgemeinerung des Maximum-Likelihood-Prinzips. \square

3.2 Ansatz von Gill

Im Folgenden wird die Idee der Likelihood-Gleichungen verallgemeinert. In dem hier zugrunde liegenden Modell existieren im Allgemeinen die Likelihood-Funktionen nicht. Können wir aber Likelihood-Gleichungen für eine Auswahl von Teilmengen $\Theta_h \subset \Theta$, $h \in H$ angeben und ist $\hat{\theta}$ eine Lösung der Gleichungen für alle $h \in H$ mit $\theta \in \Theta_h$, so ist dies eine natürliche Verallgemeinerung des Prinzips der Likelihood-Gleichungen. Die Menge H kann dann interpretiert werden als Menge der zugelassenen Richtungen, in denen die Verteilung $P_{\hat{\theta}}$ mit anderen Elementen der Verteilungsfamilie verglichen wird.

Annahme A.3.1 Es existiert eine Abbildung

$$\psi : \Theta \times H \times \mathbb{R} \rightarrow \Theta, \quad (3.2)$$

so dass für alle $\theta \in \Theta$ und $h \in H$ folgende Bedingungen erfüllt sind:

1. Es gilt $\psi(\theta, h, 0) = \theta$.
2. Es existiert ein σ -endlich Maß $\mu(\theta, h)$, so dass die Verteilungsfamilie $\{P_{\psi(\theta, h, c)} \mid c \in \mathbb{R}\}$ bzgl. $\mu(\theta, h)$ dominiert ist.
3. Die Likelihood-Funktion $L_x(c, \theta, h) = \frac{dP_{\psi(\theta, h, c)}}{d\mu(\theta, h)}(x)$ ist für alle $x \in \mathcal{X}$ differenzierbar in c an der Stelle $c = 0$.

Für eine Beobachtung $\mathbf{X} = x$ betrachten wir nun die Likelihood-Gleichungen

$$U_x(\theta, h) := \frac{\partial}{\partial c} \left(\log L_x(c, \theta, h) \right) \Big|_{c=0} \quad \text{für alle } \theta \in \Theta, h \in H. \quad (3.3)$$

Definition 3.3 (Gill (1989))

Es sei eine Beobachtung $X = x$ gegeben. Dann heißt $\hat{\theta}(x) \in \Theta$ verallgemeinerte Maximum-Likelihood-Schätzung (vMLS), falls für alle $h \in H$ gilt

$$U_x(\hat{\theta}(x), h) = 0.$$

Diese Definition hängt offensichtlich von der Wahl der Abbildung ψ ab. Weiterhin liefern die Likelihood-Gleichungen in einem klassischen dominierten Verteilungsmodell nur eine notwendige aber keine hinreichende Bedingung für ein Maximum der Likelihood-Funktion. Diese Einschränkung überträgt sich natürlich auch auf diese Verallgemeinerung der Likelihood-Gleichungen.

Bemerkung 3.4 Nehmen wir erneut an, dass die Verteilungsfamilie $\{P_\theta \mid \theta \in \Theta\}$ durch ein σ -endliches Maß μ dominiert ist (vgl. Bemerkung 3.2) und sind insbesondere für alle $x \in \mathcal{X}$, $\theta \in \Theta$, $h \in H$ und $c_1, c_2 \in \mathbb{R}$ die Aussagen

$$L_x(c_1, \theta, h) \geq L_x(c_2, \theta, h) \quad \text{und} \quad p_{\psi(\theta, h, c_1)}(x) \geq p_{\psi(\theta, h, c_2)}(x) \quad (3.4)$$

für die gewählte Festlegung p_θ der Radon-Nikodym-Dichte von P_θ bzgl. μ äquivalent. Dann ist für eine Beobachtung $\mathbf{X} = x$ eine MLS $\hat{\theta}_M(x)$ auch eine vMLS nach Gill, da für alle $h \in H$ die Likelihood-Funktion $L_x(c, \hat{\theta}_M(x), h)$ ein Maximum an der Stelle $c = 0$ besitzt. \square

3.3 Vergleich der Ansätze

Es sei \mathbf{X} eine das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$ beschreibende Zufallsvariable. Ferner sei eine Abbildung ψ (vgl. Annahme A.3.1) gegeben und es existiere für die Beobachtung $\mathbf{X} = x$ eine vMLS $\hat{\theta}(x)$ nach Kiefer und Wolfowitz. Dann erfüllen die gewählten Festlegungen $\frac{dP_\theta}{d[P_\theta + P_{\hat{\theta}}]}$ der Radon-Nikodym-Dichten für alle $h \in H$ und $c \in \mathbb{R}$ die folgende Ungleichung

$$\frac{dP_{\psi(\hat{\theta}(x), h, 0)}}{d[P_{\psi(\hat{\theta}(x), h, 0)} + P_{\psi(\hat{\theta}(x), h, c)}]}(x) \geq \frac{dP_{\psi(\hat{\theta}(x), h, c)}}{d[P_{\psi(\hat{\theta}(x), h, c)} + P_{\psi(\hat{\theta}(x), h, 0)}]}(x). \quad (3.5)$$

Nehmen wir weiterhin an, für alle $h \in H$ und $c \in \mathbb{R}$ sei die Likelihood-Funktion $L_x(c, \hat{\theta}(x), h)$ durch

$$L_x(c, \hat{\theta}(x), h) = \frac{dP_{\psi(\hat{\theta}(x), h, c)}}{d[P_{\psi(\hat{\theta}(x), h, c)} + P_{\psi(\hat{\theta}(x), h, 0)}]}(x) \frac{d[P_{\psi(\hat{\theta}(x), h, 0)} + P_{\psi(\hat{\theta}(x), h, c)}]}{d\mu(\hat{\theta}(x), h)}(x). \quad (3.6)$$

gegeben. So erfüllt auch die Likelihood-Funktion $L_x(c, \hat{\theta}(x), h)$ für alle $h \in H$ und $c \in \mathbb{R}$ die Ungleichung

$$L_x(0, \hat{\theta}(x), h) \geq L_x(c, \hat{\theta}(x), h) \quad (3.7)$$

und $\hat{\theta}(x)$ ist eine vMLS nach Gill.

Bemerkung 3.5 Die Gleichheit in (3.6) gilt nur für $\mu(\hat{\theta}(x), h)$ -f.a. $x \in \mathcal{X}$ (vgl. Bemerkung 2.1). Damit gilt (3.6) natürlich nicht für eine beliebige Version der Radon-Nikodym-Dichten $\frac{dP_\theta}{d[P_\theta + P_{\hat{\theta}}]}$ und somit ist eine vMLS nach Kiefer und Wolfowitz nicht notwendig eine vMLS nach Gill.

Die Gleichheit in (3.6) gilt insbesondere, wenn die Wahrscheinlichkeit $P_{\hat{\theta}(x)}(\{x\})$ positiv ist. Dann ist eine vMLS nach Kiefer und Wolfowitz immer auch eine vMLS nach Gill. Insbesondere ist nun die vMLS nach Kiefer und Wolfowitz nicht eindeutig bestimmt, so ist auch die vMLS nach Gill nicht eindeutig bestimmt. \square

Kapitel 4

Selbstinformativer Grenzwert

Das Bayessche Prinzip setzt die Idee um, eine vor der Entnahme einer Stichprobe vorhandene Information über das zugrunde liegende statistische Modell als eine a priori Gewichtung der Parameter aufzufassen. Dieser häufig beschriebene Vorteil ist gleichzeitig ein einschneidender Nachteil, wenn keine a priori Information vorhanden ist. In der Literatur wurden verschiedene Konzepte der Wahl einer a priori Verteilung in der Situation, dass keine a priori Information über den Parameter vorhanden ist, vorgestellt (vgl. [Hartigan \(1983\)](#) sowie [Box und Tiao \(1992\)](#)). Ein alternativer Ansatz ist die Beschreibung einer nichtinformativen Situation nach [Bunke \(2002\)](#). Die dort vorgestellte Konstruktion basiert auf der Idee, in einem gewissen Sinn infinitesimal wachsendes Gewicht der in der Beobachtung enthaltenen Information im Vergleich zur a priori angesetzten Gewichtung des Parameters zu geben. Dazu wird die iterative Prozedur betrachtet, die in jeder Iteration die a posteriori Verteilung der vorherigen Iteration als a priori Verteilung auffasst und die zugehörige a posteriori Verteilung berechnet. Konvergiert nun die Folge der zugehörigen Bayesschen Schätzungen, so wird der erhaltene Grenzwert als selbstinformativer Grenzwert bezeichnet ([Bunke \(2002\)](#)). In der Arbeit von [Bunke und Johannes \(2002\)](#) wird diese Idee weiter spezifiziert. Das dort vorgestellte Konzept wird im folgenden detailliert dargelegt.

4.1 Modellannahmen

Wir betrachten das in Abschnitt 1.3 vorgestellte Bayessche Modell. Die Zufallsvariable (\mathbf{X}, ϑ) mit der gemeinsamen Wahrscheinlichkeitsverteilung $P^{\mathbf{X}, \vartheta}$ auf der Borel- σ -Algebra $\mathfrak{B} \times \mathfrak{C}$ nehme Werte in dem polnischen Raum $\mathcal{X} \times \Theta$ an. Unter der Bedingung $\vartheta = \theta$ beschreibe die Zufallsvariable \mathbf{X} das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta | \theta \in \Theta)$ und die a priori Verteilung von ϑ sei Π .

4.2 Iterative Prozedur

Für einen festen Wert $x \in \mathcal{X}$ betrachten wir folgendes iteratives Verfahren. ϑ_1 sei eine Zufallsvariable mit Werten in (Θ, \mathfrak{C}) und der a priori Verteilung Π . Unter der Bedingung $\vartheta_1 = \theta$ beschreibe die Zufallsvariable \mathbf{X}_1 das statistische Experiment $\{\mathcal{X}, \mathfrak{B}, P_\theta | \theta \in \Theta\}$. Im ersten Schritt bestimmen wir mit Hilfe der a priori Verteilung Π von ϑ_1 und der bedingten Verteilung P_θ der Zufallsvariable \mathbf{X}_1 unter der Bedingung $\vartheta_1 = \theta$ die a posteriori Verteilung von ϑ_1 unter der Bedingung $\mathbf{X}_1 = x$. Mit Π_x^1 bezeichnen wir eine Festlegung der a posteriori Verteilung von ϑ_1 unter der Bedingung $\mathbf{X}_1 = x$. In der r -ten

Iteration ($r \geq 2$) sei ϑ_r eine Zufallsvariable mit der a priori Verteilung Π_x^{r-1} und unter der Bedingung $\vartheta_r = \theta$ beschreibe die Zufallsvariable \mathbf{X}_r das statistische Experiment $\{\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta\}$. Wir berechnen mit Hilfe der a priori Verteilung Π_x^{r-1} und der bedingten Verteilung P_θ der Zufallsvariable \mathbf{X}_r unter der Bedingung $\vartheta_r = \theta$ die a posteriori Verteilung der Zufallsvariable ϑ_r unter der Bedingung $\mathbf{X}_r = x$ und wählen eine Festlegung Π_x^r der a posteriori Verteilung.

Abschließend erhalten wir mit Hilfe der vorgestellten iterativen Prozedur eine Folge Π_x^1, Π_x^2, \dots von a posteriori Verteilungen auf (Θ, \mathfrak{C}) .

Beispiel 4.1 Wir betrachten das in Abschnitt 2.3 vorgestellte Bayessche Modell einer dominierten Verteilungsfamilie. Der zufällige Parameter ϑ besitze die a priori Verteilung Π mit der Dichte π bzgl. ν und unter der Bedingung $\vartheta = \theta$ beschreibe die Zufallsvariable \mathbf{X} das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$, wobei die Verteilungsfamilie $\{P_\theta \mid \theta \in \Theta\}$ dominiert und Θ eine Teilmenge des \mathbb{R}^k sei.

Ferner sei $x \in \mathcal{X}$ mit $p_r(x) := \int [p(x, \theta)]^r \pi(\theta) \nu(d\theta) < \infty$ für alle $r \in \mathbb{N}$ fixiert.

In der ersten Iteration sei ϑ_1 ein zufälliger Parameter mit der a priori Verteilung Π und unter der Bedingung $\vartheta_1 = \theta$ beschreibe die Zufallsvariable \mathbf{X}_1 das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$. Wir wählen nun die durch die Dichte (vgl. (2.6) im Abschnitt 2.3)

$$\pi_x^1(\theta) := \begin{cases} \frac{p(x, \theta)\pi(\theta)}{p_1(x)} & : 0 < p_1(x) < \infty \\ 0 & : \text{sonst} \end{cases} \quad (4.1)$$

bzgl. ν festgelegte Version Π_x^1 der a posteriori Verteilung von ϑ_1 unter der Bedingung $\mathbf{X}_1 = x$. In der zweiten Iteration sei ϑ_2 eine Zufallsvariable mit der a priori Verteilung Π_x^1 und unter der Bedingung $\vartheta_2 = \theta$ beschreibe \mathbf{X}_2 das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$. Analog zur ersten Iteration wählen wir die durch die Dichte

$$\pi_x^2(\theta) := \begin{cases} \frac{p(x, \theta)\pi_x^1(\theta)}{p_2(x)} & : 0 < p_2(x) < \infty \\ 0 & : \text{sonst} \end{cases} \quad (4.2)$$

bzgl. ν festgelegte Version Π_x^2 der a posteriori Verteilung von ϑ_2 unter der Bedingung $\mathbf{X}_2 = x$. Führen wir dieses iterative Verfahren fort, so erhalten wir die in Abschnitt 2.3 vorgestellte Folge von Verteilungen $(\Pi_x^r)_{r \in \mathbb{N}}$. \square

Explizite Darstellung der Festlegung der a posteriori Verteilung nach r Iterationen

Mit Hilfe des vorgestellten Algorithmus erhalten wir eine Folge Π_x^1, Π_x^2, \dots von Festlegungen von a posteriori Verteilungen, wobei die Festlegung Π_x^r von der Wahl der Festlegung der a posteriori Verteilung Π_x^j ($1 \leq j < r$) in jeder vorherigen Iteration und der a priori Verteilung Π im ersten Schritt abhängt. Im Folgenden werden wir basierend auf dem nächsten Satz eine explizite Darstellung dieser iterativen Prozedur angeben.

Annahme A.4.1 Es sind $\vartheta, \vartheta_1, \vartheta_2$ und $\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2$ Zufallsvariablen mit Werten in Θ bzw. \mathcal{X} .

1. Die a priori Verteilung von ϑ und ϑ_1 ist Π .
2. P_θ ist die bedingte Verteilung von \mathbf{Y}_1 unter der Bedingung $\vartheta_1 = \theta$ und von \mathbf{X}_1 bzw. \mathbf{X}_2 unter der Bedingung $\vartheta = \theta$. Die bedingte Verteilung des Vektors $(\mathbf{X}_1, \mathbf{X}_2)$ unter der Bedingung $\vartheta = \theta$ ist das Produktmaß $P_\theta \times P_\theta$.

3. Es ist Π_y eine Festlegung der a posteriori Verteilung $P^{\vartheta_1 | \mathbf{Y}_1=y}$ von ϑ_1 unter der Bedingung $\mathbf{Y}_1 = y$ und ξ_{x_1, x_2} eine Festlegung der a posteriori Verteilung $P^{\vartheta | \mathbf{X}_1=x_1, \mathbf{X}_2=x_2}$ von ϑ unter der Bedingung $\mathbf{X}_1 = x_1, \mathbf{X}_2 = x_2$.
4. Die bedingte Verteilung Q_{y_1} von $(\mathbf{Y}_2, \vartheta_2)$ unter der Bedingung $\vartheta_1 = \theta_1, \mathbf{Y}_1 = y_1$ ist gegeben durch
- $$Q_{y_1}(B \times C) = \int_C P_\theta(B) \Pi_{y_1}(d\theta), \quad \text{für alle } B \in \mathfrak{B}, C \in \mathfrak{C}.$$
5. Es ist Π_{y_1, y_2} eine Festlegung der bedingten Verteilung $P^{\vartheta_2 | \vartheta_1=\theta_1, \mathbf{Y}_1=y_1, \mathbf{Y}_2=y_2}$ von ϑ_2 unter der Bedingung $\vartheta_1 = \theta_1, \mathbf{Y}_1 = y_1, \mathbf{Y}_2 = y_2$.

Bemerkung 4.2 Die Festlegung der bedingten Verteilung von $(\mathbf{Y}_2, \vartheta_2)$ ist unter der Annahme A.4.1.4 unabhängig von ϑ_1 , so dass auch die Festlegung der bedingten Verteilung von ϑ_2 in der Annahme A.4.1.5 unabhängig von ϑ_1 gewählt werden kann. \square

Satz 4.3 Es seien $\vartheta, \vartheta_1, \vartheta_2$ und $\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}_1, \mathbf{Y}_2$ Zufallsvariablen mit Werten in Θ bzw. \mathcal{X} und es gelte die Annahme A.4.1. Dann ist die Randverteilung von $(\mathbf{X}_1, \mathbf{X}_2)$ und $(\mathbf{Y}_1, \mathbf{Y}_2)$ identisch und es gilt

$$P(\xi_{\mathbf{X}_1, \mathbf{X}_2}(C) = \Pi_{\mathbf{X}_1, \mathbf{X}_2}(C)) = 1, \quad \text{für alle } C \in \mathfrak{C}.$$

BEWEIS : Die Randverteilungen von $(\mathbf{X}_1, \mathbf{X}_2)$ und $(\mathbf{Y}_1, \mathbf{Y}_2)$ sind identisch, falls für jede beschränkte messbare Funktion $f : (\mathcal{X} \times \mathcal{X}, \mathfrak{B} \otimes \mathfrak{B}) \rightarrow (\mathbb{R}, \mathfrak{L})$ gilt (vgl. Billingsley (1968), S. 9, Satz 1.3)

$$\mathbb{E} f(\mathbf{X}_1, \mathbf{X}_2) = \mathbb{E} f(\mathbf{Y}_1, \mathbf{Y}_2).$$

Sei also f eine beliebige beschränkte messbare Funktion. Dann folgt

$$\begin{aligned} \mathbb{E} f(\mathbf{X}_1, \mathbf{X}_2) &= \int_{\mathcal{X} \times \mathcal{X}} f(x_1, x_2) dP^{\mathbf{X}_1, \mathbf{X}_2}(d(x_1, x_2)) \\ &= \int_{\Theta} \int_{\mathcal{X}} \left[\int_{\mathcal{X}} f(y_1, y_2) P_{\theta_2}(dy_2) \right] P_{\theta_2}(dy_1) \Pi(d\theta_2) \\ &= \int_{\mathcal{X}} \int_{\Theta} \left[\int_{\mathcal{X}} f(y_1, y_2) P_{\theta_2}(dy_2) \right] \Pi_{y_1}(d\theta_2) P^{\mathbf{Y}_1}(dy_1) \\ &= \int_{\mathcal{X}} \int_{\Theta \times \mathcal{X}} f(y_1, y_2) Q_{y_1}(d(\theta_2, y_2)) P^{\mathbf{Y}_1}(dy_1) \\ &= \int_{\mathcal{X}} \int_{\mathcal{X}} f(y_1, y_2) P^{\mathbf{Y}_2 | \mathbf{Y}_1=y_1}(dy_2) P^{\mathbf{Y}_1}(dy_1) \\ &= \mathbb{E} f(\mathbf{Y}_1, \mathbf{Y}_2). \end{aligned}$$

Folglich sind die Randverteilungen von $(\mathbf{X}_1, \mathbf{X}_2)$ und $(\mathbf{Y}_1, \mathbf{Y}_2)$ identisch, wobei die zweite und dritte Identität aus den Annahmen A.4.1.1-A.4.1.3 und die vierte und fünfte Identität aus der Annahme A.4.1.4 folgt.

Der zweite Teil des Satzes folgt aus dem Satz von Radon-Nikodym, falls Π_{y_1, y_2} eine Festlegung der

bedingten Verteilung $P^{\vartheta} | \mathbf{X}_1=y_1, \mathbf{X}_2=y_2$ ist. Insbesondere ist Π_{y_1, y_2} auf Grund der identischen Randverteilungen von $(\mathbf{X}_1, \mathbf{X}_2)$ und $(\mathbf{Y}_1, \mathbf{Y}_2)$ eine Festlegung der bedingten Verteilung $P^{\vartheta} | \mathbf{X}_1=y_1, \mathbf{X}_2=y_2$, falls Π_{y_1, y_2} die Integralgleichung

$$\int_{B_1 \times B_2} \Pi_{y_1, y_2}(C) P^{\mathbf{X}_1, \mathbf{X}_2}(d(y_1, y_2)) = P^{\vartheta, \mathbf{X}_1, \mathbf{X}_2}(C \times B_1 \times B_2)$$

für alle $B_1, B_2 \in \mathfrak{B}, C \in \mathfrak{C}$ erfüllt. Es seien also beliebige $B_1, B_2 \in \mathfrak{B}, C \in \mathfrak{C}$ gegeben, dann gilt für die Indikatorfunktion $\mathbb{I}_{C \times B_1 \times B_2}$ in Analogie zum Beweis des ersten Teiles

$$\begin{aligned} P^{\vartheta, \mathbf{X}_1, \mathbf{X}_2}(C \times B_1 \times B_2) &= \int_{\Theta} \int_{\mathfrak{X}} \int_{\mathfrak{X}} \mathbb{I}_{C \times B_1 \times B_2}(\theta, y_1, y_2) P_{\theta}(dy_2) P_{\theta}(dy_1) \Pi(d\theta) \\ &= \int_{\mathfrak{X}} \int_{\Theta} \int_{\mathfrak{X}} \mathbb{I}_{C \times B_1 \times B_2}(\theta, y_1, y_2) P_{\theta}(dy_2) \Pi_{y_1}(d\theta) P^{\mathbf{Y}_1}(dy_1) \\ &= \int_{\mathfrak{X}} \int_{\Theta \times \mathfrak{X}} \mathbb{I}_{C \times B_1 \times B_2}(\theta, y_1, y_2) Q_{y_1}(d(\theta, y_2))(d\theta) P^{\mathbf{Y}_1}(dy_1) \\ &= \int_{\mathfrak{X}} \int_{\mathfrak{X}} \int_{\Theta} \mathbb{I}_{C \times B_1 \times B_2}(\theta, y_1, y_2) \Pi_{y_1, y_2}(d\theta) P^{\mathbf{Y}_2 | \mathbf{Y}_1=y_1}(dy_2) P^{\mathbf{Y}_1}(dy_1) \\ &= \int_{B_1 \times B_2} \Pi_{y_1, y_2}(C) P^{\mathbf{X}_1, \mathbf{X}_2}(d(y_1, y_2)) \end{aligned}$$

und somit die Behauptung, wobei für die vorletzte Identität 5. der Annahme A.4.1 benutzt wurde. \square

Betrachten wir nun die in Abschnitt 4.2 vorgestellte iterative Prozedur, so können wir mit Hilfe des Satzes 4.3 eine explizite Darstellung der Festlegung Π_x^r der a posteriori Verteilung nach r Iterationen angeben.

Für $r \in \mathbb{N}$ seien der zufällige Vektor $\mathbf{X}^r = (\mathbf{X}_1, \dots, \mathbf{X}_r)$ auf $(\mathcal{X}^r, \mathfrak{B}^r)$ und die Zufallsvariable ϑ auf (Θ, \mathfrak{C}) mit der a priori Verteilung Π gegeben, wobei jede Zufallsvariable \mathbf{X}_i ($i = 1, \dots, r$) unter der Bedingung $\vartheta = \theta$ die Verteilung P_{θ} und der Vektor \mathbf{X}^r unter der Bedingung $\vartheta = \theta$ das Produktmaß $P_{\theta} \times \dots \times P_{\theta}$ besitze, d.h. unter der Bedingung $\vartheta = \theta$ seien die Zufallsvariablen $\mathbf{X}_1, \dots, \mathbf{X}_r$ unabhängig und identisch P_{θ} -verteilt. Setzen wir in jedem Schritt der iterativen Prozedur nicht den gleichen Wert $x \in \mathcal{X}$ ein, sondern wählen jeweils einen Wert $x_i \in \mathcal{X}$ ($i = 1, \dots, r$), so erhalten wir eine Folge $\Pi_{x_1}^1, \dots, \Pi_{x_r}^r$ von Festlegungen der a posteriori Verteilung. Insbesondere ist nun mit Satz 4.3 die Festlegung $\Pi_{x_r}^r$ der a posteriori Verteilung in der r -ten Iteration eine Festlegung der bedingten Verteilung $P^{\vartheta} | \mathbf{X}^r=x^r$ von ϑ unter der Bedingung $\mathbf{X}^r = x^r$, falls $\Pi_{x_r}^r$ eine Borel-messbare Funktion in $x^r = (x_1, \dots, x_r)$ ist. Für $x \in \mathcal{X}$ und $r \in \mathbb{N}$ betrachten wir nun den Spezialfall, r -mal den Wert x beobachtet zu haben. Für diese Beobachtung $(x, \dots, x) \in \mathcal{X}^r$ schreiben wir abkürzend $\mathbb{1}_r^t \otimes x$, wobei $\mathbb{1}_r$ der r -dimensionale Einsvektor und \otimes das Kroneckerprodukt ist. Wählen wir eine Festlegung der bedingten Verteilung $P^{\vartheta} | \mathbf{X}^r=\mathbb{1}_r^t \otimes x$ von ϑ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r^t \otimes x$, dann ist dies eine Festlegung der a posteriori Verteilung Π_x^r nach r Iterationen und somit eine alternative nicht iterative Beschreibung der vorgestellten Prozedur. Wir werden im Folgenden nur noch diese Beschreibung der a posteriori Verteilung nach r Iterationen benutzen.

Beispiel 4.4 Wir betrachten die im Beispiel 4.1 vorgestellte Folge $(\Pi_x^r)_{r \in \mathbb{N}}$ von iterativen Festlegungen der a posteriori Verteilungen in einem Bayesschen Modell (vgl. Abschnitt 2.3) mit dominierter Verteilungsfamilie $\{P_\theta \mid \theta \in \Theta\}$. Zusätzlich nehmen wir für $r \in \mathbb{N}$ an, dass unter der Bedingung $\vartheta = \theta$ die gemeinsame Verteilung des Vektors $\mathbf{X}^r = (\mathbf{X}_1, \dots, \mathbf{X}_r)$ das r -fache Produkt $P_\theta \times \dots \times P_\theta$ sei und definieren eine messbare Funktion $p_r : \mathcal{X}^r \rightarrow [0, \infty]$ durch

$$p_r(x^r) := \int_{\Theta} \prod_{i=1}^r p(x_i, \theta) \pi(\theta) \nu(d\theta) \quad \text{für } x^r = (x_1, \dots, x_r) \quad (4.3)$$

Unter der Annahme, dass für μ -f.a. $x^r \in \mathcal{X}^r$ gilt $p_r(x^r) < \infty$, ist p_r nun eine Dichtefunktion bzgl. μ der Randverteilung $P^{\mathbf{X}^r}$ auf $(\mathcal{X}^r, \mathfrak{B}^r)$ und für $P^{\mathbf{X}^r}$ -f.a. $x^r \in \mathcal{X}^r$ definiert

$$\pi_{x^r}(\theta) := \begin{cases} \frac{\prod_{i=1}^r p(x_i, \theta) \pi(\theta)}{p_r(x^r)} & : 0 < p_r(x^r) < \infty \\ 0 & : \text{sonst} \end{cases} \quad (4.4)$$

eine Dichtefunktion bzgl. ν einer Verteilung $\Pi_{x^r}^r$ auf (Θ, \mathfrak{C}) . $\Pi_{x^r}^r$ entspricht dann einer Festlegung der a posteriori Verteilung von ϑ unter der Bedingung $\mathbf{X}^r = x^r$.

Nehmen wir nun weiterhin an, dass $x \in \mathcal{X}$ mit $p_r(\mathbb{1}_r^t \otimes x) < \infty$ für alle $r \in \mathbb{N}$ fixiert sei. So entspricht die Festlegung $\Pi_{\mathbb{1}_r^t \otimes x}^r$ der a posteriori Verteilung von ϑ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r^t \otimes x$ der in Beispiel 4.1 vorgestellten r -ten iterativen Festlegung Π_x^r der a posteriori Verteilung. \square

Eindeutige Festlegung der a posteriori Verteilung nach r Iterationen

Eine Festlegung der bedingten Verteilung $P^\vartheta \mid \mathbf{X}^r = x^r$ von ϑ unter der Bedingung $\mathbf{X}^r = x^r$ ist im Allgemeinen nicht eindeutig, da mit Ausnahme von einigen Spezialfällen die Menge

$$B^r := \{\mathbb{1}_r^t \otimes x \mid x \in \mathcal{X}\}$$

eine Nullmenge bzgl. der Randverteilung von \mathbf{X}^r ist. Wir werden im Folgenden einen Grenzübergang für $r \rightarrow \infty$ betrachten und sind somit an einer eindeutigen Festlegung der a posteriori Verteilung Π_x^r nach r Iterationen für alle $x \in \mathcal{X}$ interessiert. Eine Möglichkeit ist die Beschränkung auf eine Festlegung der bedingten Verteilung $P^\vartheta \mid \mathbf{X}^r = x^r$, die stetig in $x^r \in \mathcal{X}^r$ ist, so dass auch die bedingte Verteilung $P^\vartheta \mid \mathbf{X}^r = \mathbb{1}_r^t \otimes x$ eindeutig festgelegt ist.

Annahme A.4.2 Die Zufallsvariable \mathbf{X} nimmt Werte im euklidischen Raum $\mathcal{X} = \mathbb{R}^p$ an. Die Verteilung von \mathbf{X} lässt sich als endliche Mischung

$$P^{\mathbf{X}} = \sum_{t \in T} k_t Q_t, \quad |T| < \infty. \quad (4.5)$$

von Verteilungen Q_t mit positiven Gewichten k_t darstellen. Die Verteilungen Q_t sind konzentriert auf messbaren Teilmengen \mathcal{X}_t von \mathcal{X} , wobei $\{\mathcal{X}_t \mid t \in T\}$ eine Partition von \mathcal{X} bildet, d.h.

$$Q_t(\mathcal{X}_t) = 1, \quad \mathcal{X} = \sum_{t \in T} \mathcal{X}_t, \quad \mathcal{X}_t \cap \mathcal{X}_s = \emptyset, \quad t \neq s, \quad t, s \in T.$$

Die Mengen \mathcal{X}_t sind enthalten in linearen Unterräumen L_t des \mathbb{R}^p und jede Verteilung Q_t besitzt eine positive stetige Dichte q_t bzgl. des Lebesgue Maßes λ_t auf L_t .

Satz 4.5 Es sei (ϑ, \mathbf{X}) eine Zufallsvariable mit Werten in $\Theta \times \mathcal{X}$ und es gelte für die Randverteilung von \mathbf{X} die Annahme A.4.2. Des weiteren sei $f : \Theta \rightarrow \mathbb{R}$ eine messbare Funktion mit $\mathbb{E}|f(\vartheta)| < \infty$. Eine in x stetige Festlegung, falls sie existiert, des bedingten Erwartungswertes $\mathbb{E}(f(\vartheta) \mid \mathbf{X} = x)$ ist dann eindeutig.

BEWEIS : Es seien $g(x)$ und $h(x)$ zwei stetige Festlegungen des bedingten Erwartungswertes $\mathbb{E}(f(\boldsymbol{\vartheta}) | \mathbf{X} = x)$. Es ist zu zeigen, dass für alle $x \in \mathcal{X}$

$$g(x) = h(x)$$

gilt. Dazu sei $x \in \mathcal{X}$ nun beliebig fixiert. Mit der Annahme A.4.2 ist $\{X_t | t \in T\}$ eine Partition von \mathcal{X} , so dass genau ein $t \in T$ existiert mit $x \in X_t$. Des weiteren sei für alle $\varepsilon > 0$

$$K_t^\varepsilon(x) = \{y \in L_t | d_{\mathcal{X}}(y, x) < \varepsilon\} \quad (4.6)$$

die offene Kugel im linearen Unterraum L_t mit Radius ε um x . Auf der Borel- σ -Algebra \mathfrak{B} über \mathcal{X} definieren wir zwei Maße μ und ν bzgl. der Randverteilung Q von \mathbf{X} durch

$$\mu(B) = \int_B g(x)Q(dx), \quad \nu(B) = \int_B h(x)Q(dx), \quad \text{für alle } B \in \mathfrak{B}. \quad (4.7)$$

Da g und h Festlegungen des bedingten Erwartungswertes sind, sind die Maße μ und ν identisch und es gilt speziell für die messbare Kugel $K_t^\varepsilon(x)$

$$\mu(K_t^\varepsilon(x)) = \nu(K_t^\varepsilon(x)) \quad (4.8)$$

für alle $\varepsilon > 0$. Weiterhin gilt mit Annahme A.4.2 für alle $\varepsilon > 0$

$$\mu(K_t^\varepsilon(x)) = k_t \int_{K_t^\varepsilon(x)} g(y)q_t(y)\lambda_t(dy) = k_t \int_{K_t^\varepsilon(x)} h(y)q_t(y)\lambda_t(dy) = \nu(K_t^\varepsilon(x)), \quad (4.9)$$

wobei $q_t(y)$ stetig in y und λ_t das Lebesgue-Maß auf dem linearen Unterraum L_t ist. Die Stetigkeit von g und h in x liefert somit

$$k_t g(x)q_t(x) = \lim_{\varepsilon \downarrow 0} \frac{\mu(K_t^\varepsilon(x))}{\lambda_t(K_t^\varepsilon(x))} = \lim_{\varepsilon \downarrow 0} \frac{\nu(K_t^\varepsilon(x))}{\lambda_t(K_t^\varepsilon(x))} = k_t h(x)q_t(x). \quad (4.10)$$

Unter der Annahme A.4.2 sind nun k_t und $q_t(x)$ positiv und es folgt die Behauptung. \square

Bemerkung 4.6 Insbesondere ist Θ eine Teilmenge des \mathbb{R}^k und existiert eine in x stetige Festlegung des bedingten Erwartungswertes $\mathbb{E}[\boldsymbol{\vartheta} | \mathbf{X} = x]$, so ist diese unter den Annahmen des Satzes 4.5 eindeutig bestimmt. Ferner bezeichnen wir eine Festlegung $P^{\boldsymbol{\vartheta} | \mathbf{X}=x}$ der bedingten Verteilung von $\boldsymbol{\vartheta}$ unter der Bedingung $\mathbf{X} = x$ als stetig in x , falls für alle beschränkten stetigen Funktionen $f : \Theta \rightarrow \mathbb{R}$ der bedingte Erwartungswert

$$\mathbb{E}[f(\boldsymbol{\vartheta}) | \mathbf{X} = x] = \int_{\Theta} f(\boldsymbol{\vartheta})P^{\boldsymbol{\vartheta} | \mathbf{X}=x}(d\boldsymbol{\vartheta}) \quad (4.11)$$

eine stetige Funktion in x ist. Existiert nun eine in x stetige Festlegung $P^{\boldsymbol{\vartheta} | \mathbf{X}=x}$ und ist die Annahme A.4.2 erfüllt, so ist diese Festlegung eindeutig. Dazu betrachten wir zwei in x stetige Festlegungen $P_1^{\boldsymbol{\vartheta} | \mathbf{X}=x}$ und $P_2^{\boldsymbol{\vartheta} | \mathbf{X}=x}$. Dann genügt es zu zeigen, dass für alle beschränkten stetigen Funktionen $f : \Theta \rightarrow \mathbb{R}$ (vgl. Billingsley (1968), S. 9, Satz 1.3)

$$h_f(x) := \int_{\Theta} f(\boldsymbol{\vartheta})P_1^{\boldsymbol{\vartheta} | \mathbf{X}=x}(d\boldsymbol{\vartheta}) = \int_{\Theta} f(\boldsymbol{\vartheta})P_2^{\boldsymbol{\vartheta} | \mathbf{X}=x}(d\boldsymbol{\vartheta}) =: g_f(x)$$

für alle $x \in \mathcal{X}$ gilt. Da die Funktionen $h_f(x)$ und $g_f(x)$ nach Voraussetzung stetig in x sind, folgt $h_f(x) = g_f(x)$ für alle x wegen Satz 4.5 und somit die Behauptung. \square

4.3 Definition des selbstinformativen Grenzwertes

Im Folgenden wählen wir den a posteriori Erwartungswert

$$\hat{\theta}_r(\mathbb{1}_r^t \otimes x) := \int_{\Theta} \theta P^{\vartheta} | \mathbf{X}^r = \mathbb{1}_r^t \otimes x (d\theta) \quad \text{für } r \in \mathbb{N},$$

falls dieser existiert, als Bayessche Schätzung und untersuchen für die Folge $\hat{\theta}_1(x), \hat{\theta}_2(\mathbb{1}_2^t \otimes x), \dots$ der Bayesschen Schätzungen und die Folge $P^{\vartheta} | \mathbf{X}^1 = x, P^{\vartheta} | \mathbf{X}^2 = \mathbb{1}_2^t \otimes x, \dots$ der Festlegungen der a posteriori Verteilungen den Grenzübergang für $r \rightarrow \infty$. Dabei möchten wir nicht nur den Fall der Existenz eines Grenzwertes betrachten sondern auch das Verhalten von nicht konvergenten oder fluktuierenden Folgen beschreiben.

Definition 4.7 (Selbstinformativer Grenzwert)

Existiert für ein $x \in \mathcal{X}$ der Grenzwert

$$\lim_{r \rightarrow \infty} \hat{\theta}_r(\mathbb{1}_r^t \otimes x) = \theta_x \in \Theta,$$

so heißt θ_x selbstinformativer Grenzwert.

Definition 4.8 (Schwacher selbstinformativer a posteriori Träger)

Eine Teilmenge Θ_x von Θ heißt schwacher selbstinformativer a posteriori Träger, falls für alle Umgebungen U von Θ_x (offene Mengen U mit $\Theta_x \subset U$)

$$\lim_{r \rightarrow \infty} P^{\vartheta} | \mathbf{X}^r = \mathbb{1}_r^t \otimes x (U) = 1$$

gilt.

Wir bezeichnen für alle $T \subset \Theta$ den Abstand von $\theta \in \Theta$ zu T mit

$$d_{\Theta}[\theta, T] = \inf_{\underline{\theta} \in T} d_{\Theta}(\underline{\theta}, \theta),$$

dabei ist d_{Θ} die Metrik auf Θ .

Definition 4.9 (Selbstinformativ Grenzmenge)

Eine Teilmenge Θ_x von Θ heißt selbstinformativ Grenzmenge, falls

$$\lim_{r \rightarrow \infty} d_{\Theta}[\hat{\theta}_r(x), \Theta_x] = 0$$

gilt.

Definition 4.10 (Selbstinformativer a posteriori Träger q -ter Ordnung)

Eine Teilmenge Θ_x von Θ heißt selbstinformativer a posteriori Träger q -ter Ordnung, falls

$$\lim_{r \rightarrow \infty} \int_{\Theta} d_{\Theta}[\theta, \Theta_x]^q P^{\vartheta} | \mathbf{X}^r = \mathbb{1}_r^t \otimes x (d\theta) = 0$$

gilt.

Bemerkung 4.11 Insbesondere ist eine einelementige Menge $\Theta_x = \{\theta_x\}$ eine selbstinformativ Grenzmengung, so ist θ_x ein selbstinformativ Grenzwert. Weiterhin ist offensichtlich die Menge Θ immer ein schwacher a posteriori Träger sowie ein selbstinformativ a posteriori Träger beliebiger Ordnung. Nehmen wir andererseits an, dass eine abgeschlossene Teilmenge $\Theta_x \neq \Theta$ von Θ ein selbstinformativ a posteriori Träger q -ter Ordnung ($q \in \mathbb{N}$) sei. Für jede offene Menge $U \neq \Theta$ mit $\Theta_x \subset U$ ist dann $\delta_{U^c} := \inf_{\theta \in U^c} d_{\Theta}(\theta, \Theta_x)$ positiv und

$$[\delta_{U^c}]^q P^{\vartheta} | \mathbf{X}^r = \mathbb{1}_r^t \otimes x (U^c) \leq \int_{\Theta} [d_{\Theta}(\theta, \Theta_x)]^q P^{\vartheta} | \mathbf{X}^r = \mathbb{1}_r^t \otimes x (d\theta).$$

Somit ist Θ_x auch ein schwacher a posteriori Träger. Im Gegensatz dazu sei $\Theta_x \neq \Theta$ nun offen, $\underline{\theta}$ ein Randpunkt von Θ_x und die Folge der a posteriori Verteilungen für alle $r \in \mathbb{N}$ das Punktmaß an der Stelle $\underline{\theta}$. Dann ist Θ_x offensichtlich ein selbstinformativ a posteriori Träger q -ter Ordnung und kein schwacher a posteriori Träger. \square

4.4 Dominierte Verteilungsfamilie

Wir untersuchen nun den Spezialfall des vorgestellten Bayesschen Modells (Abschnitt 4.1), in dem die Verteilungsfamilie $\{P_{\theta} | \theta \in \Theta\}$ durch ein σ -endliches Maß μ dominiert ist.

Der zufällige Parameter ϑ besitze die a priori Verteilung Π und unter der Bedingung $\vartheta = \theta$ beschreibe die Zufallsvariable \mathbf{X} das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_{\theta} | \theta \in \Theta)$. Ferner sei $p(x, \theta)$ eine Dichte von P_{θ} bzgl. μ .

Beispiel 4.12 Es sei zusätzlich Θ eine Teilmenge des \mathbb{R}^k und π eine Dichte der a priori Verteilung Π bzgl. ν (vgl. Kapitel 2). Insbesondere sei $x \in \mathcal{X}$ mit

$$0 < \int_{\Theta} [p(x, \underline{\theta})]^r \pi(\underline{\theta}) \nu(d\underline{\theta}) < \infty \quad \text{für alle } r \in \mathbb{N}$$

fixiert. Dann ist durch die Dichte

$$\pi_r(\theta, \mathbb{1}_r^t \otimes x) = \frac{[p(x, \theta)]^r \pi(\theta)}{\int_{\Theta} [p(x, \underline{\theta})]^r \pi(\underline{\theta}) \nu(d\underline{\theta})}$$

bzgl. ν eine Festlegung $\Pi_{\mathbb{1}_r^t \otimes x}^r$ der a posteriori Verteilung von ϑ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r^t \otimes x$ gegeben (vgl. Beispiel 4.4). Nehmen wir weiterhin an, dass eine eindeutig festgelegte Maximum-Likelihood-Schätzung $\hat{\theta}_M(x)$ existiere und die Annahmen A.2.1 und A.2.2 des Abschnittes 2.3 erfüllt seien. Dann ist $\{\hat{\theta}_M(x)\}$ ein schwacher selbstinformativ a posteriori Träger (vgl. Satz 2.4). Wählen wir die Festlegung

$$\hat{\theta}_*^r(\mathbb{1}_r^t x) := \int_{\Theta} \theta \pi_r(\theta, \mathbb{1}_r^t \otimes x) \nu(d\theta)$$

des a posteriori Erwartungswert als eine Bayessche Schätzung für θ , dann ist $\{\hat{\theta}_M(x)\}$ eine selbstinformativ Grenzmengung und somit $\hat{\theta}_M(x)$ ein selbstinformativ Grenzwert (vgl. Satz 2.6). \square

Dem Beispiel 4.4 folgend sei $x \in \mathcal{X}$ mit

$$0 < \int_{\Theta} [p(x, \underline{\theta})]^r \Pi(d\underline{\theta}) < \infty \quad \text{für alle } r \in \mathbb{N} \quad (4.12)$$

fixiert und die Festlegung

$$\Pi_x^r(C) = \int_C \frac{[p(x, \theta)]^r}{\int_{\Theta} [p(x, \underline{\theta})]^r \Pi(d\underline{\theta})} \Pi(d\theta), \quad \text{für alle } C \in \mathfrak{C} \quad (4.13)$$

der a posteriori Verteilung von $\boldsymbol{\theta}$ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r^t \otimes x$ gewählt.

Annahme A.4.3 Für eine abgeschlossene Teilmenge Θ_x von Θ gilt:

1. Für jede Umgebung $U \neq \Theta$ von Θ_x existiert eine Umgebung V von Θ_x mit $V \subset U$ und

$$\sup_{\theta \in U^c} p(x, \theta) < \inf_{\theta \in V} p(x, \theta).$$

2. Für jede Umgebung U von Θ_x gilt $\Pi(U) > 0$.

Bemerkung 4.13 Nehmen wir an, dass eine Maximum-Likelihood-Schätzung $\hat{\theta}_M(x)$ existiere und bezeichnen wir die Menge aller Maximum-Likelihood-Schätzungen mit

$$\hat{\Theta}(x) = \{\theta \in \Theta \mid p_{\theta}(x) = \sup_{\underline{\theta} \in \Theta} p_{\underline{\theta}}(x)\}, \quad (4.14)$$

so erfüllt $\hat{\Theta}(x)$ offensichtlich die Annahme A.4.3.1, falls die Funktion $p(x, \theta)$ für festes $x \in \mathcal{X}$ in θ stetig ist und Θ kompakt ist. Die Annahme der Kompaktheit kann auch hier durch lokale Kompaktheit und die Eigenschaft

$$\lim_{\theta \rightarrow \infty} p_{\theta}(x) = 0$$

ersetzt werden, wenn wir die Dichtefunktionen $p(x, \theta)$ dann zu stetigen Funktionen auf dem kompaktifizierten Raum Θ^∞ erweitern können. Nehmen wir nun o.B.d.A. an, dass ein $\theta \in \Theta$ mit $p(x, \theta) > 0$ existiere, so ändert sich die Menge der Maximum-Likelihood-Schätzungen $\hat{\Theta}(x)$ beim Übergang zu dem kompaktifizierten Raum Θ^∞ nicht und ist offensichtlich kompakt. \square

Satz 4.14 Es gelte für eine Menge Θ_x die Annahme A.4.3. Dann ist Θ_x ein schwacher selbstinformativer a posteriori Träger.

BEWEIS : Es ist zu zeigen, dass für alle Umgebungen U von Θ_x

$$\lim_{r \rightarrow \infty} \Pi_x^r(U) = 1$$

gilt. O.B.d.A. sei $\Theta_x \neq \Theta$ und $U \neq \Theta$ eine beliebige Umgebung von Θ_x (vgl. Bemerkung 4.11). Dann existiert unter der Annahme A.4.3 eine Umgebung V von Θ_x mit $V \subset U$, $\Pi(V) > 0$ und $p(x, \underline{\theta}) > 0$ für alle $\underline{\theta} \in V$. Weiterhin sei $U_+^c \subset U^c$, so dass für alle $\theta \in U_+^c$ gilt $p(x, \theta) > 0$. Insbesondere folgt dann

die Behauptung direkt, falls $\Pi(U_+^c) = 0$ ist. Somit sei $\Pi(U_+^c) > 0$ und wir erhalten

$$\begin{aligned} \Pi_x^r(U^c) &= \int_{U_+^c} \frac{[p(x, \theta)]^r}{\int_{\Theta} [p(x, \theta)]^r \Pi(d\theta)} \Pi(d\theta) \\ &\leq \int_{U_+^c} \left[\int_V \left(\frac{p(x, \theta)}{p(x, \theta)} \right)^r \Pi(d\theta) \right]^{-1} \Pi(d\theta) \\ &\leq \frac{\Pi(U_+^c)}{\Pi(V)} \left[\frac{\inf_{\theta \in V} p(x, \theta)}{\sup_{\theta \in U_+^c} p(x, \theta)} \right]^{-r}. \end{aligned}$$

Damit folgt unter Berücksichtigung der Annahme A.4.3.1 die Behauptung. \square

Bemerkung 4.15 Es gelten die Annahmen der Bemerkung 4.13. Dann erfüllt die Menge der Maximum-Likelihood-Schätzungen $\hat{\Theta}(x)$ (vgl. (4.14)) die Voraussetzungen des Satzes 4.14 und ist somit ein schwacher selbstinformativer a posteriori Träger. Insbesondere verzichten wir auf die Annahme der Eindeutigkeit der Maximum-Likelihood-Schätzung und fordern nur noch ihre Existenz. \square

Annahme A.4.4 $(\Theta, \|\cdot\|_{\Theta})$ ist ein Banachraum und es gilt

$$\int_{\Theta} \|\theta\|_{\Theta}^q \Pi(d\theta) < \infty \quad (4.15)$$

für ein q in \mathbb{N} .

Satz 4.16 Es gelte für eine beschränkte Menge Θ_x die Annahme A.4.3 und A.4.4. Dann ist Θ_x ein selbstinformativer a posteriori Träger q -ter Ordnung. Insbesondere ist θ_x ein selbstinformativer Grenzwert, wenn $\Theta_x = \{\theta_x\}$ eine einelementige Teilmenge von Θ ist.

BEWEIS : Wir zeigen zuerst, dass

$$\lim_{r \rightarrow \infty} \int_{\Theta} d_{\Theta}[\theta, \Theta_x]^q \Pi_x^r(d\theta) = 0$$

gilt. Es sei $\theta_x \neq \theta$ (vgl. Bemerkung 4.11) und $\theta_0 \in \Theta_x$ fixiert. Da Θ_x beschränkt ist, existiert eine positive Konstante $K < \infty$, so dass

$$d_{\Theta}[\theta, \Theta_x]^q \leq \left(\|\theta\|_{\Theta} + \|\theta_0\|_{\Theta} + \sup_{\theta \in \Theta_x} \|\theta - \theta_0\|_{\Theta} \right)^q \leq K(\|\theta\|_{\Theta}^q + 1) \quad (4.16)$$

für alle $\theta \in \Theta$ gilt. Sei nun $\varepsilon > 0$ beliebig fixiert, dann existiert eine Umgebung $U_{\varepsilon} \neq \emptyset$ von θ_x , so dass gilt

$$\sup_{\theta \in U_{\varepsilon}} d_{\Theta}[\theta, \theta_x]^q \leq \frac{\varepsilon}{3}. \quad (4.17)$$

Unter der Annahme A.4.3 und A.4.4 existiert eine Umgebung V von Θ_x mit $V \subset U_\varepsilon$ und ein $r_\varepsilon \geq 1$, so dass für alle $r \geq r_\varepsilon$

$$\Pi_x^r(U_\varepsilon^c) \leq \left[\frac{\sup_{\theta \in U_\varepsilon^c} p(x, \theta)}{\inf_{\theta \in V} p(x, \theta)} \right]^r \frac{\Pi(U_\varepsilon^c)}{\Pi(V)} \leq \frac{\varepsilon}{3K}, \quad (4.18)$$

$$\int_{U_\varepsilon^c} \|\theta\|_\Theta^q \Pi_x^r(d\theta) \leq \left[\frac{\sup_{\theta \in U_\varepsilon^c} p(x, \theta)}{\inf_{\theta \in V} p(x, \theta)} \right]^r \frac{1}{\Pi(V)} \int_{U_\varepsilon^c} \|\theta\|_\Theta^q \Pi(d\theta) \leq \frac{\varepsilon}{3K} \quad (4.19)$$

erfüllt ist. Für alle $r \geq r_\varepsilon$ folgt nun aus (4.16)-(4.19)

$$\int_{\Theta} d_\Theta[\theta, \Theta_x]^q \Pi_x^r(d\theta) \leq \int_{U_\varepsilon} d_\Theta[\theta, \Theta_x]^q \Pi_x^r(d\theta) + K \int_{U_\varepsilon^c} \|\theta\|_\Theta^q \Pi_x^r(d\theta) + K \Pi_x^r(U_\varepsilon^c) \leq \varepsilon,$$

so dass Θ_x ein selbstinformativer a posteriori Träger q -ter Ordnung ist.

Falls $\Theta_x = \{\theta_x\}$ ist, so bleibt zu zeigen, dass θ_x ein selbstinformativer Grenzwert ist. Wir haben gezeigt, dass unter den Annahmen A.4.3 und A.4.4 $\Theta_x = \{\theta_x\}$ ein selbstinformativer a posteriori Träger q -ter Ordnung ist, d.h. es gilt

$$\lim_{r \rightarrow \infty} \int_{\Theta} \|\theta_x - \theta\|_\Theta^q \Pi_x^r(d\theta) = 0. \quad (4.20)$$

Insbesondere existiert dann ein $r_0 \geq 1$ mit

$$\int_{\Theta} \|\theta\|_\Theta \Pi_x^r(d\theta) < \infty \quad \text{für alle } r \geq r_0, \quad (4.21)$$

so dass für alle $r \geq r_0$ das Bochner Integral

$$\hat{\theta}_r(x) := \int_{\Theta} \theta \Pi_x^r(d\theta) \quad (4.22)$$

definiert und die Ungleichung (vgl. Da Prato und Zabczyk (1992), S. 20)

$$\int_{\Theta} \|\theta\|_\Theta \Pi_x^r(d\theta) \geq \left\| \int_{\Theta} \theta \Pi_x^r(d\theta) \right\|_\Theta \quad (4.23)$$

erfüllt sind. Somit erhalten wir zusammen mit der Ungleichung von Jensen für alle $r \geq r_0$

$$\int_{\Theta} \|\theta_x - \theta\|_\Theta^q \Pi_x^r(d\theta) \geq \left[\int_{\Theta} \|\theta_x - \theta\|_\Theta \Pi_x^r(d\theta) \right]^q \geq \left\| \theta_x - \int_{\Theta} \theta \Pi_x^r(d\theta) \right\|_\Theta^q. \quad (4.24)$$

Die Behauptung

$$\lim_{r \rightarrow \infty} \|\theta_x - \hat{\theta}_r(x)\|_\Theta = 0$$

folgt nun aus der Eigenschaft (4.20). \square

Bemerkung 4.17 Viele der gewöhnlich betrachteten Modelle für diskrete oder absolut stetige Zufallsvariablen \mathbf{X} mit stetiger Dichte in θ erfüllen die Annahmen A.4.3 und A.4.4. Besitzt weiterhin die a priori Verteilung ein erstes Moment, so ist die Menge der Maximum-Likelihood-Schätzungen ein selbstinformativer a posteriori Träger erster Ordnung und falls die Maximum-Likelihood-Schätzung eindeutig ist, ist sie auch der selbstinformativen Grenzwert. Damit ist offensichtlich in diesen Fällen der selbstinformativen Grenzwert unabhängig von der Wahl der a priori Verteilung, so lange diese ein erstes Moment besitzen und die Annahme 2. in A.4.3 erfüllen. Dass dies eine sinnvolle Voraussetzung für die Beschreibung einer „nicht informativen“ Situation ist, sieht man sofort, wenn man den extremen Fall eines Punktmaßes an der Stelle $\theta_0 \in \Theta$ als a priori Verteilung betrachtet. Offensichtlich ist dann unabhängig von der Beobachtung $\mathbf{X} = x$ die a posteriori Verteilung in jeder Iteration wieder das Punktmaß an der Stelle θ_0 , so dass $\{\theta_0\}$ ein selbstinformativer a posteriori Träger beliebiger Ordnung und der selbstinformativen Grenzwert ist. \square

4.5 Nicht dominierte Verteilungsfamilie

Wir kehren nun, ohne die Dominierbarkeit der Verteilungsfamilie anzunehmen, zu dem vorgestellten Bayesschen Modell (Abschnitt 4.1) zurück.

Der zufällige Parameter $\boldsymbol{\vartheta}$ besitze die a priori Verteilung $P^\boldsymbol{\vartheta}$ und unter der Bedingung $\boldsymbol{\vartheta} = \theta$ beschreibe die Zufallsvariable \mathbf{X} das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta | \theta \in \Theta)$. O.B.d.A. sei die Zufallsvariable $(\mathbf{X}, \boldsymbol{\vartheta})$ eine messbare Abbildung von dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ nach $(\mathcal{X} \times \Theta, \mathfrak{B} \times \mathfrak{C})$ (vgl. Abschnitt 1.3), wobei die Verteilung P auf \mathfrak{A} durch

$$P(\mathbf{X}^{-1}(B) \cap \boldsymbol{\vartheta}^{-1}(C)) = \int_C P_\theta(B) P^\boldsymbol{\vartheta}(d\theta) \quad \text{für alle } B \in \mathfrak{B}, C \in \mathfrak{C}$$

gegeben sei. Betrachten wir für fixiertes $C \in \mathfrak{C}$ das Maß P_C mit $P_C(B) = P(\boldsymbol{\vartheta}^{-1}(C) \cap \mathbf{X}^{-1}(B))$ für alle $B \in \mathfrak{B}$. Dann besitzt das Maß P_C eine Radon-Nikodym Dichte $p_C(x)$ bzgl. der Randverteilung $P^\mathbf{X}$ von \mathbf{X} . Unter der Annahme, dass Θ ein polnischer Raum sei, können wir nun diese Dichte $p_C(x)$ als Funktion auf $(\mathcal{X} \times \mathfrak{C})$ so wählen, dass sie die Bedingungen einer regulären bedingten Verteilung erfüllt (vgl. Ash (1972), S. 265, Satz 6.6.5 und 6.6.6). Damit ist mit dem Satz von Radon-Nikodym gesichert, dass eine reguläre bedingte Verteilung existiert, aber es wird kein konstruktives Verfahren zur Berechnung der a posteriori Verteilung geliefert. Im Folgenden wird eine aus der Literatur bekannte Möglichkeit der Bestimmung der a posteriori Verteilung vorgestellt.

Es sei $K_\varepsilon(x) := \{y \in \mathcal{X} | d_{\mathcal{X}}(x, y) < \varepsilon\}$ die offene Kugel mit Radius $\varepsilon > 0$ um $x \in \mathcal{X}$. Wir betrachten für alle $x \in \mathcal{X}$, $C \in \mathfrak{C}$ und $\varepsilon > 0$ mit $P^\mathbf{X}(K_\varepsilon(x)) > 0$ den Quotienten

$$\frac{P^{\mathbf{X}, \boldsymbol{\vartheta}}(K_\varepsilon(x) \times C)}{P^\mathbf{X}(K_\varepsilon(x))},$$

wobei wir den Quotienten gleich Null setzten, falls $K_\varepsilon(x)$ eine $P^\mathbf{X}$ -Nullmenge ist. Das folgende Lemma zeigt nun, dass für $P^\mathbf{X}$ -fast alle $x \in \mathcal{X}$ das Maß der Kugel $K_\varepsilon(x)$ bzgl. $P^\mathbf{X}$ für $\varepsilon > 0$ positiv ist.

Lemma 4.18 *Die Menge*

$$N := \{x \in \mathcal{X} | P^\mathbf{X}(K_\varepsilon(x)) = 0 \text{ für ein } \varepsilon > 0\} \tag{4.25}$$

ist Teilmenge einer $P^\mathbf{X}$ -Nullmenge.

BEWEIS : Es ist zu zeigen, dass eine Menge $B \in \mathfrak{B}$ existiert, so dass gilt $N \subset B$ und $P^{\mathbf{X}}(B) = 0$. Wir definieren eine abzählbare Überdeckung $\{N_n \mid n \in \mathbb{N}\}$ von N durch

$$N_n := \{x \in \mathcal{X} \mid P^{\mathbf{X}}(K_{\frac{1}{n}}(x)) = 0\} \quad \text{für } n \in \mathbb{N}.$$

Es sei $n \in \mathbb{N}$ fixiert. Dann existiert unter der Annahme, dass \mathcal{X} ein separabler metrischer Raum ist, eine abzählbare Überdeckung mit disjunkten Mengen U_1, U_2, \dots von N_n , so dass für alle $x \in N_n$ ein U_i existiert mit $x \in U_i \subset K_{\frac{1}{n}}(x)$ (vgl. „Paving System“ [Hahn und Rosenthal \(1948\)](#), S. 254). Somit gilt für geeignet gewählte $x_i \in N_n, i \in \mathbb{N}$:

$$N_n \subset \sum_{i \geq 1} U_i \subset \bigcup_{i \geq 1} K_{\frac{1}{n}}(x_i) =: B_n.$$

Die Behauptung folgt für $B = \bigcup_{n \geq 1} B_n \in \mathfrak{B}$. □

Satz 4.19 (Hahn und Rosenthal (1948), S. 247 ff., Satz 17.2.2, 17.2.41 und 17.2.51)

Es sei \mathcal{X} ein polnischer Raum und P, Q Wahrscheinlichkeitsmaße auf der induzierten Borel- σ -Algebra \mathfrak{B} . Dann ist die Funktion $q : \mathcal{X} \rightarrow [0, \infty]$ mit

$$q(x) = \limsup_{\varepsilon \downarrow 0} \frac{Q(K_\varepsilon(x))}{P(K_\varepsilon(x))}$$

messbar, wobei wir den Quotienten für $P(K_\varepsilon(x)) = 0$ gleich ∞ bzw. 0 setzen, falls $Q(K_\varepsilon(x))$ positiv bzw. 0 ist. Für P -fast alle $x \in \mathcal{X}$ ist $q(x)$ dann endlich und falls Q absolut stetig bzgl. P ist, gilt weiterhin

$$Q(B) = \int_B q(x) P(dx) \quad \text{für alle } B \in \mathfrak{B}.$$

Satz 4.19 sichert, dass durch

$$p_C(x) := \limsup_{\varepsilon \downarrow 0} \frac{P^{\mathbf{X}, \vartheta}(K_\varepsilon(x) \times C)}{P^{\mathbf{X}}(K_\varepsilon(x))}, \quad x \in \mathcal{X}$$

eine Festlegung der Radon-Nikodym Dichte des Maßes P_C bzgl. der Randverteilung $P^{\mathbf{X}}$ definiert wird, da für alle $B \in \mathfrak{B}$ gilt

$$\int_B p_C(x) P^{\mathbf{X}}(dx) = P_C(B).$$

Andererseits wird für festes $x \in \mathcal{X}$ mit $P^{\mathbf{X}}(K_\varepsilon(x)) > 0$ für alle $\varepsilon > 0$ durch

$$P_{x\varepsilon}(C) := \frac{P^{\mathbf{X}, \vartheta}(K_\varepsilon(x) \times C)}{P^{\mathbf{X}}(K_\varepsilon(x))} \quad \text{für } C \in \mathfrak{C} \tag{4.26}$$

ein Wahrscheinlichkeitsmaß auf \mathfrak{C} definiert. Ferner ist die Menge der Wahrscheinlichkeitsmaße auf einem polnischen Raum bzgl. der schwachen Konvergenz vollständig (vgl. [Prohorov \(1956\)](#)). Insbesondere konvergiert nun für $\varepsilon \downarrow 0$ die Folge von Wahrscheinlichkeitsmaßen $(P_{x\varepsilon})_\varepsilon$ schwach gegen ein Maß P_x

$$P_{x\varepsilon} \xrightarrow{w} P_x \quad \text{für } \varepsilon \downarrow 0,$$

so ist P_x auch ein Wahrscheinlichkeitsmaß auf \mathfrak{C} .

Satz 4.20 (Pfanzagl (1979))

Für $P^{\mathbf{X}}$ -f.a. $x \in \mathcal{X}$ und $\varepsilon \downarrow 0$ konvergiert die Folge der Verteilungen $(P_{x\varepsilon})_\varepsilon$ (4.26) schwach gegen eine Festlegung $P^{\vartheta \mid \mathbf{X}=x}$ der a posteriori Verteilung von ϑ unter der Bedingung $\mathbf{X} = x$:

$$P_{x\varepsilon} \xrightarrow{w} P^{\vartheta \mid \mathbf{X}=x} \quad \text{für } \varepsilon \downarrow 0.$$

Bemerkung 4.21 Wir bezeichnen eine Festlegung $P^{\vartheta|\mathbf{X}=x}$ der bedingten Verteilung von ϑ unter der Bedingung $\mathbf{X} = x$ als stetig in x , falls für alle beschränkten, stetigen Funktionen $f : \Theta \rightarrow \mathbb{R}$ der bedingte Erwartungswert

$$\mathbb{E}[f(\vartheta)|\mathbf{X} = x] = \int_{\Theta} f(\theta)P^{\vartheta|\mathbf{X}=x}(d\theta) \quad (4.27)$$

eine stetige Funktion in x ist (vgl. Bemerkung 4.6). Insbesondere ist dann die Festlegung $P^{\vartheta|\mathbf{X}=x}$ eindeutig, wenn für die Randverteilung $P^{\mathbf{X}}$ von \mathbf{X} die Annahme A.4.2 erfüllt ist (vgl. Bemerkung 4.6). \square

Annahme A.4.5 Es existiert eine eindeutig bestimmte in x stetige Festlegung $P^{\vartheta|\mathbf{X}=x}$ der a posteriori Verteilung von ϑ unter der Bedingung $\mathbf{X} = x$.

Lemma 4.22 Es gelte die Annahme A.4.5. Dann konvergiert für alle $x \in \mathcal{X} \setminus N$ (4.25) und $\varepsilon \downarrow 0$ die Folge der Verteilungen $(P_{x\varepsilon})_{\varepsilon}$ (4.26) schwach gegen die in x stetige Festlegung $P^{\vartheta|\mathbf{X}=x}$ der a posteriori Verteilung von ϑ unter der Bedingung $\mathbf{X} = x$:

$$P_{x\varepsilon} \xrightarrow{w} P^{\vartheta|\mathbf{X}=x} \quad \text{für } \varepsilon \downarrow 0.$$

BEWEIS : Es sei $x \in \mathcal{X} \setminus N$ fixiert. Dann genügt es zu zeigen, dass für alle beschränkten stetigen Funktionen $f : \Theta \rightarrow \mathbb{R}$

$$\lim_{\varepsilon \downarrow 0} \int_{\Theta} f(\theta)P_{x\varepsilon}(d\theta) = \int_{\Theta} f(\theta)P^{\vartheta|\mathbf{X}=x}(d\theta) =: g_f(x)$$

gilt. Es sei also eine beliebige beschränkte stetige Funktion f gegeben. Da $P^{\vartheta|\mathbf{X}=x}$ eine Festlegung der bedingten Verteilung von ϑ unter der Bedingung $\mathbf{X} = x$ ist, gilt nun

$$\begin{aligned} \int_{\Theta} f(\theta)P_{x\varepsilon}(d\theta) &= \frac{1}{P^{\mathbf{X}}(K_{\varepsilon}(x))} \int_{K_{\varepsilon}(x)} \int_{\Theta} f(\theta)P^{\vartheta|\mathbf{X}=y}(d\theta)P^{\mathbf{X}}(dy) \\ &= \frac{1}{P^{\mathbf{X}}(K_{\varepsilon}(x))} \int_{K_{\varepsilon}(x)} g_f(y)P^{\mathbf{X}}(dy). \end{aligned}$$

Insbesondere ist $P^{\mathbf{X}}(K_{\varepsilon}(x))$ positiv für alle $\varepsilon > 0$ und es folgt offensichtlich

$$\inf_{y \in K_{\varepsilon}(x)} g_f(y) \leq \int_{\Theta} f(\theta)P_{x\varepsilon}(d\theta) \leq \sup_{y \in K_{\varepsilon}(x)} g_f(y).$$

Wegen Annahme A.4.5 ist $g_f(x)$ eine stetige Funktion in x und es folgt die Behauptung. \square

Schwacher selbstinformativer a posteriori Träger

Für alle $r \in \mathbb{N}$ seien der zufällige Vektor $\mathbf{X}^r = (\mathbf{X}_1, \dots, \mathbf{X}_r)$ auf $(\mathcal{X}^r, \mathfrak{B}^r)$ und die Zufallsvariable ϑ auf (Θ, \mathfrak{C}) mit der a priori Verteilung P^{ϑ} gegeben, wobei jede Zufallsvariable \mathbf{X}_i ($i = 1, \dots, r$) unter der Bedingung $\vartheta = \theta$ die Verteilung P_{θ} und der Vektor \mathbf{X}^r unter der Bedingung $\vartheta = \theta$ das r -fache Produktmaß $P_{\theta} \times \dots \times P_{\theta}$ besitze. Die gemeinsame Verteilung des Vektors $(\mathbf{X}^r, \vartheta)$ für alle $B^r = B_1 \times \dots \times B_r \in \mathfrak{B}^r$ und $C \in \mathfrak{C}$ ist durch

$$P^{\mathbf{X}^r, \vartheta}(B^r \times C) = \int_C \prod_{i=1}^r P_{\theta}(B_i)P^{\vartheta}(d\theta) \quad (4.28)$$

gegeben.

Lemma 4.23 *Es sei*

$$N_r := \{x^r \in \mathcal{X}^r \mid P^{\mathbf{X}^r}(K_\varepsilon(x_1) \times \cdots \times K_\varepsilon(x_r)) = 0 \text{ für ein } \varepsilon > 0\}. \quad (4.29)$$

Für alle $r \in \mathbb{N}$ und $x \in \mathcal{X} \setminus N$ gilt dann $\mathbb{1}_r^t \otimes x \in \mathcal{X}^r \setminus N_r$.

BEWEIS : Es sei $x \in \mathcal{X} \setminus N$ und $r \in \mathbb{N}$ beliebig fixiert. Es ist zu zeigen, dass für alle $\varepsilon > 0$ die Wahrscheinlichkeit $P^{\mathbf{X}^r}(K_\varepsilon(x) \times \cdots \times K_\varepsilon(x))$ positiv ist, wobei mit (4.28) gilt

$$P^{\mathbf{X}^r}(K_\varepsilon(x) \times \cdots \times K_\varepsilon(x)) = \int_{\Theta} [P_\theta(K_\varepsilon(x))]^r P^\theta(d\theta).$$

Nehmen wir nun an, dass $P^{\mathbf{X}^r}(K_\varepsilon(x) \times \cdots \times K_\varepsilon(x)) = 0$ für ein $\varepsilon > 0$ ist. Damit folgt $P_\theta(K_\varepsilon(x)) = 0$ für P^θ -f.a. $\theta \in \Theta$ und insbesondere $P^{\mathbf{X}}(K_\varepsilon(x)) = \int P_\theta(K_\varepsilon(x)) dP^\theta = 0$. Dies steht aber im Widerspruch zu der Annahme $x \in \mathcal{X} \setminus N$, so dass die Behauptung folgt. \square

Weiterhin ist für alle $x^r \in \mathcal{X}^r \setminus N_r$ und alle $\varepsilon > 0$

$$P_{x^r}^\varepsilon(C) := \frac{P^{\mathbf{X}^r, \theta}(K_\varepsilon(x_1) \times \cdots \times K_\varepsilon(x_r) \times C)}{P^{\mathbf{X}^r}(K_\varepsilon(x_1) \times \cdots \times K_\varepsilon(x_r))} \quad \text{für } C \in \mathfrak{C}$$

ein Wahrscheinlichkeitsmaß auf \mathfrak{C} . Existiert nun für alle $r \in \mathbb{N}$ eine eindeutige in x^r stetige Festlegung $P^\theta | \mathbf{X}^r = x^r$ der a posteriori Verteilung von θ gegeben $\mathbf{X}^r = x^r$, so gilt für alle $x^r \in \mathcal{X}^r \setminus N_r$ (vgl. Lemma 4.22)

$$P_{x^r}^\varepsilon \xrightarrow{w} P^\theta | \mathbf{X}^r = x^r \quad \text{für } \varepsilon \downarrow 0.$$

Insbesondere konvergiert mit Lemma 4.23 nun für alle $x \in \mathcal{X} \setminus N$ die Folge der Verteilungen $P_{\mathbb{1}_r^t \otimes x}^\varepsilon$ schwach gegen die in x stetige Festlegung $P^\theta | \mathbf{X}^r = \mathbb{1}_r^t \otimes x$ der a posteriori Verteilung von θ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r^t \otimes x$.

Bemerkung 4.24 Wir werden im Folgenden annehmen, dass $x \in \mathcal{X} \setminus N$ fixiert sei. Dann gilt für alle $r \in \mathbb{N}$ und

1. für alle offenen Mengen $G \in \mathfrak{C}$:

$$\liminf_{\varepsilon \downarrow 0} P_{\mathbb{1}_r^t \otimes x}^\varepsilon(G) \geq P^\theta | \mathbf{X}^r = \mathbb{1}_r^t \otimes x(G),$$

2. für alle abgeschlossenen Mengen $F \in \mathfrak{C}$

$$\limsup_{\varepsilon \downarrow 0} P_{\mathbb{1}_r^t \otimes x}^\varepsilon(F) \leq P^\theta | \mathbf{X}^r = \mathbb{1}_r^t \otimes x(F)$$

(vgl. Billingsley (1968), S. 11, Satz 2.1). \square

Annahme A.4.6 *Für eine abgeschlossene Teilmenge Θ_x von Θ gilt:*

1. Für jede Umgebung U von Θ_x ist $P^\theta(U)$ positiv.
2. Für jede Umgebung $U \neq \Theta$ von Θ_x existiert eine Umgebung V von Θ_x mit $V \subset U$ und:

(a) Für alle $\varepsilon \in (0, \varepsilon_U]$ mit $\varepsilon_U > 0$ und für alle $\theta \in U^c$ und $\underline{\theta} \in V$ gilt

$$P_\theta(K_\varepsilon(x)) \leq P_{\underline{\theta}}(K_\varepsilon(x)).$$

(b) Für eine Funktion $f : \mathcal{X} \times \Theta \times \Theta \rightarrow [0, \infty]$ mit $f(x, \theta, \underline{\theta}) := \limsup_{\varepsilon \downarrow 0} \frac{P_\theta(K_\varepsilon(x))}{P_{\underline{\theta}}(K_\varepsilon(x))}$ gilt

$$\sup_{\theta \in U^c} \sup_{\underline{\theta} \in V} f(x, \theta, \underline{\theta}) < 1.$$

Bemerkung 4.25 Betrachten wir für alle $\theta, \underline{\theta} \in \Theta$ eine Festlegung $\frac{dP_\theta}{d[P_\theta + P_{\underline{\theta}}]}$ der Radon-Nikodym Dichte von P_θ bzgl. $[P_\theta + P_{\underline{\theta}}]$, dann gilt wegen Satz 4.19 für $[P_\theta + P_{\underline{\theta}}]$ -fast alle $x \in \mathcal{X}$

$$f(x, \theta, \underline{\theta}) = \frac{\frac{dP_\theta}{d[P_\theta + P_{\underline{\theta}}]}(x)}{1 - \frac{dP_\theta}{d[P_\theta + P_{\underline{\theta}}]}(x)}. \quad (4.30)$$

Somit definiert $f(x, \theta, \underline{\theta})$ eine Festlegung der Radon-Nikodym Dichte von P_θ bzgl. $[P_\theta + P_{\underline{\theta}}]$, und die Bedingung $f(x, \theta, \underline{\theta}) \leq 1$ ist dann äquivalent zu $\frac{dP_\theta}{d[P_\theta + P_{\underline{\theta}}]}(x) \leq \frac{dP_{\underline{\theta}}}{d[P_\theta + P_{\underline{\theta}}]}(x)$ für diese spezielle Festlegung der Radon-Nikodym Dichte. Gilt weiterhin, dass ein $\hat{\theta} \in \Theta$ existiert mit $f(x, \theta, \hat{\theta}) \leq 1$ für alle $\theta \in \Theta$, so ist $\hat{\theta}$ eine verallgemeinerte Maximum-Likelihood-Schätzung nach Kiefer und Wolfowitz. \square

Beispiel 4.26 Wir betrachten den Spezialfall einer dominierten Verteilungsfamilie $\{P_\theta \mid \theta \in \Theta\}$ (vgl. Abschnitt 4.4). Zusätzlich nehmen wir an, dass die Dichten $p(x, \theta)$ gleichmäßig stetig in x für alle θ seien.

Es sei ein beliebiges $x \in \mathcal{X} \setminus N$ (4.25) fixiert. Dann folgt $\mu(K_\varepsilon(x)) > 0$ für alle $\varepsilon > 0$ und aus der Stetigkeit in x

$$p(x, \theta) = \lim_{\varepsilon \downarrow 0} \frac{P_\theta(K_\varepsilon(x))}{\mu(K_\varepsilon(x))}$$

für alle $\theta \in \Theta$. Insbesondere ist dann die Annahme A.4.3.1 des Abschnittes A.4.3

$$\sup_{\theta \in U^c} p(x, \theta) < \inf_{\underline{\theta} \in V} p(x, \underline{\theta})$$

äquivalent zu der Annahme A.4.6.2.b

$$\sup_{\theta \in U^c} \sup_{\underline{\theta} \in V} f(x, \theta, \underline{\theta}) = \sup_{\theta \in U^c} \sup_{\underline{\theta} \in V} \frac{p(x, \theta)}{p(x, \underline{\theta})} < 1.$$

Weiterhin sei o.B.d.A.

$$\delta = \frac{1}{2} \left[\sup_{\theta \in U^c} p(x, \theta) + \inf_{\underline{\theta} \in V} p(x, \underline{\theta}) \right].$$

Wegen der gleichmäßigen Stetigkeit in x für alle $\theta \in \Theta$ existiert nun ein $\varepsilon_U > 0$, so dass für alle $\theta \in U^c$, $\underline{\theta} \in V$ und für alle $y \in \mathcal{X}$ mit $d_X(x, y) \leq \varepsilon_U$ gilt $p(y, \theta) \leq \delta \leq p(y, \underline{\theta})$ und somit

$$P_\theta(K_\varepsilon(x)) \leq P_{\underline{\theta}}(K_\varepsilon(x)) \quad \text{für } \varepsilon \in (0, \varepsilon_U].$$

Abschließend erhalten wir damit für ein $x \in \mathcal{X} \setminus N$ (4.25) die Äquivalenz der Annahme A.4.3 des Abschnittes 4.4 und der Annahme A.4.6 in dem vorgestellten Bayesschen Modell mit dominierter Verteilungsfamilie und gleichmäßig stetiger Dichte in x . \square

Satz 4.27 *Es gelte für eine Menge Θ_x die Annahme A.4.6. Dann ist Θ_x ein schwacher a posteriori Träger.*

BEWEIS : Es ist zu zeigen, dass für jede Umgebung U von Θ_x gilt

$$\lim_{r \rightarrow \infty} P^\vartheta | \mathbf{X}^r = \mathbf{1}_r^t \otimes x (U) = 1. \quad (4.31)$$

O.B.d.A. sei $\Theta_x \neq \Theta$ und $U \neq \Theta$ eine Umgebung von Θ_x (vgl. Bemerkung 4.11). Dann ist U^c abgeschlossen und disjunkt zu Θ_x . Insbesondere existieren nun disjunkte offene Umgebungen W bzw. H von U^c bzw. Θ_x , da nach Voraussetzung Θ_x abgeschlossen und \mathcal{X} ein metrischer Raum ist. Unter der Annahme A.4.6.2 existiert zu der Umgebung H von Θ_x eine Umgebung V von Θ_x mit $V \subset H$. Damit ist die Annahme A.4.6.2 natürlich auch für die Umgebung W von U^c und die Umgebung V von Θ_x mit $V \subset W^c$ erfüllt. Insbesondere gilt für die offene Menge W (vgl. Bemerkung 4.24):

$$P^\vartheta | \mathbf{X}^r = \mathbf{1}_r^t \otimes x (W) \leq \liminf_{\varepsilon \downarrow 0} P_{\mathbf{1}_r^t \otimes x}^\varepsilon (W)$$

für alle $r \in \mathbb{N}$, so dass für die Umgebung U von Θ_x folgt:

$$\begin{aligned} \lim_{r \rightarrow \infty} P^\vartheta | \mathbf{X}^r = \mathbf{1}_r^t \otimes x (U) &\leq \lim_{r \rightarrow \infty} P^\vartheta | \mathbf{X}^r = \mathbf{1}_r^t \otimes x (W) \leq \lim_{r \rightarrow \infty} \liminf_{\varepsilon \downarrow 0} P_{\mathbf{1}_r^t \otimes x}^\varepsilon (W) \\ &\leq \lim_{r \rightarrow \infty} \limsup_{\varepsilon \downarrow 0} \int_W \frac{[P_\theta(K_\varepsilon(x))]^r}{\int_{\Theta} [P_\theta(K_\varepsilon(x))]^r P^\vartheta(d\theta)} P^\vartheta(d\theta). \end{aligned}$$

Nehmen wir an, es existiere ein $\varepsilon_0 > 0$ mit $P_\theta(K_\varepsilon(x)) = 0$ für alle $\varepsilon \leq \varepsilon_0$ und $\theta \in W$, so folgt offensichtlich $P_{\mathbf{1}_r^t \otimes x}^\varepsilon (W) = 0$ für alle $\varepsilon \leq \varepsilon_0$ sowie $r \in \mathbb{N}$ und damit die Behauptung (4.31). Es existiere also o.B.d.A. für alle $\varepsilon > 0$ ein $\theta \in W$ mit $P_\theta(K_\varepsilon(x)) > 0$. Insbesondere ist dann unter der Annahme A.4.6.2.a $P_\theta(K_\varepsilon(x))$ positiv für alle $\varepsilon > 0$ und $\theta \in V$. Weiterhin ist $P^\vartheta(V)$ auf Grund der Annahme A.4.6.1 positiv und die Ungleichung von Jensen liefert

$$\lim_{r \rightarrow \infty} P^\vartheta | \mathbf{X}^r = \mathbf{1}_r^t \otimes x (U) \leq \lim_{r \rightarrow \infty} \limsup_{\varepsilon \downarrow 0} \frac{1}{P^\vartheta(V)} \int_W \left[\frac{1}{P^\vartheta(V)} \int_V \frac{P_\theta(K_\varepsilon(x))}{P_\theta(K_\varepsilon(x))} P^\vartheta(d\theta) \right]^r P^\vartheta(d\theta).$$

Annahme A.4.6.2 führt dann zusammen mit dem Satz über die dominierte Konvergenz zu

$$\begin{aligned} \lim_{r \rightarrow \infty} P^\vartheta | \mathbf{X}^r = \mathbf{1}_r^t \otimes x (U) &\leq \lim_{r \rightarrow \infty} \frac{1}{P^\vartheta(V)} \int_W \left[\frac{1}{P^\vartheta(V)} \int_V \limsup_{\varepsilon \downarrow 0} \frac{P_\theta(K_\varepsilon(x))}{P_\theta(K_\varepsilon(x))} P^\vartheta(d\theta) \right]^r P^\vartheta(d\theta) \\ &\leq \frac{P^\vartheta(W)}{P^\vartheta(V)} \lim_{r \rightarrow \infty} [\sup_{\theta \in W} \sup_{\theta \in V} f(x, \theta, \theta)]^r = 0, \end{aligned}$$

so dass die Behauptung (4.31) folgt. \square

Bemerkung 4.28 Erfüllt die betrachtete Verteilungsfamilie $\{P_\theta | \theta \in \Theta\}$ die Regularitätsbedingungen A.4.6.2 und die a priori Verteilung wie bisher die Annahme A.4.6.1, so ist die Menge der verallgemeinerten Maximum-Likelihood-Schätzungen nach Kiefer und Wolfowitz ein schwacher selbstinformativer a posteriori Träger (vgl. Bemerkung 4.25). Nehmen wir weiterhin an, dass die vMLS $\hat{\theta}(x)$ nach Kiefer und Wolfowitz eindeutig festgelegt sei, dann konvergiert die Folge $P^\vartheta | \mathbf{X}^r = \mathbf{1}_r^t \otimes x$ schwach gegen das Punktmaß $\delta_{\hat{\theta}(x)}$ an der Stelle $\hat{\theta}(x)$. Unter zusätzlichen Regularitätsbedingungen konvergiert dann die Folge der zugehörigen Erwartungswerte gegen den Erwartungswert des Punktmaßes $\delta_{\hat{\theta}(x)}$ (z.B. Billingsley (1968), S. 32, Satz 5.4), so dass $\hat{\theta}(x)$ auch der selbstinformativer Grenzwert ist. Insbesondere ist offensichtlich in diesen Fällen der selbstinformativer Grenzwert unabhängig von der Wahl der a priori Verteilung. \square

4.6 Modell mit nicht identifizierbarem Parameter

Abschließend kehren wir zu dem Bayesschen Modell des Abschnittes 1.3 zurück. Auf dem messbaren Raum $(\mathcal{X}, \mathfrak{B})$ sei eine Familie $\{P_\theta \mid \theta \in \Theta\}$ von Wahrscheinlichkeitsverteilungen gegeben und das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$ durch die Zufallsvariable \mathbf{X} beschrieben. In einem Bayesschen Ansatz, d.h. wir nehmen eine a priori Verteilung auf dem messbaren Raum (Θ, \mathfrak{C}) an, ist ein möglicherweise nicht identifizierbarer Parameter θ keine Einschränkung. Insbesondere ist es ohne weiteres möglich einen Parameter $\gamma \in \Gamma$, von dem die Verteilung P_θ nicht abhängt, zu betrachten. Dazu sei auf dem messbaren Raum $(\Theta \times \Gamma, \mathfrak{C} \otimes \mathfrak{D})$ eine a priori Verteilung definiert. Wenn für eine Beobachtung $\mathbf{X} = x$ nun die a posteriori Verteilung auf $\mathfrak{C} \otimes \mathfrak{D}$ existiert, so können wir mit Hilfe dieser statistische Rückschlüsse auf θ als auch γ ziehen. Interessant ist dies zum Beispiel in einem Regressionsmodell in dem sich die Koeffizienten in der Zeit ändern, wir aber nur zu gewissen Zeitpunkten Beobachtungen erheben können. Die Zufallsvariable \mathbf{X} wird dann also nur von einigen Koeffizienten abhängen. Wir werden dieses Beispiel am Ende des Kapitels betrachten.

4.6.1 Bayessches Modell

Wir nehmen an, dass \mathcal{X} , Θ und Γ polnische Räume mit den induzierten Borel- σ -Algebren \mathfrak{B} , \mathfrak{C} und \mathfrak{D} seien. Insbesondere ist dann die induzierte Borel- σ -Algebra auf $\mathcal{X} \times \Theta \times \Gamma$ das kartesische Produkt $\mathfrak{B} \times \mathfrak{C} \times \mathfrak{D}$ (vgl. Billingsley (1968), S. 224). Weiterhin seien ϑ und γ zufällige Parameter mit Werten in (Θ, \mathfrak{C}) und (Γ, \mathfrak{D}) und unter der Bedingung $(\vartheta, \gamma) = (\theta, \gamma)$ beschreibe die Zufallsvariable \mathbf{X} das statistische Experiment $(\mathcal{X}, \mathfrak{B}, P_\theta \mid \theta \in \Theta)$.

A priori Annahmen

A priori sei auf dem messbaren Raum $(\Theta \times \Gamma, \mathfrak{C} \times \mathfrak{D})$ eine Wahrscheinlichkeitsverteilung $P^{\vartheta, \gamma}$ gegeben. Eine Wahrscheinlichkeitsverteilung P_{θ_0} aus der Familie der Wahrscheinlichkeitsverteilungen $\{P_\theta \mid \theta \in \Theta\}$ fassen wir als bedingte Verteilung von \mathbf{X} unter der Bedingung $(\vartheta, \gamma) = (\theta_0, \gamma_0)$ auf. Mittels der Familie der bedingten Verteilungen $\{P_\theta \mid \theta \in \Theta\}$ und der a priori Verteilung $P^{\vartheta, \gamma}$ definieren wir weiterhin die gemeinsame Verteilung des zufälligen Vektors $(\mathbf{X}, \vartheta, \gamma)$ auf dem messbaren Raum $(\mathcal{X} \times \Theta \times \Gamma, \mathfrak{B} \times \mathfrak{C} \times \mathfrak{D})$ durch

$$P^{\mathbf{X}, \vartheta, \gamma}(B \times C \times D) = \int_C \int_D P_\theta(B) P^{\vartheta, \gamma}(d\theta, d\gamma) \quad \text{für alle } B \in \mathfrak{B}, C \in \mathfrak{C}, D \in \mathfrak{D}.$$

Somit ist die a priori Verteilung $P^{\vartheta, \gamma}$ von (ϑ, γ) die Randverteilung bzgl. der gemeinsamen Verteilung $P^{\mathbf{X}, \vartheta, \gamma}$. Weiterhin bezeichnet $P^{\mathbf{X}, \vartheta}$ die Randverteilung von (\mathbf{X}, ϑ) und P^ϑ die Randverteilung von ϑ . Dann ist P_θ auch eine reguläre bedingte Verteilung von \mathbf{X} unter der Bedingung $\vartheta = \theta$. Des weiteren sei $P^{\gamma \mid \vartheta = \theta}$ eine reguläre bedingte Verteilung von γ unter der Bedingung $\vartheta = \theta$. Die Randverteilung $P^{\mathbf{X}}$ von \mathbf{X} bzgl. der gemeinsamen Verteilung $P^{\mathbf{X}, \vartheta, \gamma}$ ist wie bisher gegeben durch

$$P^{\mathbf{X}}(B) = \int_\Theta \int_\Gamma P_\theta(B) P^{\vartheta, \gamma}(d\theta, d\gamma) \quad \text{für alle } B \in \mathfrak{B}.$$

Außerdem seien $\mathbb{E}_{\mathbf{X} \mid \vartheta, \gamma}$, $\mathbb{E}_{\mathbf{X} \mid \gamma}$ und $\mathbb{E}_{\mathbf{X} \mid \vartheta}$ durch die regulären bedingten Verteilungen festgelegte Versionen der bedingten Erwartungswerte.

Lemma 4.29 Für alle beschränkten messbaren Funktionen $h : \mathcal{X} \rightarrow \mathbb{R}$ gilt

$$\mathbb{E}_{\mathbf{X} \mid \vartheta, \gamma}[h(\mathbf{X})] = \mathbb{E}_{\mathbf{X} \mid \vartheta}[h(\mathbf{X})].$$

BEWEIS : Es sei $h : \mathcal{X} \rightarrow \mathbb{R}$ eine beliebige beschränkte messbare Funktion, dann folgt aus

$$\int_{\Theta} \int_{\Gamma} \int_{\mathcal{X}} h(x) P_{\theta}(dx) P^{\vartheta, \gamma}(d\theta, d\gamma) = \int_{\Theta} \int_{\mathcal{X}} h(x) P_{\theta}(dx) P^{\vartheta}(d\theta)$$

die Behauptung. □

A posteriori Verteilung

Wir bezeichnen nun mit $P^{\vartheta, \gamma | \mathbf{X}=x}$, $P^{\vartheta | \mathbf{X}=x}$ sowie $P^{\gamma | \mathbf{X}=x}$ Festlegungen der a posteriori Verteilung von (ϑ, γ) , ϑ und γ unter der Bedingung $\mathbf{X} = x$. Weiterhin seien $\mathbb{E}_{\vartheta, \gamma | \mathbf{X}}$, $\mathbb{E}_{\vartheta | \mathbf{X}}$ und $\mathbb{E}_{\gamma | \mathbf{X}}$ bzgl. der a posteriori Verteilungen definierte Versionen der a posteriori Erwartungswerte.

Lemma 4.30 *Für alle beschränkten messbaren Funktionen $h : \Gamma \rightarrow \mathbb{R}$ gilt*

$$\mathbb{E}_{\gamma | \mathbf{X}}[h(\gamma)] = \mathbb{E}_{\vartheta | \mathbf{X}} \left[\mathbb{E}_{\gamma | \vartheta} [h(\gamma)] \right].$$

BEWEIS : Es seien $h : \Gamma \rightarrow \mathbb{R}$ und $g : \mathcal{X} \rightarrow \mathbb{R}$ beliebige beschränkte messbare Funktionen.

Insbesondere gilt dann wegen Lemma 4.29

$$\mathbb{E}_{\gamma} \left[h(\gamma) \mathbb{E}_{\mathbf{X} | \gamma} [g(\mathbf{X})] \right] = \mathbb{E}_{\vartheta} \left[\mathbb{E}_{\gamma | \vartheta} \left[h(\gamma) \mathbb{E}_{\mathbf{X} | \vartheta, \gamma} [g(\mathbf{X})] \right] \right] = \mathbb{E}_{\vartheta} \left[\mathbb{E}_{\gamma | \vartheta} [h(\gamma)] \mathbb{E}_{\mathbf{X} | \vartheta} [g(\mathbf{X})] \right]$$

und zusammen mit der folgenden Eigenschaft der bedingten Erwartungswerte für alle beschränkten messbaren Funktionen $f : \Theta \rightarrow \mathbb{R}$

$$\mathbb{E}_{\mathbf{X}} \left[g(\mathbf{X}) \mathbb{E}_{\vartheta | \mathbf{X}} [f(\vartheta)] \right] = \mathbb{E}_{\vartheta} \left[f(\vartheta) \mathbb{E}_{\mathbf{X} | \vartheta} [g(\mathbf{X})] \right],$$

folgt nun die Behauptung

$$\mathbb{E}_{\mathbf{X}} \left[g(\mathbf{X}) \mathbb{E}_{\gamma | \mathbf{X}} [h(\gamma)] \right] = \mathbb{E}_{\gamma} \left[h(\gamma) \mathbb{E}_{\mathbf{X} | \gamma} [g(\mathbf{X})] \right] = \mathbb{E}_{\mathbf{X}} \left[g(\mathbf{X}) \mathbb{E}_{\vartheta | \mathbf{X}} \left[\mathbb{E}_{\gamma | \vartheta} [h(\gamma)] \right] \right].$$

□

4.6.2 Selbstinformativer Grenzwert

Für alle $r \in \mathbb{N}$ sei der zufällige Vektor $\mathbf{X}^r = (\mathbf{X}_1, \dots, \mathbf{X}_r)$ mit Werten in $(\mathcal{X}^r, \mathfrak{B}^r)$ und der zufällige Vektor (ϑ, γ) mit Werten in $(\Theta \times \Gamma, \mathfrak{C} \times \mathfrak{D})$ gegeben. A priori besitze (ϑ, γ) die Verteilung $P^{\vartheta, \gamma}$ und jede Zufallsvariable \mathbf{X}_i ($i = 1, \dots, r$) unter der Bedingung $(\vartheta, \gamma) = (\theta, \gamma)$ die Verteilung P_{θ} . Weiterhin sei die Verteilung des Vektors \mathbf{X}^r unter der Bedingung $(\vartheta, \gamma) = (\theta, \gamma)$ das Produktmaß $P_{\theta} \times \dots \times P_{\theta}$. Dann ist die gemeinsame Verteilung des Vektors $(\mathbf{X}^r, \vartheta, \gamma)$ für alle $B^r = B_1 \times \dots \times B_r \in \mathfrak{B}^r$, $C \in \mathfrak{C}$ und $D \in \mathfrak{D}$ durch

$$P^{\mathbf{X}^r, \vartheta, \gamma}(B^r \times C \times D) = \int_D \int_C \prod_{i=1}^r P_{\theta}(B_i) P^{\vartheta, \gamma}(d\theta, d\gamma)$$

gegeben. Insbesondere bezeichnet $P^{\vartheta, \gamma | \mathbf{X}^r = x^r}$, $P^{\vartheta | \mathbf{X}^r = x^r}$ sowie $P^{\gamma | \mathbf{X}^r = x^r}$ eine Festlegung der a posteriori Verteilung von (ϑ, γ) , ϑ bzw. γ gegeben $\mathbf{X}^r = x^r$.

Annahme A.4.7 *Die Festlegung $P^{\gamma | \vartheta = \theta}$ der bedingten Verteilung von γ unter der Bedingung $\vartheta = \theta$ ist stetig in θ , d.h. für alle beschränkten stetigen Funktionen $h : \Gamma \rightarrow \mathbb{R}$ ist*

$$\mathbb{E}_{\gamma | \vartheta} [h(\gamma)] : \Theta \rightarrow \mathbb{R}$$

eine beschränkte stetige Funktion in ϑ (vgl. Bemerkung 4.6).

Bemerkung 4.31 Im Folgenden werden wir annehmen, dass für ein $x \in \mathcal{X}$ beliebig fixiert, $\{\hat{\theta}(x)\} \subset \Theta$ ein schwacher selbstinformativer a posteriori Träger für ϑ sei. Dann gilt (vgl. Definition 4.7) für alle Umgebungen U von $\hat{\theta}(x)$

$$P^{\vartheta} | \mathbf{X}^r = \mathbf{1}_r^t \otimes x (U) = 1.$$

Insbesondere folgt somit

$$\lim_{r \rightarrow \infty} P^{\vartheta} | \mathbf{X}^r = \mathbf{1}_r^t \otimes x (C) = \delta_{\hat{\theta}(x)}(C)$$

für alle $C \in \mathfrak{C}$ mit $\delta_{\hat{\theta}(x)}(\partial C) = 0$, wobei ∂C den Rand von C bezeichnet. Damit konvergiert die Folge der a posteriori Verteilungen schwach gegen das Punktmaß $\delta_{\hat{\theta}(x)}$ an der Stelle $\hat{\theta}(x)$ und es gilt für alle beschränkten stetigen Funktionen $f : \Theta \rightarrow \mathbb{R}$ (vgl. Billingsley (1968), S.11, Satz 2.1):

$$\lim_{r \rightarrow \infty} \mathbb{E}_{\vartheta} | \mathbf{X}^r = \mathbf{1}_r^t \otimes x [f(\vartheta)] = f(\hat{\theta}(x)).$$

□

Satz 4.32 Es sei $\{\hat{\theta}(x)\}$ ein schwacher a posteriori Träger für ϑ und es gelte die Annahme A.4.7 für die a priori Verteilung $P^{\vartheta, \gamma}$. Dann konvergiert die Folge der a posteriori Verteilungen $P^{\gamma} | \mathbf{X}^r = \mathbf{1}_r^t \otimes x$ schwach gegen die bedingte Verteilung $P^{\gamma} | \vartheta = \hat{\theta}(x)$ von γ unter der Bedingung $\vartheta = \hat{\theta}(x)$

$$P^{\gamma} | \mathbf{X}^r = \mathbf{1}_r^t \otimes x \xrightarrow{W} P^{\gamma} | \vartheta = \hat{\theta}(x) \quad \text{für } r \rightarrow \infty.$$

BEWEIS : Es ist zu zeigen, dass für alle beschränkten stetigen Funktionen $h : \Gamma \rightarrow \mathbb{R}$ gilt

$$\mathbb{E}_{\gamma} | \vartheta = \hat{\theta}(x) [h(\gamma)] = \lim_{r \rightarrow \infty} \mathbb{E}_{\gamma} | \mathbf{X}^r = \mathbf{1}_r \otimes x [h(\gamma)].$$

Es sei also $h : \Gamma \rightarrow \mathbb{R}$ eine beliebige beschränkte stetige Funktion. Dann gilt wegen Lemma 4.30

$$\lim_{r \rightarrow \infty} \mathbb{E}_{\gamma} | \mathbf{X}^r = \mathbf{1}_r \otimes x [h(\gamma)] = \lim_{r \rightarrow \infty} \mathbb{E}_{\vartheta} | \mathbf{X}^r = \mathbf{1}_r^t \otimes x [\mathbb{E}_{\gamma} | \vartheta [h(\gamma)]]$$

Nach Voraussetzung ist $\{\hat{\theta}(x)\}$ ein schwacher a posteriori Träger (vgl. Bemerkung 4.31) und $\mathbb{E}_{\gamma} | \vartheta [h(\gamma)] : \Theta \rightarrow \mathbb{R}$ eine beschränkte stetige Funktion (Annahme A.4.7), so dass die Behauptung folgt. □

4.6.3 Multivariates normales lineares Modell

Wir nehmen ein multivariates normales lineares Modell für die zufällige $n \times p$ Matrix \mathbf{X} an

$$\mathbf{X} \sim N_{np}(ZB, \Sigma \otimes \Lambda). \quad (4.32)$$

Die $n \times k$ Matrix Z mit den Zeilen Z_i besitze den Rang $d \leq \min(k, n)$ und sei fest vorgegeben. Die $n \times n$ Matrix Λ wird als bekannt und positiv definit vorausgesetzt. Die $k \times p$ Matrix B und die positiv definite $p \times p$ Matrix Σ sind die unbekannt Parameter. Das multivariate normale lineare Modell wurde in der Literatur intensiv studiert, ausführliche Diskussionen unterschiedlichster Aspekte aus der Bayesschen Sichtweise findet man in DeGroot (1970), Zellner (1971), Humak (1977), Anderson (1984) sowie Box und Tiao (1992).

Ferner sind in diesem Modell für eine Beobachtung $\mathbf{X} = X$ Versionen der im Allgemeinen nicht eindeutigen Maximum-Likelihood-Schätzungen für B und Σ gegeben durch

$$\hat{B} = \hat{B}(X) = (Z^t \Lambda^{-1} Z)^+ Z^t \Lambda^{-1} X, \quad (4.33)$$

$$\hat{\Sigma} = \hat{\Sigma}(X) = \frac{1}{n} (X - Z\hat{B})^t \Lambda^{-1} (X - Z\hat{B}). \quad (4.34)$$

Im Anschluss an diesen Abschnitt betrachten wir ein Modell mit von der Zeit abhängigen Parametern. Im Vorgriff werden wir hier zusätzlich zu dem möglicherweise nicht identifizierbaren Parameter B einen Parameter $U \in \mathbb{R}^{q \times p}$, von dem die Zufallsvariable \mathbf{X} nicht abhängt, hinzufügen. Insbesondere bezeichnen wir mit A die $(k+q) \times p$ Matrix $(B^t, U^t)^t$.

A priori Annahmen

Für den zufälligen Vektor (\mathbf{A}, Σ) nehmen wir eine Normal-Wishart-Verteilung

$$(\mathbf{A}, \Sigma) \sim NW(k+q, p, l, A_0, \Gamma_0, S_0)$$

als a priori Verteilung an (vgl. [Humak \(1977\)](#), S. 491, A 2.38, [Box und Tiao \(1992\)](#), S.423, Abschnitt 8.2):

1. Die Randverteilung der zufälligen Matrix Σ^{-1} ist eine Wishart-Verteilung

$$\Sigma^{-1} \sim W_p(l - (k+q), S_0^{-1}),$$

wobei $l - k - q > p + 1$ und S_0 eine positiv definite $p \times p$ Matrix ist.

2. Die bedingte Verteilung von $\mathbf{A} = (B^t, U^t)^t$ unter der Bedingung $\Sigma = \Sigma$ ist eine $(k+q) \times p$ -dimensionale Normalverteilung

$$\mathbf{A} | \Sigma = \Sigma \sim N_{(k+q)p}(A_0, \Sigma \otimes \Gamma_0^{-1}),$$

wobei $A_0 = (B_0^t, U_0^t)^t$ eine $(k+q) \times p$ Matrix und Γ_0 eine positiv definite $(k+q) \times (k+q)$ Matrix ist.

Bezeichnen wir weiterhin mit M die $n \times (k+q)$ Matrix $M = (Z; 0)$, so ist für eine Realisierung $(\mathbf{A}, \Sigma) = (A, \Sigma)$ die Verteilung von \mathbf{X} gegeben durch

$$\mathbf{X} \sim N_{np}(MA, \Sigma \otimes \Lambda). \quad (4.35)$$

A posteriori Verteilung

Insbesondere bildet nun die Familie der Normal-Wishart-Verteilungen eine konjugierte Familie für das multivariate normale lineare Modell (vgl. [Humak \(1977\)](#), S. 365, Kapitel 7). Bezeichnen wir zusätzlich zu (4.33) und (4.34) mit

$$\hat{A} = \hat{A}(X) = (M^t \Lambda^{-1} M)^+ M^t \Lambda^{-1} X = (0^t, \hat{B}^t)^t, \quad (4.36)$$

$$\tilde{\Gamma} = \Gamma_0 + M^t \Lambda^{-1} M, \quad (4.37)$$

$$\tilde{A} = \tilde{A}(X) = \tilde{\Gamma}^{-1} (\Gamma_0 A_0 + M^t \Lambda^{-1} M \hat{A}), \quad (4.38)$$

$$\tilde{S} = \tilde{S}(X) = S_0 + n \hat{\Sigma} + A_0^t \Gamma_0 A_0 + \hat{A}^t (M^t \Lambda^{-1} M) \hat{A} - \tilde{A}^t \tilde{\Gamma} \tilde{A}, \quad (4.39)$$

so ist die a posteriori Verteilung von (\mathbf{A}, Σ) unter der Bedingung $\mathbf{X} = X$ eine Normal-Wishart-Verteilung $NW(k+q, p, l+n, \tilde{A}, \tilde{\Gamma}, \tilde{S})$ mit den a posteriori Erwartungswerten (vgl. [Humak \(1977\)](#), S.489, A 2.19, [Anderson \(1984\)](#), S.270, Lemma 7.7.1)

$$\mathbb{E}[\mathbf{A} | \mathbf{X} = X] = \tilde{A}(X),$$

$$\mathbb{E}[\Sigma | \mathbf{X} = X] = \frac{1}{l - (k+q) - p + n - 1} \tilde{S}(X).$$

Selbstinformativer Grenzwert

Für alle $r \in \mathbb{N}$ betrachten wir nun r unabhängige Wiederholungen des Modells (4.32), d.h. das erweiterte Modell mit den Beobachtungen $\mathbf{X}_1, \dots, \mathbf{X}_r$ und

$$\mathbf{X}_i \sim N_{np}(ZB, \Sigma), \quad i = 1, \dots, r.$$

Weiterhin sei \mathbf{X}^r die $rn \times p$ Matrix $(\mathbf{X}_1^t, \dots, \mathbf{X}_r^t)^t$ und $Z_r = \mathbb{1}_r \otimes Z$. Dann erhalten wir mit $M_r = (Z_r, 0)$ analog zu (4.35) für eine Realisierung $(\mathbf{A}, \Sigma) = (A, \Sigma)$

$$\mathbf{X}^r \sim N_{rnp}(M_r A, \Sigma \otimes (I_r \otimes \Lambda)).$$

Unter der a priori Annahme

$$(\mathbf{A}, \Sigma) \sim NW(k + q, p, l, A_0, \Gamma_0, S_0)$$

ist eine Festlegung $P^{(\mathbf{A}, \Sigma | \mathbf{X}^r = X^r)}$ der a posteriori Verteilung von (\mathbf{A}, Σ) unter der Bedingung $\mathbf{X}^r = X^r$ eine Normal-Wishart-Verteilung $NW(k + q, p, l + rn, \tilde{A}(X^r), \tilde{\Gamma}_r, \tilde{S}(X^r))$, wobei sich die Parameter analog zum Abschnitt 4.6.3 ergeben. Insbesondere sind die a posteriori Erwartungswerte $\mathbb{E}[\mathbf{A} | \mathbf{X}^r = X^r] = \tilde{A}(X^r)$ und $\mathbb{E}[\Sigma | \mathbf{X}^r = X^r] = \frac{1}{l - (k + q) - p + rn - 1} \tilde{S}(X^r)$ Bayessche Schätzungen für die Parameter A und Σ . Wählen wir nun die originale Beobachtung X des Modells 4.35 als einen speziellen Wert $X^r = \mathbb{1}_r \otimes X$, so erhalten wir als eine Festlegung $P^{(\mathbf{A}, \Sigma | \mathbf{X}^r = \mathbb{1}_r \otimes X)}$ der a posteriori Verteilung von (\mathbf{A}, Σ) eine Normal-Wishart-Verteilung

$$(\mathbf{A}, \Sigma) | \mathbf{X}^r = \mathbb{1}_r \otimes X \sim NW(k + q, p, l + rn, \tilde{A}(\mathbb{1}_r \otimes X), \tilde{\Gamma}_r, \tilde{S}(\mathbb{1}_r \otimes X)) \quad (4.40)$$

und $\tilde{A}(\mathbb{1}_r \otimes X)$ sowie $\tilde{\Sigma}(\mathbb{1}_r^t \otimes X) = \frac{1}{l - (k + q) - p + rn - 1} \tilde{S}(\mathbb{1}_r \otimes X)$ als Bayessche Schätzungen für die Parameter A und Σ . Im Folgenden werden wir für die Folge $(P^{(\mathbf{A}, \Sigma) | \mathbf{X}^r = \mathbb{1}_r^t \otimes X})_{r \in \mathbb{N}}$ der a posteriori Verteilungen und $(\tilde{A}(\mathbb{1}_r^t \otimes X))_{r \in \mathbb{N}}$ sowie $(\tilde{\Sigma}(\mathbb{1}_r^t \otimes X))_{r \in \mathbb{N}}$ der Bayesschen Schätzungen den Grenzübergang für $r \rightarrow \infty$ betrachten. Das nächste Lemma liefert nun den Ausgangspunkt für die weiteren Untersuchungen.

Lemma 4.33 *Es sei $(\mathbf{A}_1, \Sigma_1), (\mathbf{A}_2, \Sigma_2), \dots$ eine Folge von Normal-Wishart-verteiltern Zufallsvariablen mit*

$$(\mathbf{A}_r, \Sigma_r) \sim NW(k, p, rn, A_r, rM_r^{-1}, rS_r) \quad \text{für } r \in \mathbb{N}.$$

Des weiteren existieren eine Matrix A und positiv definite Matrizen M und S mit

$$A = \lim_{r \rightarrow \infty} A_r, \quad M = \lim_{r \rightarrow \infty} M_r, \quad S = \lim_{r \rightarrow \infty} S_r.$$

Dann konvergiert die Folge der Zufallsvariablen $(\mathbf{A}_r, \Sigma_r)_{r \in \mathbb{N}}$ in Wahrscheinlichkeit gegen die Konstante (A, S)

$$(\mathbf{A}_r, \Sigma_r) \xrightarrow{P} (A, \frac{1}{n}S) \quad \text{für } r \rightarrow \infty.$$

BEWEIS : Es seien eine standardnormalverteilte Zufallsvariable $\Xi \sim N_{kp}(0, I_{kp})$ und eine Folge $\mathbf{W}_1, \mathbf{W}_2, \dots$ von Wishart-verteiltern Zufallsvariablen $\mathbf{W}_r \sim W_p(rn - k, 0)$ für $r \in \mathbb{N}$ gegeben. Dann konvergiert wegen dem Gesetz der Großen Zahlen die Folge $(\frac{n}{r} \mathbf{W}_r)_{r \in \mathbb{N}}$ fast sicher gegen die Einheitsmatrix I_p

$$\frac{n}{r} \mathbf{W}_r \xrightarrow{f.s.} I_p \quad \text{für } r \rightarrow \infty.$$

Zusammen mit

$$\mathbf{A}_r = A_r + \frac{1}{\sqrt{r}} M_r^{\frac{1}{2}} \Xi \Sigma_r^{-\frac{1}{2}} \quad \text{und} \quad \Sigma_r = \frac{1}{\sqrt{r}} S_r^{-\frac{1}{2}} \mathbf{W}_r \frac{1}{\sqrt{r}} S_r^{-\frac{1}{2}}$$

folgt dann die Behauptung. \square

Im Folgenden nehmen wir eine Reparametrisierung des Modells vor, um die Voraussetzungen für das Lemma 4.33 zu gewährleisten. Wir bezeichnen mit M_Λ die positiv semidefinite Matrix $M^t \Lambda^{-1} M$ und mit $\mathcal{N}(M_\Lambda)$ den Kern sowie mit $\mathcal{R}(M_\Lambda)$ das Bild der linearen Abbildung M_Λ . O.B.d.A. sei $\{h_1, \dots, h_{q+k}\}$ eine Orthonormalbasis des \mathbb{R}^{k+q} , wobei $\{h_1, \dots, h_d\}$ eine Orthonormalbasis des Bildes $\mathcal{R}(M_\Lambda)$ und $\{h_{d+1}, \dots, h_{q+k}\}$ eine Orthonormalbasis des Kerns $\mathcal{N}(M_\Lambda)$ darstelle. Insbesondere sind dann h_1, \dots, h_d Eigenvektoren zu den positiven reellen Eigenwerten $\lambda_1, \dots, \lambda_d$ der Matrix M_Λ . Ferner bezeichnen wir mit H_1 die Matrix mit den Spalten h_1, \dots, h_d und mit H_2 die Matrix mit den Spalten h_{d+1}, \dots, h_{q+k} . Nun erhalten wir für den Parameter $A \in \mathbb{R}^{k+q \times p}$ die Zerlegung $A = H_1 H_1^t A + H_2 H_2^t A = H_1 A_1 + H_2 A_2$ mit $A_1 \in \mathbb{R}^{d \times p}$ und $A_2 \in \mathbb{R}^{k+q-d \times p}$, wobei

$$H_1 H_1^t = M_\Lambda^+ M_\Lambda \quad \text{bzw.} \quad H_2 H_2^t = I_{k+q} - M_\Lambda^+ M_\Lambda \quad (4.41)$$

Projektionen auf die linearen Unterräume $\mathcal{R}(M_\Lambda)$ bzw. $\mathcal{N}(M_\Lambda)$ sind. Mit Hilfe dieser Zerlegung gilt dann

$$\mathbf{X}^r \sim N_{rnp}(M_r H_1 A_1, \Sigma \otimes (I_r \otimes \Lambda)), \quad (4.42)$$

so dass die Verteilung von \mathbf{X}^r unabhängig von dem Parameter A_2 ist. Bezeichnen wir nun den Parameter (A_1, Σ) als θ und den Parameter A_2 als γ , dann befinden wir uns in dem im Abschnitt 4.6.1 vorgestellten Modell. Unter den a priori Annahmen (vgl. Abschnitt 4.6.3) ist (\mathbf{A}, Σ) eine Normal-Wishart-verteilte Zufallsvariable. Damit besitzt die Zufallsvariable ϑ mit den Werten $\theta = (A_1, \Sigma)$ ebenfalls eine Normal-Wishart-Verteilung

$$NW(d, p, l - (k + q) + d, H_1^t A_0, (H_1^t \Gamma_0^{-1} H_1)^{-1}, S_0). \quad (4.43)$$

Ferner folgt γ unter der Bedingung $\vartheta = (A_1, \Sigma)$ einer Normalverteilung

$$N_{(k+q-d)p}(H_2^t A_0, \Sigma \otimes H_2^t \Gamma_0^{-1} H_2), \quad (4.44)$$

wobei diese Festlegung $P^\gamma | \vartheta = \theta$ der bedingten Verteilung von γ unter der Bedingung $\vartheta = \theta$ offensichtlich stetig in θ ist (vgl. Bemerkung 4.6). Ferner ist nun in dem Modell (4.42) mit der a priori Annahme (4.43) eine Festlegung $P^\vartheta | \mathbf{X}^r = \mathbb{1}_r \otimes X$ der a posteriori Verteilung von ϑ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r \otimes X$ eine Normal-Wishart-Verteilung

$$NW(d, p, l - (k + q) + d + rn, \tilde{A}_{r\theta}(X), \tilde{\Gamma}_{r\theta}, \tilde{S}_{r\theta}(X))$$

mit den Parametern

$$\begin{aligned} \tilde{\Gamma}_{r\theta} &= r^{-1} (H_1^t \Gamma_0^{-1} H_1)^{-1} + H_1^t M_\Lambda H_1, \\ \tilde{A}_{r\theta} &= \tilde{\Gamma}_{r\theta}^{-1} [r^{-1} (H_1^t \Gamma_0^{-1} H_1)^{-1} H_1^t A_0 + H_1^t M_\Lambda \hat{A}], \\ \tilde{S}_{r\theta} &= S_0 + rn \hat{\Sigma} + A_0^t \Gamma_0 A_0 + r \hat{A}^t M_\Lambda \hat{A} - r \tilde{A}_{r\theta}^t \tilde{\Gamma}_{r\theta} \tilde{A}_{r\theta}. \end{aligned}$$

Lemma 4.34 *Für $P^{\mathbf{X}}$ fast alle X konvergiert die Folge $(P^\vartheta | \mathbf{X}^r = \mathbb{1}_r \otimes X)_{r \in \mathbb{N}}$ der a posteriori Verteilungen von ϑ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r \otimes X$ schwach gegen das Punktmaß $\delta_{(H_1^t \hat{A}, \hat{\Sigma})}$ an der Stelle $(H_1^t \hat{A}, \hat{\Sigma})$:*

$$P^\vartheta | \mathbf{X}^r = \mathbb{1}_r \otimes X \xrightarrow{\mathcal{W}} \delta_{(H_1^t \hat{A}, \hat{\Sigma})} \quad \text{für } r \rightarrow \infty.$$

BEWEIS : Unter der Voraussetzung $(n - d) > p$ ist die Matrix $\hat{\Sigma}$ für fast alle $X \in \mathbb{R}^{n \times p}$ bzgl. des Lebesgue Maßes positiv definit (vgl. Humak (1977), S. 482, A 1.43). Des weiteren erhalten wir für die Parameter die Grenzwerte

$$H_1^t \hat{A} = \lim_{r \rightarrow \infty} \tilde{A}_{r\theta}, \quad \text{Diag}[\lambda_1 \dots, \lambda_d] = \lim_{r \rightarrow \infty} \frac{1}{r} \tilde{\Gamma}_{r\theta}, \quad n \hat{\Sigma} = \lim_{r \rightarrow \infty} \frac{1}{r} \tilde{S}_{r\theta},$$

so dass die Behauptung wegen Lemma 4.33 folgt. \square

Lemma 4.35 Für $P^{\mathbf{X}}$ -fast alle X konvergiert die Folge $(P^{\gamma | \mathbf{X}^r = \mathbb{1}_r^t \otimes X})_{r \in \mathbb{N}}$ der a posteriori Verteilungen von γ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r \otimes X$ schwach gegen die bedingte Verteilung $P^{\gamma | \vartheta = (H_1^t \hat{A}, \hat{\Sigma})}$ von γ unter der Bedingung $\vartheta = (H_1^t \hat{A}, \hat{\Sigma})$:

$$P^{\gamma | \mathbf{X}^r = \mathbb{1}_r \otimes X} \xrightarrow{\mathcal{W}} P^{\gamma | \vartheta = (H_1^t \hat{A}, \hat{\Sigma})} \quad \text{für } r \rightarrow \infty.$$

BEWEIS : Aufgrund ihrer Konstruktion erfüllen die Zufallsvariablen ϑ und γ die a priori Annahmen des Abschnittes 4.6.1. Insbesondere ist die gewählte Festlegung $P^{\gamma | \vartheta = \theta}$ stetig in θ (vgl. (4.44)), d.h. für jede beschränkte stetige Funktion h ist der bedingte Erwartungswert $\mathbb{E}_{\gamma | \vartheta = \theta}[h(\gamma)]$ eine stetige Funktion in θ . Weiterhin konvergiert nun wegen Lemma 4.34 die Folge $(P^{\vartheta | \mathbf{X}^r = \mathbb{1}_r \otimes X})_{r \in \mathbb{N}}$ schwach gegen das Punktmaß $\delta_{(H_1^t \hat{A}, \hat{\Sigma})}$. Somit ist $\{(H_1^t \hat{A}, \hat{\Sigma})\}$ ein schwacher a posteriori Träger für ϑ und es folgt wegen Satz 4.32 die Behauptung. \square

Betrachten wir nun das Modell

$$\mathbf{X} \sim N_{np}(MA, \Sigma \otimes \Lambda).$$

Wir sind an dem Grenzwert, falls er existiert, der Folge $(P^{\mathbf{A}, \Sigma | \mathbf{X}^r = \mathbb{1}_r \otimes X})_{r \in \mathbb{N}}$ der a posteriori Verteilungen von (\mathbf{A}, Σ) unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r \otimes X$ interessiert. Ferner haben wir bisher gezeigt, dass die Folgen $(P^{\vartheta | \mathbf{X}^r = \mathbb{1}_r \otimes X})_{r \in \mathbb{N}}$ und $(P^{\gamma | \mathbf{X}^r = \mathbb{1}_r \otimes X})_{r \in \mathbb{N}}$ schwach konvergieren. Aufgrund der Konstruktion von $\vartheta = (\vartheta_1, \vartheta_2)$ und γ gilt insbesondere $\mathbf{A} = H_1 \vartheta_1 + H_2 \gamma$ und $\Sigma = \vartheta_2$.

Lemma 4.36 Für $P^{\mathbf{X}}$ -fast alle X konvergiert die Folge $(P^{\Sigma | \mathbf{X}^r = \mathbb{1}_r^t \otimes X})_{r \in \mathbb{N}}$ der a posteriori Verteilungen von Σ unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r^t \otimes X$ schwach gegen das Punktmaß $\delta_{\hat{\Sigma}}$ an der Stelle $\hat{\Sigma}$:

$$P^{\Sigma | \mathbf{X}^r = \mathbb{1}_r \otimes X} \xrightarrow{\mathcal{W}} \delta_{\hat{\Sigma}} \quad \text{für } r \rightarrow \infty.$$

BEWEIS : Die Behauptung folgt direkt aus Lemma 4.34. \square

Lemma 4.37 Für $P^{\mathbf{X}}$ -fast alle X konvergiert die Folge $(P^{\mathbf{A} | \mathbf{X}^r = \mathbb{1}_r \otimes X})_{r \in \mathbb{N}}$ der a posteriori Verteilungen von \mathbf{A} unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r^t \otimes X$ schwach gegen eine Normalverteilung:

$$P^{\mathbf{A} | \mathbf{X}^r = \mathbb{1}_r \otimes X} \xrightarrow{\mathcal{W}} N_{k+q}(\hat{A} + H_2 H_2^t A_0, \Sigma \otimes H_2 H_2^t \Gamma_0^{-1} H_2 H_2^t) \quad \text{für } r \rightarrow \infty.$$

BEWEIS : Die Folge der Verteilungen $(P^{\vartheta_1 | \mathbf{X}^r = \mathbb{1}_r \otimes X})_{r \in \mathbb{N}}$ konvergiert schwach gegen das Punktmaß $\delta_{H_1^t \hat{A}}$ (vgl. Lemma 4.34) und die Folge $(P^{\gamma | \mathbf{X}^r = \mathbb{1}_r \otimes X})_{r \in \mathbb{N}}$ schwach gegen die bedingte Verteilung $P^{\gamma | \vartheta = (H_1^t \hat{A}, \hat{\Sigma})}$ von γ unter der Bedingung $\vartheta = (H_1^t \hat{A}, \hat{\Sigma})$ (vgl. Lemma 4.35), wobei $P^{\gamma | \vartheta = (H_1^t \hat{A}, \hat{\Sigma})}$ einer Normalverteilung $N_{k+q-d}(H_2 A_0, \hat{\Sigma} \otimes H_2^t \Gamma_0^{-1} H_2)$ entspricht (vgl. (4.44)). Es sei nun Ξ eine Zufallsvariable mit der Verteilung $P^{\gamma | \vartheta = (H_1^t \hat{A}, \hat{\Sigma})}$. Dann konvergiert die Folge $(P^{H_1 \vartheta_1 + H_2 \gamma | \mathbf{X}^r = \mathbb{1}_r \otimes X})_{r \in \mathbb{N}}$ schwach gegen die Verteilung $P^{H_1 H_1^t \hat{A} + H_2 \Xi}$ der Zufallsvariable $H_1 H_1^t \hat{A} + H_2 \Xi$ (vgl. Billingsley (1968), S. 27, Satz 4.4), wobei $P^{H_1 H_1^t \hat{A} + H_2 \Xi}$ offensichtlich einer Normalverteilung $N_{k+q}(\hat{A} + H_2 H_2^t A_0, \Sigma \otimes H_2 H_2^t \Gamma_0^{-1} H_2 H_2^t)$ entspricht. \square

Ferner ist unter der Annahme $\mathbf{X} \sim N_{np}(MA, \Sigma \otimes \Lambda)$ die Menge der Maximum-Likelihood-Schätzungen für A der affine Raum (vgl. (4.33))

$$\hat{\Theta}_A(X) := \{\hat{A}(X) + v \mid v \in \mathcal{N}(M_\Lambda)\} = \left\{ \begin{pmatrix} \hat{B}(X) + B \\ U \end{pmatrix} \mid U \in \mathbb{R}^{q \times p}, ZB = 0, B \in \mathbb{R}^{k \times p} \right\} \quad (4.45)$$

und für Σ die einelementige Menge (vgl. (4.34))

$$\hat{\Theta}_\Sigma(X) = \{\hat{\Sigma}(X)\}. \quad (4.46)$$

Satz 4.38 *Für $P^{\mathbf{X}}$ -fast alle X ist die Menge $\hat{\Theta}(X) = \hat{\Theta}_A(X) \times \hat{\Theta}_\Sigma(X)$ ein schwacher selbstinformativer a posteriori Träger.*

BEWEIS : Die Behauptung folgt direkt aus der schwachen Konvergenz der Folge der a posteriori Verteilungen $(P^{\mathbf{A}, \Sigma} | \mathbf{X}^r = \mathbb{1}_r \otimes X)_{r \in \mathbb{N}}$ von (\mathbf{A}, Σ) unter der Bedingung $\mathbf{X}^r = \mathbb{1}_r \otimes X$ gegen die Verteilung der Zufallsvariable $(\tilde{\Sigma}, \tilde{\mathbf{A}})$ (vgl. Lemma 4.36 und 4.37), wobei die Verteilung von $\tilde{\mathbf{A}}$ wegen Lemma 4.37 eine auf dem affinen Raum $\hat{\Theta}_A$ konzentrierte Normalverteilung ist. \square

Abschließend betrachten wir nun den Grenzübergang für die Folgen $(\tilde{A}(\mathbb{1}_r \otimes X))_{r \in \mathbb{N}}$ und $(\tilde{\Sigma}(\mathbb{1}_r \otimes X))_{r \in \mathbb{N}}$ der Bayesschen Schätzungen für die Parameter A und Σ .

Satz 4.39 *Die selbstinformativen Grenzwerte der Bayesschen Schätzungen für A und Σ sind*

$$\bar{A}(X) = \begin{pmatrix} \hat{B}(X) + (I_k - (Z^t \Lambda^{-1} Z)^+ (Z^t \Lambda^{-1} Z)) B_0 \\ U_0 \end{pmatrix} \quad \text{und} \quad \bar{\Sigma}(X) = \hat{\Sigma}(X).$$

BEWEIS : Es sei G die Matrix $\Gamma_0^{-\frac{1}{2}} M_\Lambda \Gamma_0^{-\frac{1}{2}}$. Dann ist G positiv semidefinit und besitzt den Rang d . Ferner seien v_1, \dots, v_d orthonormale Eigenvektoren zu den positiven Eigenwerten $\kappa_1, \dots, \kappa_d$ der Matrix G und o.B.d.A. bilden diese mit den orthonormalen Vektoren v_{d+1}, \dots, v_{k+q} eine Orthonormalbasis des \mathbb{R}^{k+q} . Insbesondere gilt nun für die Parameter (vgl. (4.40)):

$$\begin{aligned} \tilde{\Gamma}_r^{-1} &= \left(\frac{1}{r} \Gamma_0 + M_\Lambda \right)^{-1} = \Gamma_0^{-\frac{1}{2}} \left(\sum_{i=d+1}^{k+q} r v_i v_i^t + \sum_{i=1}^d (r^{-1} + \kappa_i)^{-1} v_i v_i^t \right) \Gamma_0^{-\frac{1}{2}}, \\ \tilde{A}_r &= \tilde{A}(\mathbb{1}_r \otimes X) = \tilde{\Gamma}_r^{-1} \left(\frac{1}{r} \Gamma_0 A_0 + M_\Lambda \hat{A} \right) \\ &= \Gamma_0^{-\frac{1}{2}} \left(\sum_{i=d+1}^{k+q} v_i v_i^t + \sum_{i=1}^d \frac{v_i v_i^t}{1 + r \kappa_i} \right) \Gamma_0^{-\frac{1}{2}} \Gamma_0 A_0 + \Gamma_0^{-\frac{1}{2}} \left(\sum_{i=1}^d \frac{r v_i v_i^t}{1 + r \kappa_i} \right) \Gamma_0^{-\frac{1}{2}} M_\Lambda \hat{A}, \\ \tilde{S}_r &= \tilde{S}(\mathbb{1}_r \otimes X) = S_0 + r n \tilde{\Sigma} + A_0^t \Gamma_0 A_0 + r \hat{A}^t M_\Lambda \hat{A} - r \tilde{A}_r^t \tilde{\Gamma}_r \tilde{A}_r. \end{aligned}$$

Insbesondere folgt nun wegen

$$\begin{aligned} \lim_{r \rightarrow \infty} \tilde{A}_r &= \Gamma_0^{-\frac{1}{2}} (I_{q+k} - G^+ G) \Gamma_0^{-\frac{1}{2}} \Gamma_0 A_0 + \Gamma_0^{-\frac{1}{2}} G^+ \Gamma_0^{-\frac{1}{2}} M_\Lambda \hat{A} = (I - M_\Lambda^+ M_\Lambda) A_0 + \hat{A}, \\ \lim_{r \rightarrow \infty} \tilde{S}_r &= n \hat{\Sigma} \end{aligned}$$

die Behauptung. \square

4.6.4 Multivariates normales zeitabhängiges lineares Modell

Für jeden Punkt t in dem Intervall $[a, b] \subset \mathbb{R}$ sei $\mathbf{Y}(t)$ eine Zufallsvariable mit Werten im \mathbb{R}^q . Des weiteren seien Funktionen $f_j : [a, b] \rightarrow \mathbb{R}^{k_j}$, $k_j \in \mathbb{N}$, $j = 1, \dots, m$, mit

$$\mathbf{Y}(t_i) = \sum_{j=1}^m Z_j(t_i) f_j(t_i) + \sigma \boldsymbol{\varepsilon}(t_i),$$

für jede endliche Menge $T := \{t_1, \dots, t_n\}$ von Punkten $t_i \in [a, b]$, $i = 1, \dots, n$, gegeben. Weiterhin seien die $q \times k_j$ Matrizen $Z_j(t_i)$ bekannt und der zufällige Vektor $\boldsymbol{\varepsilon}(T) = (\boldsymbol{\varepsilon}(t_1)^t, \dots, \boldsymbol{\varepsilon}(t_n)^t)^t$ normalverteilt mit Erwartungswert Null und bekannter Kovarianzmatrix Λ . Die Funktionen f_1, \dots, f_m und der Parameter $\sigma > 0$ sind nun die unbekannt Parameter.

A priori Annahmen

In einem Bayesschen Ansatz fassen wir die Funktionen f_1, \dots, f_m als Realisierung zufälliger Funktionen $\mathbf{f}_1, \dots, \mathbf{f}_m$ und σ^2 als Realisierung einer Zufallsgröße $\boldsymbol{\sigma}^2$ auf. Insbesondere nehmen wir an, dass unter der Bedingung $\boldsymbol{\sigma}^2 = \sigma^2$ die zufälligen Funktionen $\mathbf{f}_1, \dots, \mathbf{f}_m$ unabhängig und normalverteilt seien, d.h. für jede endliche Menge $T = \{t_1, \dots, t_n\}$ von Punkten $t_i \in [a, b]$ seien die Vektoren $\mathbf{f}_j(T) = (\mathbf{f}_j(t_1)^t, \dots, \mathbf{f}_j(t_n)^t)^t$, $j = 1, \dots, m$, unabhängig und normalverteilt. Unter der Bedingung $\boldsymbol{\sigma}^2 = \sigma^2$ ist dann die a priori Verteilung der zufälligen Funktion \mathbf{f}_j eindeutig durch ihre Erwartungswertfunktion $\mathbb{E} \mathbf{f}_j = f_j^0$ und die Kovarianzfunktion $\text{Cov} \mathbf{f}_j = \sigma^2 \Sigma_j^0$ für $j = 1, \dots, m$, bestimmt. A priori sei weiterhin der Zufallsvektor $\boldsymbol{\varepsilon}(T)$ unabhängig von dem zufälligen Vektor $\mathbf{f}(T)$ und der Zufallsgröße $\boldsymbol{\sigma}^2$. Insbesondere ist nun für $j = 1, \dots, m$, der Vektor $\mathbf{f}_j(T)$ normalverteilt mit

$$\mathbb{E} \mathbf{f}_j(T) = f_j^0(T) = (f_j^0(t_1)^t, \dots, f_j^0(t_n)^t)^t \text{ und } \text{Cov} \mathbf{f}_j(T) = \sigma^2 \Sigma_j^0(T) = \sigma^2 \left(\Sigma_j^0(t_i, t_l) \right)_{\substack{i=1, \dots, n \\ l=1, \dots, n}}.$$

Bezeichnen wir weiterhin den Vektor $(\mathbf{f}_1(T)^t, \dots, \mathbf{f}_m(T)^t)^t$ mit $\mathbf{f}(T)$. Dann ist $\mathbf{f}(T)$ unter der Bedingung $\boldsymbol{\sigma}^2 = \sigma^2$ normalverteilt mit

$$\mathbb{E} \mathbf{f}(T) = f^0(T) = (f_1^0(T)^t, \dots, f_m^0(T)^t)^t \quad \text{und} \quad \Sigma^0(T) := \sigma^2 \text{Diag}[\Sigma_1^0(T), \dots, \Sigma_m^0(T)].$$

Des weiteren ist nun die bedingte Verteilung des Zufallsvektors $\mathbf{Y}(T) = (\mathbf{Y}(t_1)^t, \dots, \mathbf{Y}(t_n)^t)^t$ durch

$$\mathbf{Y}(T) | \mathbf{f}(T), \boldsymbol{\sigma}^2 \sim N_{nq} \left(\sum_{j=1}^m \text{Diag}[Z_j(t_1), \dots, Z_j(t_n)] \mathbf{f}_j(T), \boldsymbol{\sigma}^2 \Lambda \right)$$

gegeben. Bezeichnen wir außerdem mit $Z_j(T)$ die Matrix $\text{Diag}[Z_j(t_1), \dots, Z_j(t_n)]$ und mit $Z(T)$ die Matrix $(Z_1(T) \cdots Z_m(T))$, dann erhalten wir unter der Bedingung $\boldsymbol{\sigma}^2 = \sigma^2$ das normale lineare Modell

$$\mathbf{f}(T) | \boldsymbol{\sigma}^2 \sim N_{nk}(f^0(T), \boldsymbol{\sigma}^2 \Sigma^0(T)) \quad \text{für } k = \sum_{j=1}^m k_j, \quad (4.47)$$

$$\mathbf{Y}(T) | \mathbf{f}(T), \boldsymbol{\sigma}^2 \sim N_{nq}(Z(T) \mathbf{f}(T), \boldsymbol{\sigma}^2 \Lambda). \quad (4.48)$$

Definieren wir analog für eine endliche Menge $S := \{s_1, \dots, s_d\}$ von Punkten $s_i \in [a, b]$ mit $S \cap T = \emptyset$ den zufälligen Vektor $\mathbf{f}(S)$. Dann hat offensichtlich der Vektor $\mathbf{f}(S)$ keinen Einfluss auf die Verteilung von $\mathbf{Y}(T)$, d.h.

$$P^{\mathbf{Y}(T) | \mathbf{f}(T), \boldsymbol{\sigma}^2} = P^{\mathbf{Y}(T) | \mathbf{f}(T), \mathbf{f}(S), \boldsymbol{\sigma}^2}.$$

Insbesondere ist nun unter der Bedingung $\sigma^2 = \sigma^2$ die gemeinsame Verteilung des Vektors $\mathbf{f}(T, S) = (\mathbf{f}(T)^t, \mathbf{f}(S)^t)^t$ eine Normalverteilung mit

$$\mathbb{E} \mathbf{f}(T, S) = f^0(T, S) = (f^0(T)^t, f^0(S)^t)^t \quad \text{und} \quad \text{Cov} \mathbf{f}(T, S) = \sigma^2 \Sigma^0(T, S).$$

Abschließend besitze nun die Zufallsgröße σ^2 als a priori Verteilung eine Inverse-Gamma-Verteilung (vgl. Humak (1977) S.491, A 2.38)

$$\sigma^{-2} \sim Ga\left(\frac{w - nk - dk}{2}, \sigma_0^{-2}\right), \quad \text{mit } w > nk + dk + 2 \text{ und } \sigma_0 > 0, \quad (4.49)$$

wobei eine reellwertige Zufallsvariable der Gammaverteilung $Ga(b, p)$ mit den Parameter $b > 0$ und $p > 0$ genügt, wenn sie die Dichte

$$f(x) = \begin{cases} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} & : x > 0, \\ 0 & : x \leq 0 \end{cases}$$

bzgl. des Lebesgueschen Maßes besitzt (vgl. Müller (1991)). Dabei bezeichnet Γ die Gamma-Funktion mit

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt, \quad x > 0.$$

Insbesondere befinden wir uns nun in dem Spezialfall $p = 1$ des in Abschnitt 4.6.3 vorgestellten Bayesschen Modells.

Selbstinformativer Grenzwert

In dem hier vorgestellten Modell 4.48 sind für eine Beobachtung $\mathbf{Y}(T) = Y(T)$ Versionen der im Allgemeinen nicht eindeutigen Maximum-Likelihood-Schätzungen für $f(T)$ und σ^2

$$\hat{f}(T) = (Z(T)^t \Lambda^{-1} Z(T))^+ Z(T)^t \Lambda^{-1} Y(T), \quad (4.50)$$

$$\hat{\sigma}^2(T) = \frac{1}{nq} (Y(T) - Z(T) \hat{f}(T))^t \Lambda^{-1} (Y(T) - Z(T) \hat{f}(T)). \quad (4.51)$$

Weiterhin seien die einelementige Menge

$$\hat{\Theta}_{\sigma^2}(Y(T)) = \{\hat{\sigma}^2(T)\} \quad (4.52)$$

und der affine Raum

$$\hat{\Theta}_{f(T,S)}(Y(T)) = \left\{ \begin{pmatrix} \hat{f}(T) + f(T) \\ f(S) \end{pmatrix} \mid f(S) \in \mathbb{R}^{dk}, Z(T)f(T) = 0, f(T) \in \mathbb{R}^{nk} \right\} \quad (4.53)$$

gegeben.

Korollar 4.40 Für $P^{\mathbf{Y}(T)}$ -fast alle $Y(T)$ ist $\hat{\Theta}(Y(T)) = \hat{\Theta}_{f(T,S)}(Y(T)) \times \hat{\Theta}_{\sigma^2}(Y(T))$ ein schwacher selbstinformativer a posteriori Träger für $\mathbf{f}(T, S)$.

BEWEIS : Die Behauptung folgt direkt aus Satz 4.38. \square

Korollar 4.41 Die selbstinformativen Grenzwerte der Bayesschen Schätzungen für $f(T, S)$ und σ^2 sind

$$\bar{f}(T, S) = \begin{pmatrix} \hat{f}(T) + (I_k - (Z(T)^t \Lambda^{-1} Z(T))^+ (Z(T)^t \Lambda^{-1} Z(T))) f_0(T) \\ f_0(S) \end{pmatrix} \quad \text{und} \quad \bar{\sigma}^2(T) = \hat{\sigma}^2(T).$$

BEWEIS : Die Behauptung folgt direkt aus Satz 4.39 □

Bemerkung 4.42 Nehmen wir zusätzlich an, dass die Matrix $Z(T)$ den Rang nk besitze. Dann ist die MLS $\hat{f}(T)$ für $f(T)$ eindeutig bestimmt und offensichtlich gilt für die selbstinformativen Grenzwerte der Bayesschen Schätzungen für $f(T, S)$

$$\bar{f}(T, S) = \begin{pmatrix} \hat{f}(T) \\ f_0(S) \end{pmatrix} \quad \text{und} \quad \overline{\sigma^2}(T) = \hat{\sigma}^2(T).$$

Damit entspricht für alle Punkte $t \in T$ der selbstinformativ Grenzwert $\bar{f}(t)$ der Maximum-Likelihood-Schätzung $\hat{f}(t)$ und für alle Punkte $s \notin T$ der a priori Annahme $f_0(s)$. Weiterhin ist der affine Unterraum

$$\hat{\Theta}_{f(T,S)}(Y(T)) = \left\{ \begin{pmatrix} \hat{f}(T) \\ f(S) \end{pmatrix} \mid f(S) \in \mathbb{R}^{dk} \right\}.$$

ein schwacher a posteriori Träger für den zufälligen Vektor $\mathbf{f}(T, S)$. □

Kapitel 5

Nichtparametrisches Modell

5.1 Modellannahmen

Wir betrachten unabhängige Zufallsvariablen $\mathbf{X}_1, \dots, \mathbf{X}_n$ mit Werten in dem messbaren Raum $(\mathcal{X}, \mathfrak{B})$ und identischer Randverteilung P auf \mathfrak{B} . Das n -fache Produktmaß $P^n = P \times \dots \times P$ auf \mathfrak{B}^n sei somit die gemeinsame Verteilung des Vektors $\mathbf{X}^n = (\mathbf{X}_1, \dots, \mathbf{X}_n)$. Die Randverteilung P einer Zufallsvariable \mathbf{X}_i , $i = 1, \dots, n$ sei unbekannt und variere in der Menge aller Wahrscheinlichkeitsverteilungen \mathcal{P} . Zusammenfassend wird das statistische Experiment $(\mathcal{X}^n, \mathfrak{B}^n, P^n | P \in \mathcal{P})$ durch die Zufallsvariable \mathbf{X}^n beschrieben. Insbesondere ist im Allgemeinen die Menge der Wahrscheinlichkeitsverteilungen \mathcal{P} nicht dominiert. Im Folgenden wird gezeigt, dass die bekannte empirische Verteilungsfunktion

$$\hat{F}_{x^n} = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}, \quad x^n = (x_1, \dots, x_n) \quad (5.1)$$

einer Beobachtung $\mathbf{X}^n = x^n$ die verallgemeinerte Maximum-Likelihood-Schätzung im Sinn von Kiefer und Wolfowitz oder von Gill sowie der selbstinformativ Grenzwert einer Bayesschen Schätzung ist.

5.2 Ansatz von Kiefer und Wolfowitz

Die Eigenschaft, dass die empirische Verteilungsfunktion \hat{F}_{x^n} für eine Beobachtung $\mathbf{X}^n = x^n$ in dem vorgestellten nichtparametrischen Modell eine verallgemeinerte Maximum-Likelihood-Schätzung ist, wurde zuerst von [Kiefer und Wolfowitz \(1956\)](#) notiert. Basierend auf dieser Notiz wurden verschiedene Kriterien einer Definition einer nichtparametrischen MLS angegeben. Eine Möglichkeit ist die Definition einer „Likelihood-Funktion“

$$L_{x^n}(P) := \prod_{i=1}^n P(\{x_i\})$$

für $x^n \in \mathcal{X}^n$ und $P \in \mathcal{P}$ (vgl. [Owen \(2001\)](#)). Das folgende Lemma besagt nun, dass die empirische Verteilungsfunktion die eindeutig bestimmte nichtparametrische MLS bzgl. der „Likelihood-Funktion“ L_{x^n} ist.

Lemma 5.1 (Owen (2001), S. 8, Satz 2.1)

Es sei $x^n \in \mathcal{X}^n$ gegeben. Dann ist \hat{F}_{x^n} die eindeutig bestimmte Lösung der Gleichung

$$L_{x^n}(\hat{F}_{x^n}) = \max_{P \in \mathcal{P}} L_{x^n}(P).$$

Wir werden aus zwei Gründen das bekannte Ergebnis, dass die empirische Verteilungsfunktion die eindeutig festgelegte verallgemeinerte Maximum-Likelihood-Schätzung im Sinn von Kiefer und Wolfowitz ist (vgl. Definition 3.1), im Folgenden beweisen. Zum einen wurde in der Literatur auf diesen Beweis verzichtet und zum anderen werden wir die Beweisidee auf ein semiparametrisches lineares Modell (Kapitel 6) übertragen.

Satz 5.2 *Es sei $\mathbf{X}^n = x^n$ eine Beobachtung mit $\mathbf{X} \sim P^n$ und $P \in \mathcal{P}$. Die eindeutig bestimmte verallgemeinerte Maximum-Likelihood-Schätzung $\hat{P}(x^n)$ von P nach Kiefer und Wolfowitz ist die empirische Verteilung \hat{F}_{x^n} .*

BEWEIS : Es sei $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$ fest vorgegeben. Es ist zu zeigen, dass für alle Verteilungen $P \in \mathcal{P}$ eine Festlegung der Radon-Nikodym-Dichten von P und \hat{F}_{x^n} bzgl. $[P + \hat{F}_{x^n}]$ mit

$$\frac{d[\hat{F}_{x^n}]^n}{d[(\hat{F}_{x^n})^n + P^n]}(x^n) \geq \frac{dP^n}{d[(\hat{F}_{x^n})^n + P^n]}(x^n)$$

existiert. Zunächst gilt für alle $P_1, P_2 \in \mathcal{P}$, dass $P_1^n + P_2^n$ durch das σ -endliche Maß $(P_1 + P_2)^n$ dominiert ist. Somit existiert eine Radon-Nikodym Dichte von $P_1^n + P_2^n$ bzgl. $[P_1 + P_2]^n$ mit

$$\prod_{j=1}^n \frac{dP_j}{d[P_1 + P_2]} \equiv \frac{dP_j^n}{d[P_1 + P_2]^n} \equiv \frac{dP_j^n}{d[P_1^n + P_2^n]} \frac{d[P_1^n + P_2^n]}{d[P_1 + P_2]^n}, \quad j = 1, 2.$$

Diese Gleichheit gilt für beliebige Repräsentanten jeweils bis auf $[P_1 + P_2]^n$ -Nullmengen. Können wir aber zeigen, dass für alle $P \in \mathcal{P}$ die Radon-Nikodym Dichte $\prod_{i=1}^n \frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}$ an der Stelle x^n eindeutig festgelegt und positiv ist, erhalten wir die Behauptung.

Nun ist für alle $P \in \mathcal{P}$ eine Version der Radon-Nikodym Dichte von \hat{F}_{x^n} bzgl. $[\hat{F}_{x^n} + P]$ durch

$$\frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}(y) = \frac{1}{n} \sum_{j=1}^n \frac{1}{\hat{F}_{x^n}(\{x_j\}) + P(\{x_j\})} \mathbb{I}_{\{x_j\}}(y), \quad \forall y \in \mathcal{X}$$

gegeben, da für alle $B \in \mathfrak{B}$ gilt

$$\begin{aligned}
\int_B \frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}(y) \hat{F}_{x^n}(dy) &= \int_B \frac{1}{n} \sum_{j=1}^n \frac{1}{\hat{F}_{x^n}(\{x_j\}) + P(\{x_j\})} \mathbb{I}_{\{x_j\}}(y) \hat{F}_{x^n}(dy) \\
&= \frac{1}{n} \sum_{j=1}^n \frac{1}{\hat{F}_{x^n}(\{x_j\}) + P(\{x_j\})} \delta_{x_j}(B) \int_{\{x_j\}} \hat{F}_{x^n}(dy) \\
&= \frac{1}{n} \sum_{j=1}^n \frac{\hat{F}_{x^n}(\{x_j\})}{\hat{F}_{x^n}(\{x_j\}) + P(\{x_j\})} \delta_{x_j}(B) \quad \text{und weiterhin} \\
\int_B \frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}(y) P(dy) &= \frac{1}{n} \sum_{j=1}^n \frac{P(\{x_j\})}{\hat{F}_{x^n}(\{x_j\}) + P(\{x_j\})} \delta_{x_j}(B).
\end{aligned}$$

Somit gilt für alle $P \in \mathcal{P}$ und $B \in \mathfrak{B}$, dass

$$\begin{aligned}
\int_B \frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}(y) [\hat{F}_{x^n} + P](dy) &= \int_B \frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}(y) \hat{F}_{x^n}(dy) + \int_B \frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}(y) P(dy) \\
&= \frac{1}{n} \sum_{j=1}^n \delta_{x_j}(B) = \hat{F}_{x^n}(B)
\end{aligned}$$

ist. Da $[\hat{F}_{x^n} + P](\{x_i\})$ positiv für $i = 1, \dots, n$ ist, ist die Radon-Nikodym Dichte von \hat{F}_{x^n} bzgl. $\hat{F}_{x^n} + P$ eindeutig bestimmt und festgelegt durch

$$\frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}(x_i) = \frac{\hat{F}_{x^n}(\{x_i\})}{\hat{F}_{x^n}(\{x_i\}) + P(\{x_i\})} \quad \text{für } i = 1, \dots, n.$$

Weiterhin ist mit

$$1 \equiv \frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]} + \frac{dP}{d[\hat{F}_{x^n} + P]},$$

die Radon-Nikodym Dichte von P bzgl. $\hat{F}_{x^n} + P$ ebenfalls durch

$$\frac{dP}{d[\hat{F}_{x^n} + P]}(x_i) = \frac{P(\{x_i\})}{\hat{F}_{x^n}(\{x_i\}) + P(\{x_i\})} \quad \text{für } i = 1, \dots, n$$

festgelegt. Es bleibt daher zu zeigen, dass

$$\prod_{i=1}^n \frac{\hat{F}_{x^n}(\{x_i\})}{\hat{F}_{x^n}(\{x_i\}) + P(\{x_i\})} \geq \prod_{i=1}^n \frac{P(\{x_i\})}{\hat{F}_{x^n}(\{x_i\}) + P(\{x_i\})}$$

oder äquivalent dazu

$$L_{x^n}(\hat{F}_{x^n}) = \prod_{i=1}^n \hat{F}_{x^n}(\{x_i\}) \geq \prod_{i=1}^n P(\{x_i\}) = L_{x^n}(P)$$

gilt. Lemma 5.1 liefert nun die Behauptung, da die Radon-Nikodym Dichte $\prod_{i=1}^n \frac{d\hat{F}_{x^n}}{d[\hat{F}_{x^n} + P]}$ an der Stelle x^n eindeutig festgelegt und positiv ist. \square

Bemerkung 5.3 Damit charakterisiert die vorgestellte „Likelihood-Funktion“ $L_x(P)$ den zentralen Bestandteil der Radon-Nikodym-Dichte von P bzgl. $[P + \hat{F}_{x^n}]$ ausgewertet an der Stelle x^n , so dass die Aussagen des Lemmas 5.1 und des Satzes 5.2 in dem vorgestellten nichtparametrischen Modell äquivalent sind. \square

5.3 Ansatz von Gill

Das folgende Ergebnis ist der Arbeit von Gill (1989) entnommen. Der dort dargestellte Beweis wird hier noch einmal nachvollzogen, vor allem um den Beweis im Fall eines semiparametrischen linearen Modells, der im Kapitel 6 dargestellt wird und nicht mehr auf der Arbeit von Gill beruht, zu motivieren.

Wir betrachten das in Abschnitt 5.1 vorgestellte nichtparametrische Modell.

Satz 5.4 *Es sei $\mathbf{X}^n = x^n$ eine Beobachtung mit $\mathbf{X}^n \sim P^n$ und $P \in \mathcal{P}$. Die eindeutig bestimmte verallgemeinerte Maximum-Likelihood-Schätzung $\hat{P}(x^n)$ von P nach Gill ist die empirische Verteilungsfunktion \hat{F}_{x^n} .*

BEWEIS : Wir identifizieren den Parameterraum Θ mit der Menge aller Wahrscheinlichkeitsverteilungen \mathcal{P} und geben eine Abbildung

$$\psi : \mathcal{P} \times C_b(\mathcal{X}, \mathfrak{B}) \times \mathbb{R} \rightarrow \mathcal{P}$$

an, die die geforderten Eigenschaften (vgl. Annahme A.3.1):

1. für alle $P \in \mathcal{P}$ und $h \in C_b(\mathcal{X}, \mathfrak{B})$ gilt $\psi(P, h, 0) = P$,
2. für alle $P \in \mathcal{P}$ und $h \in C_b(\mathcal{X}, \mathfrak{B})$ existiert ein σ -endliches Maß $\mu(P, h)$, so dass die Familie der Wahrscheinlichkeitsverteilungen $\{\psi(P, h, c) \mid c \in \mathbb{R}\}$ durch das Maß $\mu(P, h)$ dominiert ist,
3. für alle $x \in \mathcal{X}$ ist die Radon-Nikodym Dichte $\frac{d\psi(P, h, c)}{d\mu(P, h)}(x)$ in c an der Stelle $c = 0$ differenzierbar,

erfüllt, wobei $C_b(\mathcal{X}, \mathfrak{B})$ die Menge aller beschränkten stetigen Funktionen auf $(\mathcal{X}, \mathfrak{B})$ sei. Des weiteren seien

$$L_{x^n}(c, P, h) := \prod_{i=1}^n \frac{d\psi(P, h, c)}{d\mu(P, h)}(x_i),$$

$$U_{x^n}(P, h) := \left. \frac{\partial}{\partial c} \left(\log L_{x^n}(c, P, h) \right) \right|_{c=0}.$$

Eine Schätzfunktion $\hat{P}(x^n)$ heißt nun verallgemeinerte Maximum-Likelihood-Schätzung nach Gill, falls gilt (vgl. Definition 3.3)

$$\forall h \in C_b(\mathcal{X}, \mathfrak{B}) : U_{x^n}(\hat{P}(x^n), h) = 0.$$

Wir betrachten die durch

$$\frac{d\psi(P, h, c)}{dP}(x) := \prod_{i=1}^n \frac{1 + ch(x_i)}{\int_{\mathcal{X}} (1 + ch(x)) P(dx)}$$

definierte Abbildung ψ . Dabei erfüllt ψ die geforderten Eigenschaften, falls c hinreichend nahe an 0 gewählt wird. Sei $P \in \mathcal{P}$ und $h \in C_b(\mathcal{X}, \mathfrak{B})$ fest gegeben und c hinreichend nahe an 0. Dann ist $1 + ch(x)$ nicht negativ für alle $x \in \mathcal{X}$ und das Integral $\int_{\mathcal{X}} 1 + ch(x) dP_P(x)$ endlich. Damit ist $\frac{d\psi(P, h, c)}{dP}$ die Dichtefunktion eines Wahrscheinlichkeitsmaßes bzgl. P . Des weiteren gilt offensichtlich, dass $\psi(P, h, 0) = P$

ist und dass die Dichte $\frac{d\psi(P,h,c)}{dP}(x)$ für alle $x \in \mathfrak{X}$ in c differenzierbar ist. Insbesondere folgt nun

$$\begin{aligned} L_{x^n}(c, P, h) &= \prod_{i=1}^n \frac{1 + ch(x_i)}{\int_{\mathfrak{X}} (1 + ch(x)) P(dx)}, \\ \log L_{x^n}(c, P, h) &= \sum_{i=1}^n \log(1 + ch(x_i)) - \sum_{i=1}^n \log \int_{\mathfrak{X}} (1 + ch(x)) P(dx), \\ U_{x^n}(P, h) &= \sum_{i=1}^n h(x_i) - n \int_{\mathfrak{X}} h(x) P(dx). \end{aligned}$$

Eine Schätzfunktion $\hat{P}(x^n)$ ist somit eine verallgemeinerte Maximum-Likelihood-Schätzung nach Gill, falls gilt

$$\forall h \in C_b(\mathfrak{X}, \mathfrak{B}) \quad : \quad \int_{\mathfrak{X}} h(x) \left[\frac{1}{n} \sum_{i=1}^n \delta_{x_i} - \hat{P}(x^n) \right] (dx) = 0,$$

so dass $\hat{P}(x^n) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ folgt und $\hat{P}(x^n)$ eindeutig bestimmt ist (vgl. [Billingsley \(1968\)](#), S.9, Satz 1.3). □

5.4 Selbstinformativer Grenzwert

Es sei eine das statistische Experiment $(\mathcal{X}^n, \mathfrak{B}^n, P^n | P \in \mathcal{P})$ beschreibende Zufallsvariable $\mathbf{X}^n = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ gegeben (vgl. Abschnitt 5.1). Die Randverteilung P einer Beobachtung \mathbf{X}_i ($i = 1, \dots, n$) ist der unbekannte Parameter und variiert in der Menge aller Wahrscheinlichkeitsverteilungen \mathcal{P} . Im Bayesschen Sinn nehmen wir an, dass die Randverteilung P selbst Realisierung einer Zufallsvariable \mathbf{P} ist. Dann bezeichnen wir \mathbf{P} als ein zufälliges Maß und unter der Bedingung $\mathbf{P} = P$ beschreibt \mathbf{X}^n das statistische Experiment $(\mathcal{X}^n, \mathfrak{B}^n, P^n | P \in \mathcal{P})$. Dazu ist es notwendig, auf der Menge der Wahrscheinlichkeitsverteilungen \mathcal{P} eine a priori Verteilung zu definieren. Ein häufig angewendetes Konzept ist der von [Ferguson \(1973\)](#) vorgestellte Dirichlet Prozess. Zum Dirichlet Prozess alternative Konzepte werden zum Beispiel in [Dubins und Freedman \(1963\)](#), [Kraft und van Eeden \(1964\)](#), [Kraft \(1964\)](#), [Dubins und Freedman \(1967\)](#) untersucht. Eine detaillierte Beschreibung der Eigenschaften des Dirichlet Prozesses sowie dessen Anwendung in den unterschiedlichsten Modellen findet man zum Beispiel in [Blackwell und MacQueen \(1973\)](#), [Fabius \(1973a\)](#), [Fabius \(1973b\)](#), [Korwar und Hollander \(1973\)](#), [Doksum \(1974\)](#), [Antoniak \(1974\)](#), [Ghorai und Rubin \(1982\)](#), [Lo \(1984\)](#) und [Rolin \(1992\)](#). Da der Dirichlet Prozess ein wesentlicher Bestandteil dieser Arbeit ist, wird er im Abschnitt 5.4.2 ausführlicher eingeführt.

5.4.1 Zufälliges Wahrscheinlichkeitsmaß

Es sei $(\mathcal{X}, \mathcal{B})$ wie bisher ein polnischer Raum mit der induzierten Borel- σ -Algebra, ein Wahrscheinlichkeitsmaß P auf \mathcal{B} ist dann ein Element der Menge $[0, 1]^{\mathfrak{B}}$ aller Abbildung von \mathfrak{B} in das Intervall $[0, 1]$. Möchten wir das zufällige Wählen eines Elementes aus der Menge $\mathcal{P} \subset [0, 1]^{\mathfrak{B}}$ aller Wahrscheinlichkeitsmaße auf \mathcal{B} beschreiben, so ist es nötig, eine σ -Algebra auf \mathcal{P} zu definieren.

σ -Algebra bzgl. der Zylindermengen

Für einen beliebigen messbaren Raum (Ω, \mathfrak{A}) und eine nichtleere Indexmenge T definieren wir den Raum Ω^T aller Funktionen $f : T \rightarrow \Omega$. Für jede endliche Teilmenge $\{t_1, \dots, t_n\}_{n \in \mathbb{N}} \subset T$ ist Ω^n das kartesische Produkt $\times_{i=1}^n \Omega$ und \mathfrak{A}^n die zugehörige Produkt- σ -Algebra $\otimes_{i=1}^n \mathfrak{A}$. Eine Untermenge von Ω^T der Gestalt $\{f \in \Omega^T : (f(t_1), \dots, f(t_n)) \in A\}$, wobei $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$ und $A \in \mathfrak{A}^n$ ist, nennt man Zylindermenge in Ω^T . Die durch die Zylindermengen in Ω^T erzeugte σ -Algebra bezeichnen wir mit \mathfrak{A}^T . Betrachten wir andererseits für alle $n \in \mathbb{N}$ und $t_1, \dots, t_n \in T$ die natürlichen Projektionen $\pi_{t_1 \dots t_n}$ von Ω^T nach Ω^n mit $\pi_{t_1 \dots t_n}(f) := (f(t_1), \dots, f(t_n))$, dann ist \mathfrak{A}^T die kleinste σ -Algebra, so dass für alle $n \in \mathbb{N}$ und $t_1, \dots, t_n \in T$ die Abbildung $\pi_{t_1 \dots t_n}$ von Ω^T nach Ω^n eine \mathfrak{A}^T - \mathfrak{A}^n -messbare Funktion ist. Nehmen wir weiterhin an, dass Ω ein separabler metrischer Raum und \mathfrak{A} die induzierte Borel- σ -Algebra sei. Dann ist die zugehörige Produkt- σ -Algebra \mathfrak{A}^n das n -fache kartesische Produkt $\mathfrak{A} \times \dots \times \mathfrak{A}$ für alle $n \in \mathbb{N}$ (vgl. [Billingsley \(1968\)](#), S. 224). Damit lässt sich jede Zylindermenge in Ω^T als $\bigcap_{i=1}^n \{f \in \Omega^T : f(t_i) \in A_i\}$ darstellen, wobei $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$ und $A^n = A_1 \times \dots \times A_n \in \mathfrak{A}^n$ ist. \mathfrak{A}^T entspricht somit der kleinsten σ -Algebra, so dass für alle $t \in T$ die Abbildung $\pi_t(f) = f(t)$ von Ω^T nach Ω eine \mathfrak{A}^T - \mathfrak{A} -messbare Funktion ist. Für eine Teilmenge $\mathcal{F} \subset \Omega^T$ ist weiterhin die Spur $\mathfrak{A}_{\mathcal{F}}^T := \mathfrak{A}^T \cap \mathcal{F}$ die durch die Zylindermengen in \mathcal{F} erzeugte σ -Algebra und für alle $t \in T$ ist die Abbildung $\pi_t(f) = f(t)$ von \mathcal{F} nach Ω eine $\mathfrak{A}_{\mathcal{F}}^T$ - \mathfrak{A} -messbare Funktion.

Bemerkung 5.5 Alternativ zu dem vorgestellten Konzept der Definition einer σ -Algebra bzgl. der Zylindermengen ist es auch möglich, die induzierte Borel- σ -Algebra bzgl. der schwachen bzw. starken Topologie zu betrachten. Die σ -Algebra bzgl. der Zylindermengen ist dagegen im Allgemeinen nicht durch eine Topologie induziert, so dass im Allgemeinen die Existenz einer regulären bedingten Verteilung auf der bzgl. der Zylindermengen induzierten σ -Algebra nicht gesichert ist. Ist dagegen die

schwache bzw. starke Topologie metrisierbar und der metrische Raum \mathcal{P} weiterhin polnisch, dann ist zu mindestens die Existenz einer regulären bedingten Verteilung auf der induzierten Borel- σ -Algebra gesichert.

Borel- σ -Algebra bzgl. der starken Topologie: Für zwei Maße $P, Q \in \mathcal{P}$ definiert

$$d_V(P, Q) := \sup_{B \in \mathfrak{B}} |P(B) - Q(B)| \quad (5.2)$$

eine Metrik auf \mathcal{P} . d_V wird Variationsabstand genannt. Es gilt offensichtlich $0 \leq d_V(P, Q) \leq 1$ für alle $P, Q \in \mathcal{P}$. Falls eine Folge P_1, P_2, \dots von Wahrscheinlichkeitsmaßen nun im Variationsabstand gegen ein Wahrscheinlichkeitsmaß P konvergiert, so heißt sie stark konvergent. Weiterhin bezeichnet \mathfrak{T}_V bzw. \mathfrak{B}_V die durch den Variationsabstand d_V induzierte starke Topologie bzw. Borel- σ -Algebra auf \mathcal{P} . Insbesondere ist der metrische Raum (\mathcal{P}, d_V) genau dann separabel, wenn \mathcal{P} dominiert und \mathfrak{B} abzählbar erzeugt ist (Strasser (1985), S. 98, Satz 21.3).

Borel- σ -Algebra bzgl. der schwachen Topologie: Für $\varepsilon > 0$ und $B \in \mathfrak{B}$ sei

$$B^\varepsilon := \{y \in \mathcal{X} \mid d_X(x, y) < \varepsilon, \quad \text{für ein } x \in B\},$$

dann definiert für zwei Maße $P, Q \in \mathcal{P}$

$$d_P(P, Q) := \inf_{\varepsilon > 0} \{\varepsilon \mid P(B) \leq Q(B^\varepsilon) + \varepsilon, \quad \text{für alle } B \in \mathfrak{B}\} \quad (5.3)$$

eine Metrik auf \mathcal{P} , falls \mathcal{X} separabel ist (vgl. Billingsley (1968), S. 236 ff.). d_P wird Prohorov Abstand genannt und für alle $P, Q \in \mathcal{P}$ gilt $0 \leq d_P(P, Q) \leq 1$ (vgl. Dudley (1989)). Weiterhin bezeichnet \mathfrak{T}_P bzw. \mathfrak{B}_P die durch d_P induzierte Topologie bzw. Borel- σ -Algebra auf \mathcal{P} . Ferner sei $C_b(\mathcal{X}, \mathfrak{B})$ der Vektorraum aller beschränkten, stetigen, reellen Funktionen auf \mathcal{X} . Dann heißt eine Folge P_1, P_2, \dots von Wahrscheinlichkeitsmaßen schwach konvergent gegen ein Wahrscheinlichkeitsmaß P , falls gilt

$$\lim_{n \rightarrow \infty} \int f dP_n = \int f dP, \quad \text{für alle } f \in C_b(\mathcal{X}, \mathfrak{B}).$$

Die zugrunde liegende Topologie auf der Menge \mathcal{P} bezeichnen wir als Topologie der schwachen Konvergenz. Ferner sei \mathcal{X} ein separabel metrischer Raum. Dann ist (\mathcal{P}, d_P) ein separabel metrischer Raum und \mathfrak{T}_P ist die Topologie der schwachen Konvergenz (vgl. Billingsley (1968), S. 236 ff.). Nehmen wir zusätzlich an, dass \mathcal{X} ein vollständiger metrischer Raum sei. Dann ist die Menge aller Wahrscheinlichkeitsmaße \mathcal{P} auf \mathcal{X} vollständig bzgl. d_P (vgl. Prohorov (1956)). \square

Betrachten wir nun die Menge $[0, 1]^{\mathfrak{B}}$ der Abbildungen von der Borel- σ -Algebra \mathfrak{B} über dem polnischen Raum \mathcal{X} in das Intervall $[0, 1]$, wobei das Intervall $[0, 1]$ mit der Borel- σ -Algebra \mathfrak{L} versehen sei. Eine endliche Teilmenge $N = \{B_1, \dots, B_m\}$ von \mathfrak{B} ist dann eine endliche Folge von Mengen B_1, \dots, B_m aus \mathfrak{B} . Insbesondere hat nun eine Zylindermenge für $A \in \mathfrak{L}^m$ die Gestalt

$$\{P \in [0, 1]^{\mathfrak{B}} \mid (P(B_1), \dots, P(B_m)) \in A\}.$$

Die durch die Zylindermengen auf $[0, 1]^{\mathfrak{B}}$ erzeugte σ -Algebra bezeichnen wir mit $\mathfrak{L}^{\mathfrak{B}}$. Der Raum $[0, 1]$ versehen mit dem euklidischen Abstand ist ferner separabel, so dass $\mathfrak{L}^{\mathfrak{B}}$ schon als die kleinste σ -Algebra, für die die Abbildung $P(B)$ von $[0, 1]^{\mathfrak{B}}$ nach $[0, 1]$ für alle $B \in \mathfrak{B}$ eine $\mathfrak{L}^{\mathfrak{B}}$ - \mathfrak{L} -messbare Funktion ist, eindeutig definiert. Für eine Menge \mathcal{P}_0 von Wahrscheinlichkeitsmaßen $\mathcal{P}_0 \subset [0, 1]^{\mathfrak{B}}$ und die Spur $\mathfrak{L}_{\mathcal{P}_0}^{\mathfrak{B}} = \mathfrak{L}^{\mathfrak{B}} \cap \mathcal{P}_0$ von $\mathfrak{L}^{\mathfrak{B}}$ in \mathcal{P}_0 charakterisieren wir nun die Verteilung P einer Zufallsvariable P_0 mit Werten in $(\mathcal{P}_0, \mathfrak{L}_{\mathcal{P}_0}^{\mathfrak{B}})$.

Es sei Ω ein polnischer Raum und \mathfrak{A} die induzierte Borel- σ -Algebra. Ferner sei \mathbf{f} eine Zufallsvariable mit Werten in dem messbaren Raum $(\Omega^T, \mathfrak{A}^T)$ und der Verteilung P auf \mathfrak{A}^T . Insbesondere ist für alle $t \in T$ die Abbildung $\pi_t(\mathbf{f}) = f(t)$ von \mathcal{F} nach Ω eine \mathfrak{A}_t^T - \mathfrak{A} -messbare Funktion. Wir können somit \mathbf{f} auch als Familie $\{\mathbf{f}(t) := \pi_t(\mathbf{f}) \mid t \in T\}$ von Zufallsvariablen mit Werten in Ω auffassen. $\mathbf{f} = \{\mathbf{f}(t) \mid t \in T\}$ bezeichnen wir als zufällige Funktion oder stochastischen Prozess. Für eine Teilmenge $\{t_1, \dots, t_n\}_{n \in \mathbb{N}}$ von T bezeichnen wir das Maß $P_{t_1 \dots t_n}(A) := P(\pi_{t_1 \dots t_n}^{-1}(A))$ für $A \in \mathfrak{A}^n$ als Projektion oder endlich dimensionale Verteilung des Maßes P auf $(\Omega^n, \mathfrak{A}^n)$. Von zentraler Bedeutung für das Studium der zufälligen Funktion \mathbf{f} ist die Familie $\{P_{t_1 \dots t_n} \mid t_1, \dots, t_n \in T, n \geq 1\}$ der endlich dimensionalen Verteilungen, insbesondere genügen sie den Verträglichkeitsbedingungen des folgenden Satzes von Kolmogoroff.

Satz 5.6 (Kolmogoroff (1933))

Sei Ω ein vollständiger separabler metrischer Raum, T eine beliebige nichtleere Indexmenge und P_{t_1, \dots, t_n} , $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$, eine Familie von Wahrscheinlichkeitsmaßen auf den messbaren Räumen $(\Omega^n, \mathfrak{A}^n)$, die den Bedingungen

1. (Verträglichkeit)

$$P_{t_1, \dots, t_{n+m}}(A \times \Omega^m) = P_{t_1, \dots, t_n}(A), \quad A \in \mathfrak{A}^n$$

2. (Symmetrie)

$$P_{t_1, \dots, t_n}(A_1 \times \dots \times A_n) = P_{t_{i(1)}, \dots, t_{i(n)}}(A_{i(1)} \times \dots \times A_{i(n)}),$$

für beliebige $n \in \mathbb{N}$, $t_1, \dots, t_n \in T$, $A_1, \dots, A_n \in \mathfrak{A}$ und jede Permutation $(i(1), \dots, i(n))$

genügen. Dann existiert genau ein Wahrscheinlichkeitsmaß P auf $(\Omega^T, \mathfrak{A}^T)$, so dass für jede endliche Teilmenge $\{t_1, \dots, t_n\}$ von T die Projektion von P auf $(\Omega^n, \mathfrak{A}^n)$ mit $P_{t_1 \dots t_n}$ übereinstimmt.

Existieren das erste und zweite Moment der endlichdimensionalen Verteilungen, so bezeichnen wir eine Abbildung $\mathbb{E} \mathbf{f}$ von T nach Ω als Erwartungswertfunktion, wenn $\mathbb{E} \mathbf{f}(t) = \mathbb{E}[\mathbf{f}(t)]$ für alle $t \in T$ gilt, und eine Abbildung $\text{Cov} \mathbf{f}$ von $T \times T$ nach \mathbb{R}^+ als Kovarianzfunktion, falls $\text{Cov} \mathbf{f}(t, s) = \text{Cov}(\mathbf{f}(t), \mathbf{f}(s))$ für alle $t, s \in T$ gilt.

Kehren wir nun zu dem vorgestellten messbaren Raum $([0, 1]^{\mathfrak{B}}, \mathfrak{L}^{\mathfrak{B}})$ zurück. Der Satz von Kolmogoroff zeigt, dass eine Verteilung P auf $\mathfrak{L}^{\mathfrak{B}}$ eindeutig durch eine Familie von endlich dimensionalen Verteilungen festgelegt ist, falls diese die vorgestellten Regularitätsbedingungen erfüllt. Insbesondere interessieren wir uns nicht für eine beliebige zufällige Funktion \mathbf{f} mit Werten in $([0, 1]^{\mathfrak{B}}, \mathfrak{L}^{\mathfrak{B}})$, sondern wir möchten ein zufälliges Wahrscheinlichkeitsmaß charakterisieren, d.h. eine zufällige Funktion \mathbf{P} mit Werten in $([0, 1]^{\mathfrak{B}}, \mathfrak{L}^{\mathfrak{B}})$ und einer Verteilung $P^{\mathbf{P}}$ auf $\mathfrak{L}^{\mathfrak{B}}$, die konzentriert ist auf $\mathfrak{L}_{\mathfrak{P}}^{\mathfrak{B}}$.

Definition 5.7 (Ferguson (1973), Doksum (1974), Simar (1984))

Eine zufällige Funktion \mathbf{P} mit Werten in $([0, 1]^{\mathfrak{B}}, \mathfrak{L}^{\mathfrak{B}})$ und der Verteilung $P^{\mathbf{P}}$ auf $\mathfrak{L}^{\mathfrak{B}}$ heißt zufälliges Wahrscheinlichkeitsmaß, falls

1. $\mathbf{P}(\mathcal{X}) = 1$ $P^{\mathbf{P}}$ -f.s. gilt.
2. Für jede fallende Folge messbarer Mengen $(B_n)_{n \in \mathbb{N}}$ mit $\lim_{n \rightarrow \infty} B_n = \emptyset$ gilt $\lim_{n \rightarrow \infty} \mathbf{P}(B_n) = 0$ $P^{\mathbf{P}}$ -f.s.
3. Die Verteilung des zufälligen Vektors $(\mathbf{P}(B_1), \dots, \mathbf{P}(B_k))$ entspricht der Verteilung des zufälligen Vektors $(\sum_i \mathbf{P}(B_{1,i}), \dots, \sum_i \mathbf{P}(B_{k,i}))$, wobei
 - B_1, \dots, B_k eine messbare Partition von \mathcal{X} ,
 - $B_{j,1}, \dots, B_{j,k_j}$ eine messbare Partition von B_j für alle $j = 1, \dots, k$ ist.

Andererseits sei $\{\mathbf{P}(B) \mid B \in \mathfrak{B}\}$ eine Familie von Zufallsvariablen $\mathbf{P}(B)$ mit Werten in $[0, 1]$ für alle $B \in \mathfrak{B}$, die die Bedingungen 1.-3. der Definition 5.7 erfüllt. So existiert eine eindeutig bestimmte Verteilung $P^{\mathbf{P}}$ auf $\mathfrak{L}^{\mathfrak{B}}$, so dass \mathbf{P} eine zufällige Funktion mit der Verteilung $P^{\mathbf{P}}$ ist (vgl. Ferguson (1973)).

5.4.2 Dirichlet Prozess

Dirichletverteilung

Es sei $(\mathcal{X}, \mathfrak{B}, \mu)$ ein Wahrscheinlichkeitsraum mit Borel- σ -Algebra \mathfrak{B} und Wahrscheinlichkeitsmaß μ . Ferner seien α eine positive reelle Zahl sowie $\{B_1, \dots, B_k\}$ ($k \in \mathbb{N}$) eine messbare Partition von \mathcal{X} , so dass $\sum_{i=1}^k \mu(B_i) = 1$ gilt. Für jedes $i = 1, \dots, k$ definieren wir eine gammaverteilte Zufallsvariable

$$\eta_i \sim Ga(1, \alpha\mu(B_i)),$$

wobei eine reellwertige Zufallsvariable der Gammaverteilung $Ga(b, p)$ mit den Parameter $b > 0$ und $p > 0$ genügt, wenn sie die Dichte

$$f(x) = \begin{cases} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} & : x > 0, \\ 0 & : x \leq 0 \end{cases}$$

bzgl. des Lebesgueschen Maßes besitzt (vgl. Müller (1991)). Dabei bezeichnet Γ die Gamma-Funktion mit

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt, \quad x > 0.$$

Eine $Ga(1, 0)$ -verteilte Zufallsgröße ist per Definition degeneriert bei Null. Für alle $i = 1, \dots, k$ gilt damit $\mathbb{E} \eta_i = \alpha\mu(B_i)$ und $\text{Var} \eta_i = \alpha\mu(B_i)$. Weiterhin nehmen wir an, dass die Zufallsgrößen η_1, \dots, η_k unabhängig sind und definieren die Zufallsgröße

$$\boldsymbol{\eta} := \sum_{i=1}^k \eta_i.$$

Die Zufallsgröße $\boldsymbol{\eta}$ ist dann gammaverteilt mit $\boldsymbol{\eta} \sim Ga(1, \alpha)$, da eine abzählbare Summe von unabhängigen gammaverteilten Zufallsgrößen mit Parametern (b, p_i) gammaverteilt mit den Parametern $(b, \sum_{i \geq 1} p_i)$ ist. Somit sind der Erwartungswert und die Varianz von $\boldsymbol{\eta}$ gleich α . Es sei außerdem eine Folge von reellwertigen Zufallsvariablen ξ_1, \dots, ξ_k durch

$$\xi_i := \frac{\eta_i}{\boldsymbol{\eta}} \quad \text{für } i = 1, \dots, k$$

definiert. Die Zufallsgrößen $\boldsymbol{\eta}, \xi_1, \dots, \xi_k$ sind dann unabhängig und es gilt offensichtlich

$$\sum_{i=1}^k \xi_i = 1.$$

Die Zufallsgrößen ξ_i sind auf dem Intervall $[0, 1]$ betaverteilt

$$\xi_i \sim Beta(\alpha\mu(B_i), \alpha\mu(B_i^c)) \quad \text{mit } B_i^c = \mathcal{X} \setminus B_i$$

für alle $i = 1, \dots, k$. Dabei heißt eine Zufallsgröße betaverteilt $Beta(p, q)$ mit den Parameter $p > 0$ und $q > 0$ auf dem Intervall $[a, b]$, wenn sie die Dichte

$$f(x) = \begin{cases} \frac{(b-a)^{1-p-q}}{B(p, q)} (x-a)^{p-1} (b-x)^{q-1} & : a < x < b, \\ 0 & : x \leq 0 \end{cases}$$

bzgl. des Lebesgueschen Maßes besitzt (vgl. Müller (1991)). B bezeichnet hierbei die Beta-Funktion mit

$$B(p, q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt, \quad p > 0, \quad q > 0.$$

Per Definition ist eine $Beta(p, 0)$ - bzw. $Beta(0, q)$ -Zufallsgröße degeneriert bei Eins bzw. Null. Damit gilt für $1 \leq i < j \leq k$

$$\mathbb{E} \xi_i = \mu(B_i), \quad \text{Var} \xi_i = \frac{\mu(B_i)(1-\mu(B_i))}{1+\alpha} \quad \text{und} \quad \text{Cov}(\xi_i, \xi_j) = \frac{\mu(B_i)\mu(B_j)}{1+\alpha}.$$

Die gemeinsame Verteilung der Zufallsgrößen ξ_1, \dots, ξ_k bezeichnen wir als Dirichletverteilung

$$\mathcal{D}(\alpha\mu(B_1), \dots, \alpha\mu(B_k)),$$

wobei ein Zufallsvektor $(\mathbf{X}_1, \dots, \mathbf{X}_k)$ der Dirichletverteilung $\mathcal{D}(p_1, \dots, p_k)$ mit den positiven Parametern (p_1, \dots, p_k) genügt, wenn $\sum_{i=1}^k \mathbf{X}_i = 1$ gilt und die Verteilung des $(k-1)$ -dimensionalen Zufallsvektors $(\mathbf{X}_1, \dots, \mathbf{X}_{k-1})$ die Dichte

$$f(x_1, \dots, x_{k-1}) = \begin{cases} \frac{\Gamma(\sum_{i=1}^k p_i)}{\prod_{i=1}^k \Gamma(p_i)} \prod_{i=1}^{k-1} x_i^{p_i-1} (1 - \sum_{i=1}^{k-1} x_i)^{p_k-1} & : 0 \leq x_1, \dots, x_{k-1}, \sum_{i=1}^{k-1} x_i \leq 1, \\ 0 & : \text{sonst} \end{cases}$$

bzgl. des Lebesgueschen Maßes auf dem \mathbb{R}^{k-1} besitzt (vgl. Müller (1991)). Zusätzlich sei die Zufallsgröße \mathbf{X}_i degeneriert an der Stelle 0, falls der zugehörige Parameter $p_i = 0$ ist.

Dirichlet Prozess

Die im Folgenden definierte Familie $\{\mathbf{P}(B) \mid B \in \mathfrak{B}\}$ von Zufallsgrößen $\mathbf{P}(B)$ mit Werten in $[0, 1]$ für alle $B \in \mathfrak{B}$ erfüllt die Bedingungen 1.-3. der Definition 5.7 eines zufälligen Wahrscheinlichkeitsmaßes (vgl. Ferguson (1973)). Somit definiert die Familie $\{\mathbf{P}(B) \mid B \in \mathfrak{B}\}$ eine eindeutig bestimmte Verteilung $P^{\mathbf{P}}$ auf $\mathfrak{L}^{\mathfrak{B}}$. Insbesondere ist \mathbf{P} dann eine zufällige Funktion (stochastischer Prozess) mit der Verteilung $P^{\mathbf{P}}$.

Definition 5.8 (Dirichlet Prozess)

Es sei α eine positive reelle Zahl und μ ein Wahrscheinlichkeitsmaß auf $(\mathcal{X}, \mathfrak{B})$. Ein Familie von Zufallsgrößen $\{\mathbf{P}(B) \mid B \in \mathfrak{B}\}$ wird Dirichlet Prozess

$$\mathbf{P} \sim \mathcal{D}_{\alpha\mu}$$

mit Parameter $\alpha\mu$ genannt, falls für jede messbare Partition B_1, \dots, B_k , $k \in \mathbb{N}$, von \mathcal{X} der zufällige Vektor

$$(\mathbf{P}(B_1), \dots, \mathbf{P}(B_k)) \sim \mathcal{D}(\alpha\mu(B_1), \dots, \alpha\mu(B_k))$$

dirichletverteilt mit den Parameter $(\alpha\mu(B_1), \dots, \alpha\mu(B_k))$ ist.

Sei \mathbf{P} ein Dirichlet Prozess auf $(\mathcal{X}, \mathfrak{B})$ mit den Parameter α und μ . Dann gilt für die Erwartungswertfunktion $\mathbb{E} \mathbf{P}(B) = \mu(B)$ für alle $B \in \mathfrak{B}$ und die Kovarianzfunktion $\text{Cov} \mathbf{P}$ ist für alle $B_1, B_2 \in \mathfrak{B}$ gegeben durch $\text{Cov} \mathbf{P}(B_1, B_2) = \frac{\mu(B_1)\mu(B_2)}{1+\alpha}$ falls $B_1 \neq B_2$ bzw. $\text{Cov} \mathbf{P}(B_1, B_1) = \frac{\mu(B_1)(1-\mu(B_1))}{1+\alpha}$ falls $B_1 = B_2$ (vgl. Ferguson (1973)).

5.4.3 Bayessche Schätzung

Wir betrachten nun das vorgestellte nichtparametrische Modell (vgl. Abschnitt 5.1). Dazu sei \mathbf{P} ein zufälliges Wahrscheinlichkeitsmaß und unter der Bedingung $\mathbf{P} = P$ beschreibe die Zufallsvariable \mathbf{X}^n das statistische Experiment $(\mathcal{X}^n, \mathfrak{B}^n, P^n | P \in \mathcal{P})$. A priori sei \mathbf{P} ein Dirichlet Prozess $\mathbf{P} \sim \mathcal{D}_{\alpha\mu}$, wobei α positiv und μ ein Wahrscheinlichkeitsmaß auf \mathcal{X} sei.

Satz 5.9 (Ferguson (1973))

Es sei eine Beobachtung $\mathbf{X}^n = x^n$ gegeben. Dann ist das zufällige Wahrscheinlichkeitsmaß \mathbf{P} unter der Bedingung $\mathbf{X}^n = x^n$ ein Dirichlet Prozess mit

$$\mathbf{P} | \mathbf{X}^n = x^n \sim \mathcal{D}_{\alpha_n \mu_n},$$

wobei $\alpha_n = \alpha + n$ und $\mu_n = \frac{\alpha}{\alpha+n}\mu + \frac{n}{\alpha+n}\hat{F}_{x^n}$ ist.

Als Bayessche Schätzung für die unbekannte Verteilung P betrachten wir die a posteriori Erwartungswertfunktion $\mathbb{E}[\mathbf{P} | \mathbf{X}^n = x^n]$.

Satz 5.10 (Ferguson (1973))

Es sei eine Beobachtung $\mathbf{X}^n = x^n$ gegeben. Dann ist die Bayessche Schätzung

$$\hat{P}(x^n) = \frac{\alpha}{\alpha+n}\mu + \frac{n}{\alpha+n}\hat{F}_{x^n}.$$

Bemerkung 5.11 Wählen wir die quadratische Verlustfunktion

$$L_Q(P, \hat{P}) := \sup_{B \in \mathfrak{B}} (P(B) - \hat{P}(B))^2,$$

dann wird das a posteriori Risiko

$$\mathbb{E}[L_Q(\mathbf{P}, \hat{P}) | \mathbf{X}^n = x^n] = \sup_{B \in \mathfrak{B}} \mathbb{E}[(\mathbf{P}(B_i) - \hat{P}(B_i))^2 | \mathbf{X}^n = x^n]$$

durch die a posteriori Erwartungswertfunktion $\mathbb{E}[\mathbf{P} | \mathbf{X}^n = x^n]$ minimiert, falls die Zufallsvariablen $\mathbf{P}(B) | \mathbf{X}^n = x^n$ für alle $B \in \mathfrak{B}$ ein endliches zweites Moment besitzen. \square

5.4.4 Selbstinformativer Grenzwert

Im Folgenden werden wir die im Kapitel 4 vorgestellte iterative Prozedur anwenden. Es sei \mathbf{P} ein zufälliges Wahrscheinlichkeitsmaß und unter der Bedingung $\mathbf{P} = P$ seien die Zufallsvariablen $\mathbf{X}_1^n, \dots, \mathbf{X}_r^n$ unabhängig und jede Zufallsvariable \mathbf{X}_i^n beschreibe das statistische Experiment

$$(\mathcal{X}^n, \mathfrak{B}^n, P^n | P \in \mathcal{P}).$$

Insbesondere beschreibt somit unter der Bedingung $\mathbf{P} = P$ der Vektor $\mathbf{X}^{rn} = (\mathbf{X}_1^n, \dots, \mathbf{X}_r^n)$ das statistische Experiment $(\mathcal{X}^{rn}, \mathfrak{B}^{rn}, P^{rn} | P \in \mathcal{P})$. A priori nehmen wir weiterhin an, dass \mathbf{P} ein Dirichlet Prozess mit

$$\mathbf{P} \sim \mathcal{D}_{\alpha\mu}$$

sei, wobei α eine positive Zahl und μ ein Wahrscheinlichkeitsmaß ist. Damit ergibt sich direkt aus dem Satz 5.9, dass die a posteriori Verteilung $P^{rn} | \mathbf{X}^{rn} = x^{rn}$ von \mathbf{P} unter der Bedingung $\mathbf{X}^{rn} = x^{rn}$ durch einen Dirichlet Prozess mit den Parametern $\alpha + rn$ und $\frac{\alpha}{\alpha+rn}\mu + \frac{rn}{\alpha+rn}\hat{F}_{x^{rn}}$ beschrieben wird.

Für eine Beobachtung $\mathbf{X}^{rn} = x^{rn}$ bezeichnen wir mit $\hat{F}_{x^{rn}}$ die empirische Verteilung der Werte $x^{rn} = (x_1, \dots, x_{rn})$. Weiterhin ergibt sich aus dem Satz 5.10 die Bayessche Schätzung

$$\hat{P}(x^{rn}) := \frac{\alpha}{\alpha + rn} \mu + \frac{rn}{\alpha + rn} \hat{F}_{x^{rn}}$$

von P . Nehmen wir nun an, dass r -mal der Vektor $x^n \in \mathcal{X}^n$ beobachtet wurde

$$\mathbf{X}^{rn} = \mathbb{1}_r^t \otimes x^n = (x^n, \dots, x^n),$$

so ist die a posteriori Verteilung $P^{\mathbf{P}} | \mathbf{X}^{rn} = \mathbb{1}_r^t \otimes x^n$ von \mathbf{P} durch einen Dirichlet Prozess $\mathcal{D}_{\alpha_r, \mu_r}$ mit den Parametern $\alpha_r := \alpha + rn$ und $\mu_r := \frac{\alpha}{\alpha + rn} \mu + \frac{rn}{\alpha + rn} \hat{F}_{x^n}$ beschrieben und die Bayessche Schätzung $\hat{P}_r(\mathbb{1}_r^t \otimes x^n)$ von P ist gegeben durch

$$\hat{P}_r(\mathbb{1}_r^t \otimes x^n) = \frac{\alpha}{\alpha + rn} \mu + \frac{rn}{\alpha + rn} \hat{F}_{x^n}.$$

Satz 5.12 *Für alle $x^n \in \mathcal{X}^n$ ist die empirische Verteilung \hat{F}_{x^n} der selbstinformativ Grenzwert der Folge der Bayesschen Schätzungen $\hat{P}_1(x^n), \hat{P}_2(\mathbb{1}_2^t \otimes x^n), \dots$*

BEWEIS : Offensichtlicher Weise gilt die punktweise Konvergenz

$$\hat{F}_{x^n}(B) = \lim_{r \rightarrow \infty} \hat{P}_r(\mathbb{1}_r^t \otimes x^n)(B) \quad \text{für alle } B \in \mathfrak{B}.$$

Weiterhin folgt aus

$$\sup_{B \in \mathfrak{B}} |\hat{P}_r(\mathbb{1}_r^t \otimes x^n)(B) - \hat{F}_{x^n}(B)| = \frac{\alpha}{\alpha + rn} \sup_{B \in \mathfrak{B}} |\mu(B) + \hat{F}_{x^n}(B)| \leq 2 \frac{\alpha}{\alpha + rn}$$

die gleichmäßige Konvergenz für alle x^n im Variationsabstand. □

Kapitel 6

Semiparametrisches lineares Modell

6.1 Modellannahmen

Es seien $\mathbf{X}_1, \dots, \mathbf{X}_n$ unabhängige Zufallsvariablen mit Werten in dem messbaren Raum $(\mathbb{R}^p, \mathfrak{L}^p)$, wobei \mathfrak{L}^p die Borel- σ -Algebra über \mathbb{R}^p bezeichnet. Die Vektoren $\mathbf{X}_1, \dots, \mathbf{X}_n$ fassen wir als Zeilenvektoren auf, deren Randverteilung durch

$$\mathbb{E} \mathbf{X}_i = Z_i B, \quad \text{Cov} \mathbf{X}_i = \Sigma, \quad \mathbf{X}_i \sim P_i, \quad i = 1, \dots, n, \quad (6.1)$$

festgelegt sei. Die $n \times k$ Matrix Z mit den Zeilen Z_i besitze den vollen Spaltenrang $k < n$ und sei fest vorgegeben. Weiterhin enthalte der Spaltenraum $R(Z)$ den n -dimensionalen Einheitsvektor $\mathbb{1}_n$. Ferner bezeichnen wir mit \mathbf{X} die zufällige $n \times p$ Matrix $(\mathbf{X}_1^t, \dots, \mathbf{X}_n^t)^t$. Die $k \times p$ Matrix B , die $p \times p$ Matrix Σ und die Randverteilungen P_1, \dots, P_n auf \mathfrak{L}^p sind unbekannt.

In den nächsten beiden Abschnitten werden wir die vMLS nach Kiefer und Wolfowitz sowie nach Gill bestimmen. Dazu wählen wir die folgende Parametrisierung

$$\mathbf{X}_i = Z_i B + U_i \quad \text{für } i = 1, \dots, n.$$

Die p -dimensionalen Zufallsvariablen U_1, \dots, U_n seien unabhängig mit identischer Randverteilung G auf \mathfrak{L}^p . Die Verteilung G variiert in der Menge \mathcal{P}_+^0 aller Wahrscheinlichkeitsverteilungen auf \mathfrak{B} mit Erwartungswert 0 und endlichem zweiten Moment. Offensichtlich ist für $i = 1, \dots, n$ die Randverteilung von \mathbf{X}_i eindeutig durch den Parameter $\theta = (B, G)$ und den Vektor Z_i bestimmt, so dass wir die Randverteilung der Zufallsvariable \mathbf{X}_i mit P_{θ, Z_i} für $i = 1, \dots, n$ und die Verteilung der Zufallsvariable \mathbf{X} mit $P_{\theta, Z} := \prod_{i=1}^n P_{\theta, Z_i}$ bezeichnen. Der Parameterraum ist dann gegeben durch $\Theta = \mathbb{R}^{k \times p} \times \mathcal{P}_+^0$ und die zufällige Matrix \mathbf{X} beschreibt das statistische Experiment

$$(\mathbb{R}^{n \times p}, \mathfrak{B}^{n \times p}, P_{\theta, Z} \mid \theta \in \Theta). \quad (6.2)$$

Weiterhin sind wir an dem unbekanntem Parameter Σ interessiert. Wählen wir die vorgestellte Parametrisierung $\theta \in \Theta$, so ist Σ ein abgeleiteter Parameter

$$\Sigma = \varphi(\theta).$$

Existiert insbesondere eine vMLS $\hat{\theta}$ im Sinne von Kiefer und Wolfowitz bzw. Gill, so ist durch

$$\hat{\Sigma} = \varphi(\hat{\theta}) \quad (6.3)$$

eine vMLS für Σ festgelegt.

Bemerkung 6.1 Das semiparametrische lineare Modell ist im Allgemeinen kein Spezialfall des im Kapitel 5 vorgestellten nichtparametrischen Modells, da die gemeinsame Verteilung $\prod_{i=1}^n P_{\theta, Z_i}$ des Vektors $\mathbf{X} = (\mathbf{X}_1^t, \dots, \mathbf{X}_n^t)^t$ sich mit Ausnahme einiger Spezialfälle nicht mehr als Produkt $\prod_{i=1}^n P$ identischer Randverteilungen P darstellen lässt. Somit ist ein Übertragen der im Kapitel 5 erzielten Ergebnisse nicht direkt möglich. Einen Spezialfall in dem das semiparametrische lineare Modell eine Verteilungsfamilie definiert, die in der Verteilungsfamilie des nichtparametrischen Modells enthalten ist, betrachten wir im Abschnitt 6.5. \square

6.2 Ansatz von Kiefer und Wolfowitz

Es sei \mathbf{X} eine das statistische Experiment (vgl. Abschnitt 6.1)

$$(\mathbb{R}^{n \times p}, \mathfrak{B}^{n \times p}, \prod_{i=1}^n P_{\theta, Z_i} | \theta \in \Theta)$$

beschreibende zufällige $n \times p$ Matrix.

In Analogie zum nichtparametrischen Modell (vgl. Abschnitt 5.2) definieren wir für alle $x = (x_1^t, \dots, x_n^t)^t \in \mathbb{R}^{n \times p}$ durch

$$L_x(B, G) := \prod_{i=1}^n G(\{x_i - Z_i B\}) \quad \text{für } B \in \mathbb{R}^{k \times p} \text{ und } G \in \mathcal{P}$$

eine „Likelihood-Funktion“. Wir werden nachweisen, dass $\hat{\theta}(x) \in \Theta$ genau dann eine vMLS nach Kiefer und Wolfowitz ist, wenn gilt

$$L_x(\hat{\theta}(x)) = \max_{\theta \in \Theta} L_x(\theta).$$

Ferner bezeichnen wir für alle $x \in \mathbb{R}^{n \times p}$ und $B \in \mathbb{R}^{k \times p}$ mit $\hat{F}_{B,x}$ die empirische Verteilung der Werte $(x_1 - Z_1 B, \dots, x_n - Z_n B)$ sowie mit $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ und $\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i$.

Lemma 6.2 Für alle $x \in \mathbb{R}^{n \times p}$ ist $\hat{\theta}(x) = (\hat{B}, \hat{G}) \in \Theta$ genau dann eine Lösung der Gleichung

$$L_x(\hat{\theta}(x)) = \max_{\theta \in \Theta} L_x(\theta),$$

wenn gilt

1. $L_x(\hat{B}, \hat{F}_{\hat{B},x}) = \max\{L_x(B, \hat{F}_{B,x}) \mid \bar{x} = \bar{Z}B, B \in \mathbb{R}^{k \times p}\}$,
2. \hat{G} ist die empirische Verteilung $\hat{F}_{\hat{B},x}$ der Residuen $(x_1 - Z_1 \hat{B}, \dots, x_n - Z_n \hat{B})$.

BEWEIS : Es sei $x = (x_1^t, \dots, x_n^t)^t \in \mathbb{R}^{n \times p}$ beliebig fixiert. Für alle $\theta = (B, G) \in \Theta$ gilt nun mit Lemma 5.1

$$L_x(B, \hat{F}_{B,x}) \geq L_x(B, G), \tag{6.4}$$

wobei die Gleichheit nur für $G = \hat{F}_{B,x}$ erfüllt ist. Im Allgemeinen ist $\hat{F}_{B,x}$ nicht in der Menge \mathcal{P}_+^0 enthalten. Insbesondere gilt offensichtlich $\hat{F}_{B,x} \in \mathcal{P}_+^0$ genau dann, wenn $\bar{x} = \bar{Z}B$ gilt. O.B.d.A. sei $\bar{x} \neq \bar{Z}B$, dann folgt $\bar{x} = \bar{Z}\tilde{B}$ für $\tilde{B} = B + (Z^t Z)^{-1} Z^t \mathbb{1}_n(\bar{x} - \bar{Z}B)$ und weiterhin

$$L_x(\tilde{B}, \hat{F}_{\tilde{B},x}) = L_x(B, \hat{F}_{B,x}) \geq L_x(B, G), \tag{6.5}$$

wobei die Gleichheit nur für $(B, G) = (\tilde{B}, \hat{F}_{\tilde{B},x})$ gilt. Damit folgt die Behauptung. \square

Satz 6.3 *Es sei eine Beobachtung $\mathbf{X} = x$ mit $\mathbf{X} \sim P_{\theta, Z}$ und $\theta \in \Theta$ gegeben. Eine Schätzung $\hat{\theta}(x) = (\hat{B}(x), \hat{G}(x))$ ist genau dann eine verallgemeinerte Maximum-Likelihood-Schätzung nach Kiefer und Wolfowitz, wenn sie die folgenden Bedingungen erfüllt:*

1. $L_x(\hat{B}(x), \hat{F}_{\hat{B}(x), x}) = \max\{L_x(B, \hat{F}_{B, x}) \mid \bar{x} = \bar{Z}B, B \in \mathbb{R}^{k \times p}\}$,
2. $\hat{G}(x)$ ist die empirische Verteilung $\hat{F}_{\hat{B}(x), x}$ der Residuen $(x_1 - Z_1 \hat{B}(x), \dots, x_n - Z_n \hat{B}(x))$.

BEWEIS : Es sei $x = (x_1^t, \dots, x_n^t)^t$ beliebig fixiert und es gelten für $\hat{\theta}(x)$ die 1. und 2. Bedingung, dann ist für alle $\theta \in \Theta$ zu zeigen (vgl. Definition 3.1)

$$\frac{dP_{\hat{\theta}(x), Z}}{d(P_{\hat{\theta}(x), Z} + P_{\theta, Z})}(x) \geq \frac{dP_{\theta, Z}}{d(P_{\hat{\theta}(x), Z} + P_{\theta, Z})}(x).$$

In Analogie zum Beweis des Satzes 5.2 genügt es zu zeigen, dass für alle $\theta \in \Theta$

$$\prod_{i=1}^n \frac{dP_{\hat{\theta}(x), Z_i}}{d(P_{\hat{\theta}(x), Z_i} + P_{\theta, Z_i})}(x_i) \geq \prod_{i=1}^n \frac{dP_{\theta, Z_i}}{d(P_{\hat{\theta}(x), Z_i} + P_{\theta, Z_i})}(x_i)$$

gilt, wenn für alle $\theta \in \Theta$ die Radon-Nikodym Dichte $\prod_{i=1}^n \frac{dP_{\hat{\theta}(x), Z_i}}{d[P_{\hat{\theta}(x), Z_i} + P_{\theta, Z_i}]}$ an der Stelle x eindeutig festgelegt und positiv ist.

Nun ist für $i = 1, \dots, n$ und alle $\theta \in \Theta$

$$\frac{dP_{\hat{\theta}(x), Z_i}}{d(P_{\hat{\theta}(x), Z_i} + P_{\theta, Z_i})}(y) = \frac{1}{n} \sum_{j=1}^n \frac{\mathbb{1}_{\{x_j - Z_j \hat{B}(x) + Z_i \hat{B}(x)\}}(y)}{[P_{\hat{\theta}(x), Z_i} + P_{\theta, Z_i}](\{x_j - Z_j \hat{B}(x) + Z_i \hat{B}(x)\})} \quad \text{für alle } y \in \mathbb{R}^p$$

eine Festlegung der Radon-Nikodym Dichte von $P_{\hat{\theta}(x), Z_i}$ bzgl. $[P_{\hat{\theta}(x), Z_i} + P_{\theta, Z_i}]$. Weiterhin ist $[P_{\hat{\theta}(x), Z_i} + P_{\theta, Z_i}](\{x_i\})$ positiv und somit die Radon-Nikodym Dichte an der Stelle x eindeutig durch

$$\frac{dP_{\hat{\theta}(x), Z_i}}{d[P_{\hat{\theta}(x), Z_i} + P_{\theta, Z_i}]}(x_i) = \frac{P_{\hat{\theta}(x), Z_i}(\{x_i\})}{P_{\hat{\theta}(x), Z_i}(\{x_i\}) + P_{\theta, Z_i}(\{x_i\})} \quad \text{für alle } i = 1, \dots, n$$

festgelegt. Es bleibt daher zu zeigen, dass für alle $\theta \in \Theta$

$$\prod_{i=1}^n \frac{P_{\hat{\theta}(x), Z_i}(\{x_i\})}{P_{\hat{\theta}(x), Z_i}(\{x_i\}) + P_{\theta, Z_i}(\{x_i\})} \geq \prod_{i=1}^n \frac{P_{\theta, Z_i}(\{x_i\})}{P_{\hat{\theta}(x), Z_i}(\{x_i\}) + P_{\theta, Z_i}(\{x_i\})}$$

oder äquivalent dazu

$$L_x(\hat{B}(x), \hat{F}_{\hat{B}(x), x}) \geq L_x(\theta)$$

gilt. Die Behauptung folgt somit direkt aus Lemma 6.2. \square

Korollar 6.4 *Es sei $\hat{\theta}(x) = (\hat{B}, \hat{G})$ für eine Beobachtung $\mathbf{X} = x$ eine vMLS nach Kiefer und Wolfowitz. Dann gilt*

$$P_{\hat{\theta}(x), Z}(\{x\}) > 0.$$

BEWEIS : Zusammen mit

$$P_{\hat{\theta}(x), Z}(\{x\}) = \prod_{i=1}^n \hat{G}(\{x_i - Z_i \hat{B}\})$$

folgt die Behauptung aus Satz 6.3. \square

Bemerkung 6.5 Nehmen wir an, dass die Beobachtung x im Spaltenraum $\mathcal{R}(Z)$ von Z enthalten ist, dann erhalten wir eine eindeutig definierte Lösung durch

$$\hat{B}_q(x) = (Z^t Z)^{-1} Z^t x.$$

Offensichtlich erfüllt $\hat{B}_q(x)$ die Bedingung $\bar{x} = \bar{Z} \hat{B}_q(x)$ und ist das eindeutig bestimmte Maximum von $L_x(\theta)$ für alle $\theta \in \Theta$. Andererseits für $x \notin \mathcal{R}(Z)$ ist im Allgemeinen die Lösung nicht eindeutig und entspricht nicht der Kleinsten Quadrate Lösung $\hat{B}_q(x)$. Betrachten wir zum Beispiel ein Zweistichprobenmodell mit $p = 1$, $n = 4$,

$$Z = \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & \mathbb{1}_2 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

sowie die Beobachtung $x^t = (1; -1; -2; 2)$. Die Kleinste Quadrate Lösung ist $\hat{B}_q(x) = (0; 0)^t$ und führt zu der Menge $\{1; -1; -2; 2\}$ der Residuen $x_i - Z_i \hat{B}_q(x)$, $i = 1, \dots, 4$. Das Maximum der Funktion $L_x(B, \hat{F}_{B,x})$ für $B \in \mathbb{R}^4$ wird zum Beispiel von $B_1 = (0.5, -0.5)$ mit der Residuenmenge $\{0.5; -1.5; -1.5; 1.5\}$ oder $B_2 = (-0.5, 0.5)$ mit der Residuenmenge $\{1.5; -0.5; -2.5; 1.5\}$ aber nicht von $\hat{B}_q(x)$ angenommen, da

$$\frac{1}{64} = L(B_1, \hat{F}_{B_1,x}) = \max_{\substack{B \in \mathbb{R}^4 \\ \bar{x} = \bar{Z} B}} L_x(B, \hat{F}_{B,x}) \geq L_x(\hat{B}_q(x), \hat{F}_{\hat{B}_q(x),x}) = \frac{1}{256}$$

gilt. Analoge Situationen kann man auch für andere Beobachtungen $x \notin \mathcal{R}(Z)$ angeben. \square

6.3 Ansatz von Gill

Es sei \mathbf{X} eine das statistische Experiment (vgl. Abschnitt 6.1)

$$(\mathbb{R}^{n \times p}, \mathfrak{B}^{n \times p}, \prod_{i=1}^n P_{\theta, Z_i} | \theta \in \Theta)$$

beschreibende zufällige $n \times p$ Matrix. Weiterhin sei für $i = 1, \dots, n$ und alle $B \in \mathbb{R}^{k \times p}$ die messbare Abbildung

$$T_{Z_i B}(x) = x + Z_i B \quad \text{für alle } x \in \mathbb{R}^p$$

definiert. Dann gilt für alle $\theta = (\mu, G) \in \Theta$ und alle $C \in \mathfrak{B}$

$$P_{\theta, Z_i}(C) = G(T_{Z_i B}^{-1}(C)).$$

Wir geben nun eine Abbildung ψ an, die die geforderten Eigenschaften (vgl. Annahme 3.2) erfüllt und bestimmen mit Hilfe der Abbildung ψ die verallgemeinerte Maximum-Likelihood-Schätzung nach Gill. Ferner haben wir in dem Satz 6.3 gezeigt, dass im Allgemeinen keine eindeutig definierte vMLS nach Kiefer und Wolfowitz existiert (vgl. Bemerkung 6.5). Somit können wir auch keine eindeutig definierte vMLS nach Gill erwarten (vgl. Bemerkung 3.5 und Korollar 6.4).

Annahme A.6.1 Wir bezeichnen mit $C_b(\mathbb{R}^p, \mathfrak{L}^p)$ die Menge aller beschränkten stetigen Funktionen auf \mathbb{R}^p und definieren für alle $\theta = (B, G) \in \Theta$, $h \in C_b(\mathbb{R}^p, \mathfrak{L}^p)$ sowie $c \in \mathbb{R}$ nahe genug an Null durch

$$\frac{dP_{\psi(\theta, h, c), Z_i}(x)}{dP_{\theta, Z_i}} := \frac{1 + ch\left(T_{Z_i B}^{-1}(x)\right)}{\int_{\mathbb{R}^p} \left[1 + ch\left(T_{Z_i B}^{-1}(y)\right)\right] dP_{\theta}(y)} = \frac{1 + ch(x - Z_i B)}{\int_{\mathbb{R}^p} \left[1 + ch(y)\right] G(dy)} \quad (6.6)$$

die Abbildung ψ .

Lemma 6.6 Es gelte die Annahme A.6.1, dann ist $\psi(\theta, h, c) \in \Theta$ und die Annahme 3.2 ist erfüllt.

BEWEIS : Es seien $\theta = (\mu, G) \in \Theta$ und $h \in C_b(\mathbb{R}^p, \mathfrak{L}^p)$ beliebig fixiert und c hinreichend nahe an 0 gewählt. Nun ist $1 + ch(x)$ nicht negativ für alle $x \in \mathbb{R}^p$ und das Integral $\int_{\mathbb{R}^p} 1 + ch(x)G(dx)$ endlich.

Ferner definiert

$$f(x) := \frac{1 + ch(x)}{\int_{\mathbb{R}^p} \left[1 + ch(x)\right] G(dx)}$$

eine Dichtefunktion eines Wahrscheinlichkeitsmaßes F bzgl. des dominierenden Maßes G . Weiterhin besitzt G und somit auch F ein endliches erstes und zweites Moment. Wir bezeichnen mit $\mu_F \in \mathbb{R}^p$ den Erwartungswert von F und $B + (Z^t Z)^{-1} Z^t \mathbb{1}_n \mu_F$ mit \tilde{B} . Dann gilt

$$P_{\psi(\theta, h, c), Z_i}(C) = F(T_{Z_i B}^{-1}(C)) = F^{T_{\mu_F}^{-1}}(T_{Z_i \tilde{B}}^{-1}(C)) \quad \text{für alle } C \in \mathfrak{B} \text{ und } i = 1, \dots, n$$

und somit $\psi(\theta, h, c) \in \Theta$. Offensichtlich ist $\psi(\theta, h, 0) = \theta$ und die Dichte $\frac{dP_{\psi(\theta, h, c), Z_i}(x)}{dP_{\theta}}(x)$ für alle $x \in \mathbb{R}^p$ in c an der Stelle $c = 0$ differenzierbar. \square

Des weiteren sind für $x = (x_1^t, \dots, x_n^t)^t \in \mathbb{R}^{n \times p}$, $h \in C_b(\mathbb{R}^p, \mathfrak{L}^p)$ und $\theta \in \Theta$ die Likelihood-Funktion bzw. Likelihood-Gleichung (vgl. (3.3)) durch

$$L_x(c, \theta, h, Z) = \prod_{i=1}^n \frac{1 + ch(x_i - Z_i B)}{\int_{\mathbb{R}^p} [1 + ch(y)] G(dy)},$$

$$U_{x^n}(\theta, h) = n \int_{\mathbb{R}^p} h(y) \left(\frac{1}{n} \sum_{i=1}^n \delta_{(x_i - Z_i B)} \right) (dy) - n \int_{\mathbb{R}^p} h(y) G(dy)$$

gegeben.

Satz 6.7 Es sei eine Beobachtung $\mathbf{X} = x$ mit $\mathbf{X} \sim P_{\theta, Z}$ und $\theta \in \Theta$ gegeben. Eine Schätzung $\hat{\theta}(x) = (\hat{B}(x), \hat{G}(x))$ ist eine verallgemeinerte Maximum-Likelihood-Schätzung nach Gill, falls sie die folgenden Bedingungen erfüllt:

1. Es gilt $\bar{x} = \bar{Z} \hat{B}(x)$.
2. $\hat{G}(x)$ ist die empirische Verteilung der Residuen $(x_1 - Z_1 \hat{B}(x), \dots, x_n - Z_n \hat{B}(x))$.

BEWEIS : Es sei $x = (x_1^t, \dots, x_n^t)^t \in \mathbb{R}^{n \times p}$ beliebig fixiert. Eine Schätzung $\hat{\theta}$ ist eine vMLS nach Gill, falls sie Lösung der Likelihood-Gleichungen (vgl. Definition 3.3)

$$\forall h \in C_b(\mathbb{R}^p, \mathfrak{L}^p) \quad : \quad n \int_{\mathbb{R}^p} h(y) \left(\frac{1}{n} \sum_{i=1}^n \delta_{(x_i - Z_i B)} - \hat{G} \right) (dy) = 0$$

ist. Unter der Bedingung $\hat{G} \in \mathcal{P}_+^0$ folgt nun die Behauptung des Satzes. \square

Bemerkung 6.8 Die Menge der vMLS nach Kiefer und Wolfowitz ist wie erwartet eine Teilmenge der Menge der vMLS nach Gill. Betrachten wir den Fall, dass die Beobachtung x im Spaltenraum $\mathcal{R}(Z)$ der Matrix Z liegt. Dann ist die vMLS für B und somit auch für G und Σ nach Kiefer und Wolfowitz eindeutig bestimmt und entspricht der Kleinsten Quadrate Lösung \hat{B}_q (vgl. Bemerkung 6.5). Dagegen ist im Allgemeinen die vMLS nach Gill auch für $x \in \mathcal{R}(Z)$ nicht eindeutig. Andererseits ist die Kleinste Quadrate Lösung \hat{B}_q für B immer eine vMLS nach Gill aber nur in speziellen Fällen eine vMLS von Kiefer und Wolfowitz (vgl. Bemerkung 6.5). Natürlich ist dieser Vergleich nicht ganz fair, da eine besser geeignete Abbildung ψ existieren könnte, für die mindestens die Gleichheit beider Konzepte folgt. Im Abschnitt 6.5 werden wir den Spezialfall eines Lokationsmodells betrachten und sehen, dass die hier vorgestellte Abbildung ψ in dem Lokationsmodell eine eindeutige Lösung liefert. \square

6.4 Selbstinformativer Grenzwert

6.4.1 A priori Annahmen

Es sei \mathbf{X} eine das semiparametrische lineare Modell beschreibende Zufallsvariable (vgl. Abschnitt 6.1). Unter der Annahme, dass die Zufallsvariablen $\mathbf{X}_1, \dots, \mathbf{X}_n$ einer Normalverteilung folgen und a priori der Parameter (B, Σ) eine Normal-Wishart-Verteilung besitzt, sind die Bayesschen Schätzungen für B und Σ bekannt (vgl. Abschnitt 4.6.3). Ausgehend von diesem Ergebnis werden wir annehmen, dass die unbekannte Verteilung der Zufallsvariablen $\mathbf{X}_1, \dots, \mathbf{X}_n$ durch eine Realisierung eines Dirichlet Prozesses und einer Normal-Wishart-Verteilung festgelegt sei. Der Dirichlet Prozess $\mathbf{P} \sim \mathcal{D}_{\alpha\mu}$ (vgl. Abschnitt 5.4.2) wird durch den Parameter $\alpha \in \mathbb{R}^+$ und das Wahrscheinlichkeitsmaß μ auf $(\mathbb{R}^p, \mathcal{L}^p)$ festgelegt. Insbesondere besitzen $\mathcal{D}_{\alpha\mu}$ -fast alle Realisierungen P des Dirichlet Prozesses ein endliches zweites Moment, falls das zweite Moment von μ endlich ist. (vgl. Ferguson (1973), Hartigan (1983), Simar (1984)).

Wir nehmen an, dass für die Zufallsvariablen $\mathbf{X}_1, \dots, \mathbf{X}_n$ gelte

$$\mathbf{X}_i = Z_i A + U_i \Lambda^{-\frac{1}{2}} \quad \text{für } i = 1, \dots, n. \quad (6.7)$$

Die p -dimensionalen Zufallsvariablen U_1, \dots, U_n seien unabhängig und identisch G -verteilt. Mit Hilfe der Parameter A, Λ und G ist die Verteilung der Zufallsvariable \mathbf{X}_i für $i = 1, \dots, n$ festgelegt. A priori fassen wir $\theta = (A, \Lambda)$ und G als Realisierung der unabhängigen Zufallsvariablen ϑ und \mathbf{P} auf. Die Zufallsvariable ϑ sei Normal-Wishart-verteilt (vgl. Abschnitt 4.6.3)

$$\vartheta \sim NW(k, p, l, A_0, M_0, S_0), \quad (6.8)$$

wobei $l - k > p$, $A_0 \in \mathbb{R}^{k \times p}$, M_0 bzw. S_0 eine positiv definite $k \times k$ bzw. $p \times p$ Matrix sei. Das zufällige Maß \mathbf{P} sei ein Dirichlet Prozess (vgl. Abschnitt 5.4.2)

$$\mathbf{P} \sim \mathcal{D}_{\alpha N_p(0, I_p)}, \quad (6.9)$$

dabei sei α eine positive reelle Zahl und $N_p(0, I_p)$ die Standardnormalverteilung auf \mathbb{R}^p .

Bemerkung 6.9 Im Allgemeinen besitzt eine Realisierung G des Dirichlet Prozesses nicht den Erwartungswert Null und die Einheitsmatrix als Kovarianzmatrix, so dass hier eine Reparametrisierung vorgenommen wird. Die Parameter A, Λ und die Verteilung G legen mit Gleichung (6.7) die Verteilung $P_i = P_i(A, \Lambda, G)$ der Zufallsvariable \mathbf{X}_i für $i = 1, \dots, n$ fest. Die interessierenden Parameter des

Modells (6.1) ergeben sich dann als

$$B = A + Z^+ \mathbb{1} \mu_G \Lambda^{-\frac{1}{2}}, \quad \Sigma = \Lambda^{-\frac{1}{2}} C_G \Lambda^{-\frac{1}{2}}, \quad (6.10)$$

wobei μ_G der Erwartungswert, C_G die Kovarianz von G und $Z^+ = (Z^t Z)^{-1} Z^t$ ist. Natürlich sind damit A und Λ im Allgemeinen nicht identifizierbar. Dies ist aber keine Einschränkung für einen Bayesschen Ansatz (vgl. Kapitel 4.6). Ferner sind wir nicht an A und Λ sondern nur an den identifizierbaren ursprünglichen Parametern B und Σ interessiert. \square

6.4.2 Bayessche Schätzungen

In Bunke (2002) werden die im Folgenden dargestellten Bayesschen Schätzungen bestimmt. Wir werden nachweisen, dass für diese Festlegung der Bayesschen Schätzungen der vorgestellte selbstinformative Grenzwert existiert und den bekannten Kleinsten Quadrate Lösungen für die Parameter B und Σ sowie der empirischen Verteilung der Residuen entspricht.

Es sei $N_{1v}, \dots, N_{|v|v}$ ein Partition v der Menge $N = \{1, \dots, n\}$. Die Menge aller Partitionen v sei \mathbb{D} . Der (i, j) Eintrag der Inzidenzmatrix K_v einer Partition v ist 1, falls i, j in einem Segment der Partition v liegen und in allen anderen Fällen gleich 0. Die Anzahl der Elemente in einem Segment N_{iv} sei n_{iv} und o.B.d.A. gelte $n_{1v} \geq \dots \geq n_{|v|v}$. Die Partition mit nur einelementigen Segmenten sei $*$, d.h. $n_{i*} = 1$ für $i = 1, \dots, n$. Dann ist

$$K_* = I_n.$$

Für alle Partitionen $v \in \mathbb{D} \setminus \{*\}$ existiert o.B.d.A. ein $q_v \in \mathbb{N}$ mit

$$\forall j = 1, \dots, q_v : n_{jv} > 1 \text{ und } \forall j = q_v + 1, \dots, |v| : n_{jv} = 1.$$

Des weiteren seien $n_v := \sum_{i=1}^{q_v} n_{iv}$ und $B_v = \text{Diag}[\mathbb{1}_{n_{1v}}, \dots, \mathbb{1}_{n_{q_v v}}]$. Dann existiert eine $n \times n_v$ Matrix C_v , deren Spalten Einheitsvektoren des \mathbb{R}^n sind mit $C_v^t C_v = I_{n_v}$ und

$$K_v = I_n + C_v (B_v B_v^t - I_{n_v}) C_v^t. \quad (6.11)$$

Lemma 6.10 *Es sei M eine positiv definite $k \times k$ Matrix, Z eine $n \times k$ Matrix mit Rang k . Dann gilt:*

$$\text{Rang}[Z M Z^t + K_v] = n \quad \Leftrightarrow \quad \text{Rang}[(C_v^t Z : B_v)] = n_v.$$

BEWEIS : Offensichtlich gilt $\text{Rang}[Z M Z^t + K_v] = n$ genau dann, wenn

$$\text{Rang} \left[(Z^t C_v | Z^t \bar{C}_v)^t M (Z^t C_v | Z^t \bar{C}_v) + \begin{pmatrix} B_v B_v^t & 0 \\ 0 & I_{n-n_v} \end{pmatrix} \right] = n$$

gilt, wobei \bar{C}_v durch $\bar{C}_v \bar{C}_v^t = I_n - C_v C_v^t$ definiert sei. Insbesondere ist damit

$$\text{Rang} \left[\begin{pmatrix} C_v^t Z & B_v & 0 \\ \bar{C}_v^t Z & 0 & I_{n-n_v} \end{pmatrix} \right] = n$$

äquivalent zu $\text{Rang}[Z M Z^t + K_v] = n$, so dass die Behauptung folgt. \square

Bemerkung 6.11 Eine zufällige $k \times p$ Matrix \mathbf{B} besitzt eine verallgemeinerte t-Verteilung

$$\mathbf{B} \sim GMt(v, k, p, M_0, S_0, B_0),$$

wenn die Dichte bzgl. des Lebesgue Maßes im $\mathbb{R}^{k \times p}$ gegeben ist durch (vgl. Humak (1977), S.491, A 2.39, Box und Tiao (1992), S.441)

$$f(\mathbf{B}) = \pi^{-\frac{kp}{2}} \frac{\prod_{i=1}^p \Gamma[\frac{v-i+1}{2}]}{\prod_{i=1}^p \Gamma[\frac{(v-k-i+1)}{2}]} |M_0|^{\frac{v-k}{2}} |S_0^{-1}|^{\frac{p}{2}} |M_0 + (\mathbf{B} - B_0)^t S_0^{-1} (\mathbf{B} - B_0)|^{-\frac{v}{2}}$$

mit den reellwertigen Parametern $v, k, p \geq 1$, einer $k \times p$ Matrix B_0 , sowie positiv definiten $p \times p$ und $k \times k$ Matrizen M_0 und S_0 , wobei Γ die Gammafunktion ist. \square

Annahme A.6.2 Für alle $X \in \mathbb{R}^{n \times p}$ definieren wir die Parameter

$$\begin{aligned} \Omega_v &:= ZM_0Z^t + K_v, & \alpha_v &:= \frac{\alpha^{|\nu|}}{[\alpha \cdots (\alpha + n - 1)]} \prod_{i=1}^{|\nu|} n_{iv}!, \\ D &:= \{v \in \mathbb{D} \mid \text{Rang}[\Omega_v] = n\}, & \tilde{A}_v &:= A_0 + M_0Z^t\Omega_v^{-1}(X - ZA_0), \\ \widetilde{\Lambda}_v^{-1} &:= \frac{1}{n+l-k} \tilde{S}_v, & \tilde{S}_v &:= S_0 + (X - ZA_0)^t \Omega_v^{-1} (X - ZA_0), \\ \tilde{\Delta}_v &:= M_0 - M_0Z^t\Omega_v^{-1}ZM_0, & \tilde{\rho}_v &:= \text{sp}[\tilde{\Delta}_v(\frac{1}{n}Z^tZ - \frac{n}{\alpha+n}\bar{Z}^t\bar{Z})], \\ h_{iv} &:= Z_i\tilde{\Delta}_vZ_i^t, & h_{ijv} &:= (Z_i - Z_j)\tilde{\Delta}_v(Z_i - Z_j)^t, \end{aligned}$$

und die Verteilungen F_{iv} bzw. G_{ijv} auf \mathfrak{L}^p . Die Verteilung F_{iv} ist eine verallgemeinerte t-Verteilung $GMt(n+l-k, 1, p, h_{iv}+1, \tilde{S}_v, Z_i\tilde{A}_v)$. Die Verteilung G_{ijv} ist eine verallgemeinerte t-Verteilung $GMt(n+l-k, 1, p, h_{ijv}, \tilde{S}_v, X_j + (Z_i - Z_j)\tilde{A}_v)$, falls h_{ijv} positiv ist und in allen anderen Fällen das Punktmaß $\delta_{X_j + (Z_i - Z_j)\tilde{A}_v}$.

Satz 6.12 (Bunke (2002)) Es sei eine Beobachtung $\mathbf{X} = X$ gegeben. Dann sind

$$\tilde{B} = \sum_{v \in D} k_v \tilde{B}_v, \quad \tilde{\Sigma} = \sum_{v \in D} k_v \tilde{\Sigma}_v, \quad \tilde{P}_i = \sum_{v \in D} k_v \tilde{P}_{iv} \quad \text{für } i = 1, \dots, n$$

Bayessche Schätzungen von $B, \Sigma, P_1, \dots, P_n$, wobei mit der Annahme A.6.2 gilt:

$$\begin{aligned} k_v &= \frac{\alpha_v |\Omega_v|^{-\frac{p}{2}} |\tilde{S}_v|^{-\frac{n+l-k}{2}}}{\sum_{l \in D} \alpha_l |\Omega_l|^{-\frac{p}{2}} |\tilde{S}_l|^{-\frac{n+l-k}{2}}}, \\ \tilde{B}_v &= \tilde{A}_v + \frac{n}{\alpha+n} Z^+ \mathbb{1}_n (\bar{X} - \bar{Z}\tilde{A}_v), \\ \tilde{P}_{iv} &= \frac{\alpha}{\alpha+n} F_{iv} + \frac{n}{\alpha+n} \left[\frac{1}{n} \sum_{j=1}^n G_{ijv} \right], \\ \tilde{\Sigma}_v &= \frac{\alpha+n\tilde{\rho}_v}{\alpha+n+1} \widetilde{\Lambda}_v^{-1} \\ &\quad + \frac{1}{\alpha+n+1} (X - Z\tilde{A}_v)^t (X - Z\tilde{A}_v) - \frac{n^2}{(n+\alpha)(n+\alpha+1)} (\bar{X} - \bar{Z}\tilde{A}_v)^t (\bar{X} - \bar{Z}\tilde{A}_v). \end{aligned}$$

Bemerkung 6.13 In Bunke (2002) wird gezeigt, dass die Randverteilung $P^{\mathbf{X}}$ des zufälligen Vektors \mathbf{X} unter den a priori Annahmen (6.7) - (6.9) für ϑ und \mathbf{P} eine endliche Mischung von Verteilungen $P_v^{\mathbf{X}}$

$$P^{\mathbf{X}} = \sum_{v \in D} k_v P_v^{\mathbf{X}}$$

ist. Die Verteilungen $P_v^{\mathbf{X}}$ sind konzentriert auf Mengen \mathcal{X}_v , d.h., es gilt

$$P_v^{\mathbf{X}}(\mathcal{X}_v) = 1.$$

Insbesondere sind die Mengen \mathcal{X}_v disjunkt und in linearen Unterräumen L_v des $\mathbb{R}^{n \times p}$ enthalten. Jede Verteilung $P_v^{\mathbf{X}}$ ist weiterhin absolut stetig bzgl. des Lebesgue Maßes λ_v auf L_v . Weiterhin sind die vorgestellten Bayesschen Schätzungen Festlegungen der bedingten Erwartungswerte. Insbesondere sind wegen Satz 4.5 die Festlegungen der bedingten Erwartungswerte eindeutig, da wir in der Beobachtung $\mathbf{X} = X$ stetige Festlegungen gewählt haben. \square

6.4.3 Selbstinformativer Grenzwert

Für $r \in \mathbb{N}$ betrachten wir r unabhängige Wiederholungen des linearen Modells, d.h. das erweiterte lineare Modell mit rn Zufallsvariablen $\tilde{X}_1^r, \dots, \tilde{X}_{rn}^r$ und der Regressionsmatrix $Z_r = \mathbb{1}_r \otimes Z$. In diesem Modell ergeben sich die Bayesschen Schätzungen $\tilde{B}_r, \tilde{\Sigma}_r, \tilde{P}_{ir}$ direkt aus Satz 6.12. Wählen wir die originale Beobachtung X des Modells (6.1) als einen speziellen Wert $\tilde{X}^r = \mathbb{1}_r \otimes X$ in dem Modell mit r Wiederholungen, so erhalten wir eine Festlegung $\tilde{B}_{(r)}, \tilde{\Sigma}_{(r)}, \tilde{P}_{i(r)}$ der Bayesschen Schätzung nach r Iterationen. Mit Bemerkung 6.13 ist diese Festlegung eindeutig, wenn wir nur stetige Versionen betrachten. Im Folgenden werden wir zeigen, dass die Folgen $(\tilde{B}_{(r)})_{r \in \mathbb{N}}, (\tilde{\Sigma}_{(r)})_{r \in \mathbb{N}}, (\tilde{P}_{i(r)})_{r \in \mathbb{N}}$ konvergieren und somit der selbstinformativ-Grenzwert existiert.

Die Bayesschen Schätzungen sind nach Satz 6.12 gegeben durch

$$\tilde{B}_r = \sum_{v_r \in D_r} k_{v_r} \tilde{B}_{v_r}, \quad \tilde{\Sigma}_r = \sum_{v_r \in D_r} k_{v_r} \tilde{\Sigma}_{v_r}, \quad \tilde{P}_{ir} = \sum_{v_r \in D_r} k_{v_r} \tilde{P}_{iv_r}.$$

Zuerst werden wir zeigen, dass eine von r unabhängige Menge $[D]$ existiert, so dass für die Bayesschen Schätzungen mit dem speziellen Wert $\tilde{X}^r = \mathbb{1}_r \otimes X$ gilt

$$\tilde{B}_{(r)} = k_{*r} \tilde{B}_{*r} + \sum_{[v] \in [D]} \tilde{k}_{[v]r} \tilde{B}_{[v]r}, \quad \tilde{\Sigma}_{(r)} = k_{*r} \tilde{\Sigma}_{*r} + \sum_{[v] \in [D]} \tilde{k}_{[v]r} \tilde{\Sigma}_{[v]r}, \quad \tilde{P}_{i(r)} = k_{*r} \tilde{P}_{i*r} + \sum_{[v] \in [D]} \tilde{k}_{[v]r} \tilde{P}_{i[v]r}.$$

Im Anschluss daran werden wir nachweisen, dass für alle $[v] \in [D]$ die Folgen $(k_{*r})_{r \in \mathbb{N}}, (\tilde{B}_{*r})_{r \in \mathbb{N}}, (\tilde{\Sigma}_{*r})_{r \in \mathbb{N}}, (\tilde{P}_{i*r})_{r \in \mathbb{N}}$ und $(\tilde{k}_{[v]r})_{r \in \mathbb{N}}, (\tilde{B}_{[v]r})_{r \in \mathbb{N}}, (\tilde{\Sigma}_{[v]r})_{r \in \mathbb{N}}, (\tilde{P}_{i[v]r})_{r \in \mathbb{N}}$ für $r \rightarrow \infty$ konvergieren. Abschließend werden wir dann die selbstinformativen Grenzwerte bestimmen.

Es sei $r \in \mathbb{N}$ fixiert und \mathbb{D}_r bezeichne die Menge aller Partitionen von $\{1, \dots, rn\}$. Die Partition mit einelementigen Segmenten bezeichnen wir mit $*_r$. Ferner existieren für alle Partitionen $v_r \in \mathbb{D}_r \setminus \{*_r\}$ Matrizen

$$B_{v_r} = \text{Diag}[\mathbb{1}_{n_{1v_r}}, \dots, \mathbb{1}_{n_{qv_r v_r}}] \quad \text{und} \quad C_{v_r} = (e_{i_1}^r, \dots, e_{i_{n_{v_r}}}^r),$$

wobei e_i^r den i -ten Einheitsvektor des \mathbb{R}^{rn} bezeichnet, mit

$$K_{v_r} = I_{rn} + C_{v_r} (B_{v_r} B_{v_r}^t - I_{n_{v_r}}) C_{v_r}^t.$$

Nun definieren wir für alle $v_r \in \mathbb{D}_r \setminus \{*_r\}$ mit $C_{v_r} = (e_{i_1}^r, \dots, e_{i_{n_{v_r}}}^r)$ eine $n \times n_{v_r}$ Matrix

$$C_{[v_r]_r} := (e_{[i_1]_r}, \dots, e_{[i_{n_{v_r}}]_r}), \quad (6.12)$$

dabei ist $e_{[j]_r}$ der $[j]_r$ -te Einheitsvektor des \mathbb{R}^n mit

$$[j]_r := j \bmod r \in \{1, \dots, n\}, \quad \forall j \in \{1, \dots, rn\}. \quad (6.13)$$

Insbesondere gilt dann für den speziellen Wert $\tilde{X}^r = \mathbb{1}_r \otimes X$ und $Z_r = \mathbb{1}_r \otimes Z$

$$C_{v_r}^t Z_r = C_{[v_r]_r}^t Z \quad \text{und} \quad C_{v_r}^t \tilde{X}^r = C_{[v_r]_r}^t X. \quad (6.14)$$

Definition 6.14 Zwei Partitionen $v_r, \tilde{v}_r \in \mathbb{D}_r \setminus \{*_r\}$ sind äquivalent ($v_r \sim \tilde{v}_r$), falls gilt

$$B_{v_r} = B_{\tilde{v}_r} \quad \text{und} \quad C_{[v_r]_r} = C_{[\tilde{v}_r]_r}.$$

Des weiteren sei $D_r \subset \mathbb{D}_r$ die Menge aller Partitionen v_r mit $\text{Rang}[Z_r M_0 Z_r^t + K_{v_r}] = rn$ und $[D_r]$ die Menge aller Äquivalenzklassen $D_r(v_r)$ bzgl. \sim in $D_r \setminus \{*_r\}$.

Lemma 6.15 Es sei $\tilde{X}^r = \mathbb{1}_r \otimes X$, dann gilt für alle Partitionen \tilde{v}_r einer Äquivalenzklasse $D_r(v_r)$ mit $v_r \in D_r \setminus \{*_r\}$

$$k_{v_r} = k_{\tilde{v}_r}, \quad \tilde{B}_{v_r} = \tilde{B}_{\tilde{v}_r}, \quad \tilde{\Sigma}_{v_r} = \tilde{\Sigma}_{\tilde{v}_r}, \quad \tilde{P}_{iv_r} = \tilde{P}_{i\tilde{v}_r} \quad \text{für } i = 1, \dots, n.$$

Die Anzahl $\#D_r(v_r)$ der Partitionen in einer Äquivalenzklasse $D_r(v_r)$ ist

$$\#D_r(v_r) = r^{n_{v_r}} d_{v_r} \quad \text{mit} \quad \lim_{r \rightarrow \infty} d_{v_r} = 1.$$

BEWEIS : Es seien $v_r \in D_r \setminus \{*_r\}$ und $\tilde{v}_r \in D_r(v_r)$ beliebig fixiert. Dann besitzen \tilde{v}_r und v_r die gleiche Anzahl an Segmenten und die gleiche Anzahl an Indizes pro Segment, so dass $\alpha_{\tilde{v}_r} = \alpha_{v_r}$ gilt. Auf Grund der Äquivalenz von \tilde{v}_r und v_r (vgl. (6.14)) folgt mit $X_r = \mathbb{1} \otimes X$

$$B_{v_r} = B_{\tilde{v}_r}, \quad C_{v_r}^t Z_r = C_{\tilde{v}_r}^t Z_r \quad \text{und} \quad C_{v_r}^t X_r = C_{\tilde{v}_r}^t X_r. \quad (6.15)$$

Insbesondere gilt nun $\Omega_v^{-1} = P_{Z_r}^{-1} + P_{Z_r}^{-1} C_v H_v^{-1} C_v^t P_{Z_r}^{-1}$ für alle $v \in D_r \setminus \{*_r\}$ mit

$$\begin{aligned} P_{Z_r}^{-1} &= I_{rn} - Z_r (M_0^{-1} + Z_r^t Z_r)^{-1} Z_r^t, \\ H_v &= C_v^t Z_r (M_0^{-1} + Z_r^t Z_r)^{-1} Z_r^t C_v + B_v \text{Diag}[(1 - n_i)^{-1}] B_v^t. \end{aligned}$$

Wegen (6.15) folgt dann $H_{v_r} = H_{\tilde{v}_r}$ und somit auch

$$X_r^t \Omega_{v_r}^{-1} X_r = X_r^t \Omega_{\tilde{v}_r}^{-1} X_r, \quad X_r^t \Omega_{v_r}^{-1} Z_r = X_r^t \Omega_{\tilde{v}_r}^{-1} Z_r, \quad Z_r^t \Omega_{v_r}^{-1} Z_r = Z_r^t \Omega_{\tilde{v}_r}^{-1} Z_r.$$

Weiterhin gilt unter der Annahme A.6.2 nun

$$A_{v_r} = A_{\tilde{v}_r}, \quad S_{v_r} = S_{\tilde{v}_r}, \quad \Delta_{v_r} = \Delta_{\tilde{v}_r}, \quad \rho_{v_r} = \rho_{\tilde{v}_r}.$$

Des weiteren existieren Matrizen \bar{C}_{v_r} und $\bar{C}_{\tilde{v}_r}$ mit $(C_{v_r} \bar{C}_{v_r})^{-1} = (C_{v_r} \bar{C}_{v_r})^t$ sowie $(C_{\tilde{v}_r} \bar{C}_{\tilde{v}_r})^{-1} = (C_{\tilde{v}_r} \bar{C}_{\tilde{v}_r})^t$ und $\bar{C}_{v_r}^t Z_r = \bar{C}_{\tilde{v}_r}^t Z_r$ bzw. $\bar{C}_{v_r}^t X_r = \bar{C}_{\tilde{v}_r}^t X_r$, so dass auch

$$|\Omega_{v_r}| = |(C_{v_r} \bar{C}_{v_r})^t \Omega_{v_r} (C_{v_r} \bar{C}_{v_r})| = |(C_{\tilde{v}_r} \bar{C}_{\tilde{v}_r})^t \Omega_{\tilde{v}_r} (C_{\tilde{v}_r} \bar{C}_{\tilde{v}_r})| = |\Omega_{\tilde{v}_r}| \quad (6.16)$$

folgt. Damit folgt der erste Teil der Behauptung. Mittels kombinatorischen Abzählens lässt sich die Anzahl $\#D_r(v_r)$ der Partitionen in der Äquivalenzklasse $D_r(v_r)$ bestimmen. Die zur Matrix $C_{[v_r]_r}$ zugehörige Aufteilung der Indexmenge sei $N_{1[v_r]_r}, \dots, N_{q_{v_r}[v_r]_r}$. Die Anzahl der Indizes in dem Segment $N_{i[v_r]_r}$ sei n_{iv_r} , dabei können Indizes in dieser Aufteilung mehrfach auftreten. Es seien nun t_1, \dots, t_m die paarweise verschiedenen Indizes mit Häufigkeiten h_1, \dots, h_m in der Menge $T = \bigcup_{i=1}^{q_{v_r}} N_{i[v_r]_r}$. Gesucht ist die Anzahl von Aufteilungen der Indexmenge $N_r = \{1, \dots, rn\}$ in Segmente $N_{1v_r}, \dots, N_{q_{v_r}v_r}$, wobei die Anzahl der Indizes pro Segment erneut n_{iv_r} ist und kein Index mehrfach auftreten kann. Zu jedem Index $t_i \in T$ existieren r Indizes l in der Menge N_r mit $[l]_r = t_i$. Aus diesen r Indizes werden h_i verschiedene ausgewählt, dafür gibt es $r(r-1) \cdots (r-h_i+1)$ Möglichkeiten. Da dies für jeden Index t_1, \dots, t_m gilt, existieren insgesamt

$$\#D_r(v) = \prod_{i=1}^m \prod_{j=0}^{h_i-1} (r-j) = r^{n_v} \prod_{i=1}^m \prod_{j=0}^{h_i-1} \left(1 - \frac{j}{r}\right)$$

Möglichkeiten der Aufteilung. □

Zu jeder Äquivalenzklasse $D_r(v_r)$ existieren Matrizen B und C mit (vgl. Definition 6.14)

$$B = B_{\tilde{v}_r} \quad \text{und} \quad C = C_{[\tilde{v}_r]_r}, \quad \text{für alle } \tilde{v}_r \in D_r(v_r).$$

Wir fassen nun $[v] = (B, C)$ als einen Repräsentanten der Äquivalenzklasse $D_r(v_r)$ auf und geben ein vollständiges Repräsentantensystem $[D]$ an, welches unabhängig von r ist, aber natürlich von der Matrix Z abhängt. Das folgende Lemma liefert die zentrale Voraussetzung für die darauf folgende Definition des vollständigen Repräsentantensystems.

Lemma 6.16 *Für alle Partitionen $v_r \in D_r \setminus \{*_r\}$ gilt*

$$\sum_{i=1}^{q_{v_r}} n_{iv_r} < k + q_{v_r}, \quad q_{v_r} < k, \quad n_{iv_r} \leq k \quad \text{für } i = 1, \dots, q_{v_r}.$$

BEWEIS : Wegen Lemma 6.10 gilt $\text{Rang}[C_{v_r}^t Z_r : B_{v_r}] = n_{v_r}$ für alle Partitionen $v_r \in D_r \setminus \{*\}$. O.B.d.A. seien $Z_{11}, \dots, Z_{1n_1}, \dots, Z_{qn_q}$ die Zeilen einer $n \times k$ Matrix Z mit $\text{Rang}[Z] = k$ und $n_i > 1$, $i = 1, \dots, q$ sowie $\sum_{i=1}^q n_i = n$. Weiterhin liege der Einsvektor $\mathbb{1}_n$ im Spaltenraum $\mathcal{R}(Z)$ von Z und insbesondere gelte

$$n = \text{Rang} \left[\begin{pmatrix} Z_{11} & 1 & 0 \\ \vdots & \vdots & \vdots \\ Z_{1n_1} & 1 & 0 \\ \vdots & \vdots & \ddots \\ Z_{q1} & 0 & 1 \\ \vdots & \vdots & \vdots \\ Z_{qn_q} & 0 & 1 \end{pmatrix} \right] = \text{Rang} \left[\begin{pmatrix} Z_{11} & 1 & 0 \\ Z_{12} - Z_{11} & 0 & 0 \\ \vdots & \vdots & \vdots \\ Z_{1n_1} - Z_{11} & 0 & 0 \\ \vdots & \vdots & \ddots \\ Z_{q1} & 0 & 1 \\ Z_{q2} - Z_{q1} & 0 & 0 \\ \vdots & \vdots & \vdots \\ Z_{qn_q} - Z_{q1} & 0 & 0 \end{pmatrix} \right].$$

Somit sind die $n - q$ Zeilenvektoren

$$Z_{12} - Z_{11}, \dots, Z_{1n_1} - Z_{11}, \dots, Z_{q2} - Z_{q1}, \dots, Z_{qn_q} - Z_{q1} \tag{6.17}$$

linear unabhängig. Wir werden zeigen, dass die Behauptungen

$$n < k + q, \quad q < k \quad n_i \leq k \quad \text{für } i = 1, \dots, q$$

aus der Annahme der Unabhängigkeit der $n - q$ Zeilenvektoren (6.17) folgt.

O.B.d.A. sei $\{Z_1, \dots, Z_k\}$ eine linear unabhängige Teilmenge der n Zeilenvektoren von Z . Dann lässt sich jede Zeile als Linearkombination $Z_{ij} = \sum_{m=1}^k c_{ijm} Z_m$ darstellen. Ferner sind die $n - q$ Zeilenvektoren in (6.17) unabhängig, falls die Gleichung

$$0 = \sum_{i=1}^q \sum_{j=2}^{n_i} a_{ij} (Z_{ij} - Z_{i1}) = \sum_{m=1}^k \left[\sum_{i=1}^q \sum_{j=2}^{n_i} a_{ij} (c_{ijm} - c_{i1m}) \right] Z_m$$

nur die triviale Lösung oder äquivalent dazu die $k \times n - q$ Matrix

$$M := \begin{pmatrix} (c_{121} - c_{111}) & \cdots & (c_{qn_q1} - c_{q11}) \\ \vdots & & \vdots \\ (c_{12k} - c_{11k}) & \cdots & (c_{qn_qk} - c_{q1k}) \end{pmatrix}$$

den Rang $n - q$ besitzt. Unter der Annahme, dass der Einsvektor $\mathbb{1}_n$ im Spaltenraum $\mathcal{R}(Z)$ enthalten ist, folgt $0 = \sum_{m=1}^k (c_{ijm} - c_{i1m})$ für $i = 1, \dots, q$, so dass die Zeilen der Matrix linear abhängig sind. Damit folgt die Behauptung $n < k + q$. Nehmen wir nun an, dass $q \geq k$ ist, dann gilt $n - q \geq q \geq k$, da $n_i > 1$ für $i = 1, \dots, q$ ist, was aber im Widerspruch zur Annahme das $\text{Rang}[M] = n - q$ steht, so dass die Behauptung $q < k$ folgt. Wählen wir schließlich o.B.d.A. $q = 1$, so folgt aus $n < k + q$ die Behauptung $n_i \leq k$ für $i = 1, \dots, q$. \square

Wir bezeichnen mit \mathbb{B} die Menge aller Matrizen

$$B = \text{Diag}[\mathbb{1}_{n_1}, \dots, \mathbb{1}_{n_{m_B}}] \quad (6.18)$$

mit $n_1 \geq \dots \geq n_{m_B} > 1$, $m_B, n_1, \dots, n_{m_B} \in \mathbb{N}$ und $n_B := \sum_{i=1}^{m_B} n_i$. Weiterhin definieren wir für alle $B \in \mathbb{B}$ die Menge $\mathbb{C}(B, Z)$ aller $n \times n_B$ Matrizen $C = (e_{i_1}, \dots, e_{i_{n_B}})$ mit

$$\text{Rang}[C^t Z; B] = n_B. \quad (6.19)$$

Insbesondere sei nun \mathbb{B}_k die Menge aller Matrizen $B \in \mathbb{B}$ mit

$$m_B < k, \quad n_B < k + m_B, \quad n_{i_B} \leq k \quad \text{für } i = 1, \dots, m_B \quad (6.20)$$

und

$$[D] := \{[v] = (B, C) \mid C \in \mathbb{C}(B, Z), B \in \mathbb{B}_k\}. \quad (6.21)$$

Lemma 6.17 *Für alle $r \geq k$ ist $[D]$ ein vollständiges Repräsentantensystem für die Menge $[D_r]$ der Äquivalenzklassen $D_r(v_r)$.*

BEWEIS : Es sei $r \geq k$ fixiert. Für alle Partitionen $v_r \in D_r \setminus \{*_r\}$ existiert nun mit Lemma 6.16 ein $[v] = (B, C) \in [D]$ mit $B = B_{v_r}$ und $C = C_{[v_r]_r}$. Andererseits können wir zu jedem $[v] = (B, C) \in [D]$ eine Partition $v_r \in D_r \setminus \{*_r\}$ mit

$$B = B_{v_r} \quad \text{und} \quad C = C_{[v_r]_r}$$

konstruieren. Dazu sei $B \in \mathbb{B}_k$ mit den Konstanten $m_B, n_B, n_{1B}, \dots, n_{m_B B}$ und $C \in \mathbb{C}(B, Z)$ mit $C = (e_{i_1}, \dots, e_{i_{n_B}})$ beliebig fixiert. Wir betrachten nun die Spalten $e_{i_1}, \dots, e_{i_{n_B}}$ der Matrix C und definieren die Mengen

$$N_j := \{i_k \mid \sum_{l=1}^{j-1} n_{lB} + 1 \leq k \leq \sum_{l=1}^{j-1} n_{lB} + n_{jB}\}, \quad j = 1, \dots, m_B$$

der Indizes der Einheitsvektoren e_{i_k} . Insbesondere gilt dann $\#N_j = n_{jB}$. Definieren wir nun für $j = 1, \dots, m_B$ die Mengen

$$N_j^r := \{i + (j-1)n \mid i \in N_j\} \subset \{(j-1)n + 1, \dots, jn\},$$

dann gilt $N_j^r \cap N_l^r = \emptyset$ für $j \neq l$ und $\#N_j^r = n_{jB}$. Insbesondere gilt $m_B < k$ (vgl. Definition \mathbb{B}_k (6.20)), so dass

$$N^r := \sum_{j=1}^{m_B} N_j^r \subset \{1, \dots, rn\} \quad \text{für alle } r \geq k$$

folgt. Damit ist

$$v_r : \sum_{j=1}^{m_B} N_j^r + \sum_{i \notin N^r} \{i\} = \{1, \dots, rn\}$$

eine Partition aus $D_r \setminus \{*_r\}$ mit den zugehörigen Matrizen B_{v_r} und C_{v_r} , wobei nach Konstruktion

$$B = B_{v_r} \quad \text{und} \quad C = C_{[v_r]_r}$$

gilt. □

Betrachten wir somit eine Äquivalenzklasse $D_r(v_r)$ für $v_r \in D_r \setminus \{*_r\}$. So können wir diese mit dem zugehörigen Element $[v] \in [D]$ identifizieren. Weiterhin sind für den Wert $\tilde{X}^r = \mathbb{1}_r \otimes X$ mit Lemma 6.15 die zugehörigen Parameter für alle Elemente der Äquivalenzklasse $D_r(v_r)$ identisch, so dass wir sie ab jetzt mit $\tilde{B}_{[v]_r}$, $\tilde{\Sigma}_{[v]_r}$ und $\tilde{P}_{i[v]_r}$ bezeichnen. Die Gewichte $\tilde{k}_{[v]_r}$ ergeben sich dann ebenfalls aus Lemma 6.15 mit

$$\tilde{k}_{[v]_r} = \frac{\alpha_{[v]_r} r^{n[v]} d_{[v]_r} |\Omega_{[v]_r}|^{-\frac{p}{2}} |\tilde{S}_{[v]_r}|^{-\frac{rn+l-k}{2}}}{\alpha_{*_r} |\Omega_{*_r}|^{-\frac{p}{2}} |\tilde{S}_{*_r}|^{-\frac{rn+l-k}{2}} + \sum_{[l] \in [D]} r^{n[l]} d_{[l]_r} \alpha_{[l]_r} |\Omega_{[l]_r}|^{-\frac{p}{2}} |\tilde{S}_{[l]_r}|^{-\frac{rn+l-k}{2}}}. \quad (6.22)$$

Im Folgenden gehen wir nun zur der angekündigten Darstellung

$$\tilde{B}_{(r)} = k_{*_r} \tilde{B}_{*_r} + \sum_{[v] \in [D]} \tilde{k}_{[v]_r} \tilde{B}_{[v]_r}, \quad \tilde{\Sigma}_{(r)} = k_{*_r} \tilde{\Sigma}_{*_r} + \sum_{[v] \in [D]} \tilde{k}_{[v]_r} \tilde{\Sigma}_{[v]_r}, \quad \tilde{P}_{i_r} = k_{*_r} \tilde{P}_{i*_r} + \sum_{[v] \in [D]} k_{[v]_r} \tilde{P}_{i[v]_r}$$

über und zeigen, dass die auftretenden Folgen konvergieren. Insbesondere benötigen wir die folgende Eigenschaft der verallgemeinerten t-Verteilung.

Lemma 6.18 *Es sei $(F_n)_{n \in \mathbb{N}}$ eine Folge von verallgemeinerten t-Verteilungen $GMt(nh_n, 1, p, n^{-1}d_n, nB_n, c_n)$ für $n \in \mathbb{N}$, wobei für die Parameter gelte:*

1. $(B_n)_{n \in \mathbb{N}}$ sei eine Folge positiver definiten $p \times p$ Matrizen mit dem Grenzwert B_0 und $sp[B_0] > 0$,
2. $(c_n)_{n \in \mathbb{N}}$ sei eine Folge von Vektoren $c_n \in \mathbb{R}^p$ mit dem Grenzwert $c_0 \in \mathbb{R}^p$,
3. $(h_n)_{n \in \mathbb{N}}$ und $(d_n)_{n \in \mathbb{N}}$ seien positive Folgen mit den Grenzwerten $h_0 > 1$ und $d_0 \geq 0$, wobei $h_n > \frac{p}{n}$ für alle $n \in \mathbb{N}$ gelte.

Dann konvergiert die Folge $(F_n)_{n \in \mathbb{N}}$ schwach gegen das Punktmass δ_{c_0} an der Stelle c_0

$$F_n \xrightarrow{w} \delta_{c_0} \quad \text{für } n \rightarrow \infty.$$

BEWEIS : Wir zeigen mit Hilfe des Satzes von der dominierten Konvergenz, dass für das Komplement jeder Kugel $B_\varepsilon(c_0)$ mit Radius $\varepsilon > 0$ um den Punkt c_0 gilt

$$\lim_{n \rightarrow \infty} F_n(B_\varepsilon^c(c_0)) = 0.$$

Es sei $\varepsilon > 0$ fixiert und $(f_n)_{n \in \mathbb{N}}$ die Folge der zugehörigen Dichtefunktionen (vgl. Bemerkung 6.11). Wir werden zeigen:

1. für alle $y \in B_\varepsilon^c(c_0)$ gilt $\lim_{n \rightarrow \infty} f_n(y) = 0$,
2. es existiert eine integrierbare Majorante $g : B_\varepsilon^c(c_0) \rightarrow [0, \infty]$ mit $|f_n| \leq g$,

so dass die Behauptung aus dem Satz von der dominierten Konvergenz folgt.

Mittels algebraischer Umformungen erhalten wir die folgende Abschätzung für die Dichte f_n einer verallgemeinerten t-Verteilung $GMt(nh_n, 1, p, n^{-1}d_n, c_n, nB_n)$ (vgl. Bemerkung 6.11):

$$f_n(y) \leq \pi^{-\frac{p}{2}} \prod_{i=1}^c \left(\frac{nh_n - p}{2} + i \right) (sp[B_n]d_n)^{-\frac{p}{2}} (1 + p^{-1}(sp[B_n]d_n)^{-1} \|y - c_n\|^2)^{-\frac{nh_n}{2}},$$

wobei $c \in \mathbb{N}$ mit $c \geq \frac{p}{2}$ ist. Nach Voraussetzung sind $a_n := \pi^{-\frac{p}{2}} \prod_{i=1}^c (h_n + \frac{2i-p}{2n}) > 0$ und $b_n = sp[B_n]d_n > 0$ konvergente Folgen mit den Grenzwerten $a_0 > 0$, $b_0 \geq 0$ und es gilt

$$f_n(y) \leq n^c a_n b_n^{-\frac{p}{2}} (1 + (pb_n)^{-1} \|y - c_n\|^2)^{-\frac{nh_n}{2}}.$$

Für alle $y \in B_\varepsilon^c(c_0)$ existieren dann Konstanten $q > 1$ und $M < \infty$, so dass für alle $n \geq n_0$ gilt

$$f_n(y) \leq M n^c q^{-\frac{n}{2}}.$$

Insbesondere gilt somit der erste Teil der benötigten Voraussetzungen. Andererseits existiert eine Konstante $M < \infty$, so dass für alle $n \geq n_0$ gilt

$$f_n(y) \leq \frac{M}{(\|y - c_0\| - \frac{\varepsilon}{2})^{p+2}} =: g(y) \quad \text{für alle } y \in B_\varepsilon^c(c_0).$$

Damit folgt der zweite Teil, da die Majorante g integrierbar ist. Somit sind die Voraussetzungen für den Satz von der dominierten Konvergenz erfüllt und es folgt die Behauptung. \square

Es sei M ein $k \times l$ Matrix mit $\text{Rang}[M] = l$. Wir bezeichnen mit $P_{\mathcal{R}(M)}$ die Projektion $M(M^t M)^{-1} M^t$ auf den Bildraum $\mathcal{R}(M)$ und mit $P_{\mathcal{N}(M)}$ die Projektion $I_k - M(M^t M)^{-1} M^t$ auf den Nullraum $\mathcal{N}(M)$ der linearen Abbildung M . Des weiteren bezeichnen wir mit M^+ die Moore-Penrose Inverse $(M^t M)^{-1} M^t$ von M .

Lemma 6.19 *Es sei $X \in \mathbb{R}^{n \times p}$ gegeben. Dann gilt*

$$\begin{aligned} \bar{B}_* &= \lim_{r \rightarrow \infty} \tilde{B}_{*r} = Z^+ X, \\ \bar{\Sigma}_* &= \lim_{r \rightarrow \infty} \tilde{\Sigma}_{*r} = \frac{1}{n} X^t P_{\mathcal{N}(Z)} X. \end{aligned}$$

BEWEIS : Unter den Voraussetzungen des Satzes 6.12 gilt

$$\begin{aligned}\tilde{S}_{*r} &= S_0 + (X_r - Z_r A_0)^t \Omega_{*r}^{-1} (X_r - Z_r A_0), \\ \tilde{A}_{*r} &= A_0 + M_0 Z_r^t \Omega_{*r}^{-1} (X_r - Z_r A_0), \\ \tilde{\Delta}_{*r} &= M_0 - M_0 Z_r^t \Omega_{*r}^{-1} Z_r M_0, \\ \tilde{\rho}_{*r} &= sp[\tilde{\Delta}_{*r} (\frac{1}{rn} Z_r^t Z_r - \frac{rn}{\alpha + rn} \bar{Z}_r^t \bar{Z}_r)].\end{aligned}$$

Die umfangreichen algebraischen Berechnungen, die zu den Grenzwerten

$$\begin{aligned}\bar{S}_* &= \lim_{r \rightarrow \infty} r^{-1} \tilde{S}_{*r} = X^t P_{\mathcal{N}(Z)} X, \\ \bar{A}_* &= \lim_{r \rightarrow \infty} \tilde{A}_{*r} = Z^+ X, \\ \bar{\Delta}_* &= \lim_{r \rightarrow \infty} r \tilde{\Delta}_{*r} = (Z^t Z)^{-1}, \\ \bar{\rho}_* &= \lim_{r \rightarrow \infty} r \tilde{\rho}_{*r} = sp[\bar{\Delta}_* (\frac{1}{n} Z^t Z - \bar{Z}^t \bar{Z})]\end{aligned}$$

führen, werden im Lemma 6.25 im Anschluss an diesen Abschnitt dargelegt. In Lemma 6.25 zeigen wir außerdem $\bar{Z} \bar{A}_* = \bar{X}$ und $\bar{S}_* = (X - Z \bar{A}_*)^t (X - Z \bar{A}_*)$ für

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{und} \quad \bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i.$$

Insbesondere folgt nun mit $\lim_{r \rightarrow \infty} \widetilde{\Lambda}_{*r}^{-1} = \frac{1}{n} \bar{S}_*$ und

$$\begin{aligned}\bar{B}_* &= \lim_{r \rightarrow \infty} \tilde{B}_{*r} = \lim_{r \rightarrow \infty} \tilde{A}_{*r} + \lim_{r \rightarrow \infty} Z_r^+ \frac{rn}{\alpha + rn} \mathbb{1}_{rn} (\bar{X}_r - \bar{Z}_r \tilde{A}_{*r}) = \bar{A}_* = Z^+ X, \\ \bar{\Sigma}_* &= \lim_{r \rightarrow \infty} \tilde{\Sigma}_{*r} \\ &= \lim_{r \rightarrow \infty} \frac{\alpha + rn \tilde{\rho}_{*r}}{\alpha + rn + 1} \widetilde{\Lambda}_{*r}^{-1} - \lim_{r \rightarrow \infty} \frac{(rn)^2}{(\alpha + rn)(\alpha + rn + 1)} (\bar{X}_r - \bar{Z}_r \tilde{A}_{*r})^t (\bar{X}_r - \bar{Z}_r \tilde{A}_{*r}) \\ &\quad + \lim_{r \rightarrow \infty} \frac{1}{\alpha + rn + 1} (X_r - Z_r \tilde{A}_{*r})^t (X_r - Z_r \tilde{A}_{*r}) \\ &= \frac{1}{n} (X - Z \bar{A}_*)^t (X - Z \bar{A}_*) = X^t P_{\mathcal{N}(Z)} X\end{aligned}$$

die Behauptung. □

Lemma 6.20 *Es sei $X \in \mathbb{R}^{n \times p}$ gegeben. Dann gilt*

$$\begin{aligned}\bar{B}_{[v]} &= \lim_{r \rightarrow \infty} \tilde{B}_{[v]r} = Z^+ X + Z^+ C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X, \\ \bar{\Sigma}_{[v]} &= \lim_{r \rightarrow \infty} \tilde{\Sigma}_{[v]r} = \frac{1}{n} X^t P_{\mathcal{N}(Z)} [I_n + C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t] P_{\mathcal{N}(Z)} X,\end{aligned}$$

mit $D_{[v]} = (C_{[v]}^t Z (Z^t Z)^{-1} Z^t C_{[v]} + B_{[v]} B_{[v]}^t)^{-\frac{1}{2}}$ für alle $[v] = (C_{[v]}, B_{[v]}) \in [D]$.

BEWEIS : Es sei $[v] = (C_{[v]}, B_{[v]}) \in [D]$ fixiert und für alle $r \in \mathbb{N}$ sei v_r ein Element der zugehörigen Äquivalenzklasse. Unter den Voraussetzungen des Satzes 6.12 gilt

$$\begin{aligned}\tilde{S}_{v_r} &= S_0 + (X_r - Z_r A_0)^t \Omega_{v_r}^{-1} (X_r - Z_r A_0), \\ \tilde{A}_{v_r} &= A_0 + M_0 Z_r^t \Omega_{v_r}^{-1} (X_r - Z_r A_0), \\ \tilde{\Delta}_{v_r} &= M_0 - M_0 Z_r^t \Omega_{v_r}^{-1} Z_r M_0, \\ \tilde{\rho}_{v_r} &= sp[\tilde{\Delta}_{v_r} (\frac{1}{rn} Z_r^t Z_r - \frac{rn}{\alpha + rn} \bar{Z}_r^t \bar{Z}_r)].\end{aligned}$$

Die umfangreichen algebraischen Berechnungen, die zu den Grenzwerten

$$\begin{aligned}\bar{S}_{[v]} &= \lim_{r \rightarrow \infty} r^{-1} \tilde{S}_{v_r} = X^t P_{\mathcal{N}(Z)} (I_n + C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t) P_{\mathcal{N}(Z)} X, \\ \bar{A}_{[v]} &= \lim_{r \rightarrow \infty} \tilde{A}_{v_r} = Z^+ X + Z^+ C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X, \\ \bar{\Delta}_{[v]} &= \lim_{r \rightarrow \infty} r \tilde{\Delta}_{v_r} = Z^+ (I_n - C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t) Z^{+t}, \\ \bar{\rho}_{[v]} &= \lim_{r \rightarrow \infty} r \tilde{\rho}_{v_r} = sp[\bar{\Delta}_{[v]} (\frac{1}{n} Z^t Z - \bar{Z}^t \bar{Z})]\end{aligned}$$

mit $D_{[v]} = (C_{[v]}^t Z (Z^t Z)^{-1} Z^t C_{[v]} + B_{[v]} B_{[v]}^t)^{-\frac{1}{2}}$ führen, werden im Lemma 6.26 im Anschluss an diesen Abschnitt dargelegt. In Lemma 6.26 zeigen wir außerdem $\bar{Z} \bar{A}_{[v]} = \bar{X}$ und $\bar{S}_{[v]} = (X - Z \bar{A}_{[v]})^t (X - Z \bar{A}_{[v]})$. Insbesondere folgt nun mit $\lim_{r \rightarrow \infty} \tilde{\Lambda}_{v_r}^{-1} = \frac{1}{n} \bar{S}_{[v]}$ und

$$\begin{aligned}\bar{B}_{[v]} &= \lim_{r \rightarrow \infty} \tilde{B}_{v_r} = \lim_{r \rightarrow \infty} \tilde{A}_{v_r} + \lim_{r \rightarrow \infty} \frac{rn}{\alpha + rn} Z_r^+ \mathbb{1}_{rn} (\bar{X}_r - \bar{Z}_r \tilde{A}_{v_r}) = \bar{A}_{[v]}, \\ \bar{\Sigma}_{[v]} &= \lim_{r \rightarrow \infty} \tilde{\Sigma}_{v_r} \\ &= \lim_{r \rightarrow \infty} \frac{\alpha + rn \tilde{\rho}_{v_r}}{\alpha + rn + 1} \tilde{\Lambda}_{v_r}^{-1} - \lim_{r \rightarrow \infty} \frac{(rn)^2}{(\alpha + rn)(\alpha + rn + 1)} (\bar{X}_r - \bar{Z}_r \tilde{A}_{v_r})^t (\bar{X}_r - \bar{Z}_r \tilde{A}_{v_r}) \\ &\quad + \lim_{r \rightarrow \infty} \frac{1}{\alpha + rn + 1} (X_r - Z_r \tilde{A}_{v_r})^t (X_r - Z_r \tilde{A}_{v_r}) \\ &= \frac{1}{n} (X - Z \bar{A}_{[v]})^t (X - Z \bar{A}_{[v]})\end{aligned}$$

die Behauptung. □

Im Folgenden bezeichnen wir für jedes $i = 1, \dots, n$ und für alle $X \in \mathbb{R}^{n \times p}$, $B \in \mathbb{R}^{k \times p}$ mit $\hat{F}_{B, X}^i$ die empirische Verteilung der Werte $(X_1 - Z_1 B + Z_i B, \dots, X_n - Z_n B + Z_i B)$.

Lemma 6.21 *Es sei $X \in \mathbb{R}^{n \times p}$ mit $|X^t P_{\mathcal{N}(Z)} X| > 0$ gegeben. Dann gilt für $i = 1, \dots, n$*

$$\tilde{P}_{i*} \xrightarrow{w} \hat{F}_{B^*, X}^i \quad \text{für } r \rightarrow \infty$$

und für alle $[v] \in [D]$

$$\tilde{P}_{iv_r} \xrightarrow{w} \hat{F}_{B_{[v]}, X}^i \quad \text{für } r \rightarrow \infty.$$

BEWEIS : Es sei $[v] = (C_{[v]}, B_{[v]}) \in [D]$ fixiert und für alle $r \in \mathbb{N}$ sei v_r ein Element der zugehörigen Äquivalenzklasse. Unter den Voraussetzungen des Satzes 6.12 gilt für $i = 1, \dots, n$

$$\tilde{P}_{iv_r} = \frac{r\alpha}{\alpha + rn} \left[\frac{1}{r} F_{iv_r} \right] + \frac{rn}{\alpha + rn} \left[\frac{1}{rn} \sum_{j=1}^{rn} G_{ijv_r} \right].$$

Wobei unter der Annahme $X_r = \mathbb{1}_r \otimes X$ und $Z_r = \mathbb{1}_r \otimes Z$ in der zweiten Summe die ersten n Summanden r -mal auftreten, da unter der Annahme A.6.2

$$G_{ijv_r} = \begin{cases} \delta_{X_j - (Z_j - Z_i) \tilde{A}_{v_r}} & : h_{ijv_r} = 0 \\ GMt(rn + l - k, 1, p, h_{ijv_r}, \tilde{S}_{v_r}, X_j - (Z_j - Z_i) \tilde{A}_{v_r}) & : h_{ijv_r} > 0. \end{cases}$$

gilt. Wegen Lemma 6.26 existieren nun für alle $[v] \in [D]$ die Grenzwerte

$$\bar{A}_{[v]} = \lim_{r \rightarrow \infty} \tilde{A}_{[v]r}, \quad \bar{S}_{[v]} = \lim_{r \rightarrow \infty} r^{-1} \tilde{S}_{[v]r}, \quad \bar{\Delta}_{[v]} = \lim_{r \rightarrow \infty} r \tilde{\Delta}_{[v]r},$$

wobei $\bar{\Delta}_{[v]}$ eine positiv semidefinite und $\bar{S}_{[v]} = \bar{S}_* + H_{[v]}$ für eine geeignet gewählte positiv semidefinite Matrix $H_{[v]}$ (vgl. Beweis des Lemmas 6.26) nach Voraussetzung eine positiv definite Matrix ist. Zusammen mit der a priori Annahme $l - k > p$ sind damit die Voraussetzungen für das Lemma 6.18 erfüllt und es folgt für $j = 1, \dots, n$ die schwache Konvergenz

$$\begin{aligned} \delta_{X_j + (Z_i - Z_j)\bar{A}_{[v]}} &= \lim_{r \rightarrow \infty} \delta_{X_j + (Z_i - Z_j)\tilde{A}_{[v]r}}, \\ \delta_{X_j + (Z_i - Z_j)\bar{A}_{[v]}} &= \lim_{r \rightarrow \infty} \text{GMt}\left(rn + l - k, 1, p, r^{-1}(rh_{ij[v]r}), r(r^{-1}\tilde{S}_{[v]r}), X_j + (Z_i - Z_j)\tilde{A}_{[v]r}\right). \end{aligned}$$

Insbesondere folgt nun die Behauptung mit

$$\tilde{P}_{i[v]r} \xrightarrow{w} \frac{1}{n} \sum_{j=1}^n \delta_{X_j + (Z_i - Z_j)\bar{A}_{[v]}} \quad \text{für } r \rightarrow \infty.$$

Analoge Schritte führen auf Grund der Existenz der Grenzwerte der auftretenden Folgen (vgl. Lemma 6.25) zu der Behauptung $\tilde{P}_{i*} \xrightarrow{w} \hat{F}_{B^*, X}^i$ für $r \rightarrow \infty$. \square

Lemma 6.22 *Es sei $X \in \mathbb{R}^{n \times p}$ mit $|X^t P_{N(Z)} X| > 0$ gegeben. Dann gilt*

$$\lim_{r \rightarrow \infty} k_{*r} = \bar{k}_* \in [0, 1] \quad \text{und} \quad \lim_{r \rightarrow \infty} \tilde{k}_{[v]r} = \bar{k}_{[v]} \in [0, 1]$$

für alle $[v] \in [D]$. Insbesondere gilt für alle $[v] \in [D]$

$$\bar{k}_{[v]} > 0 \quad \Rightarrow \quad \bar{S}_{[v]} = \bar{S}_*.$$

BEWEIS : Es sei $[v] = (C_{[v]}, B_{[v]}) \in [D]$ fixiert und für alle $r \in \mathbb{N}$ sei v_r ein Element der zugehörigen Äquivalenzklasse. Wir betrachten für $X_r = \mathbb{1}_r \otimes X$ und $Z_r = \mathbb{1}_r \otimes Z$ das Gewicht (vgl. (6.22))

$$\begin{aligned} \tilde{k}_{[v]r} &= \left[\frac{\alpha_{*r}}{\alpha_{[v]r}} \frac{1}{r^{n_{[v]}} d_{[v]r}} \left(\frac{|\Omega_{[v]r}|}{|\Omega_{*r}|} \right)^{\frac{p}{2}} \left(\frac{|\tilde{S}_{[v]r}|}{|\tilde{S}_{*r}|} \right)^{\frac{rn+l-k}{2}} + 1 + \right. \\ &\quad \left. + \sum_{[l] \in [D] \setminus \{[v]\}} \frac{r^{n_{[l]} d_{[l]r}}}{r^{n_{[v]} d_{[v]r}} \alpha_{[v]r}} \left(\frac{|\Omega_{[v]r}|}{|\Omega_{[l]r}|} \right)^{\frac{p}{2}} \left(\frac{|\tilde{S}_{[v]r}|}{|\tilde{S}_{[l]r}|} \right)^{\frac{rn+l-k}{2}} \right]^{-1}. \end{aligned}$$

Insbesondere zeigen wir im Lemma 6.28 am Ende dieses Abschnittes

$$|\Omega_{*r}| = r^k h_{*r} \quad \text{und} \quad |\Omega_{[v]r}| = r^{k - (n_{[v]} - q_{[v]})} h_{[v]r}$$

für positive Folgen h_{*r} und $h_{[v]r}$ mit den positiven Grenzwerten h_* und $h_{[v]}$. Weiterhin gilt nun (vgl. Annahme A.6.2))

$$\frac{\alpha_{*r}}{\alpha_{[v]r}} = \frac{\alpha^{n_{[v]} - q_{[v]}}}{\prod_{i=1}^{q_{[v]}} (n_{i[v]} - 1)!} \quad \text{und} \quad \frac{\alpha_{[l]r}}{\alpha_{[v]r}} = \frac{\alpha^{n_{[v]} - q_{[v]}} \prod_{i=1}^{q_{[l]}} (n_{i[l]} - 1)!}{\alpha^{n_{[l]} - q_{[l]}} \prod_{i=1}^{q_{[v]}} (n_{i[v]} - 1)!} \quad \text{für alle } [l] \in [D].$$

Ferner existieren dann positive reellwertige Folgen $g_r(*, [v])$ und $g_r([l], [v])$ mit den positiven Grenzwerten $g(*, [v])$ und $g([l], [v])$ für alle $[l] \in [D]$, so dass der erste Teil der Behauptung aus

$$\begin{aligned} \tilde{k}_{[v]r} &= \left[g_r(*, [v]) \frac{1}{r^{n_{[v]} + (n_{[v]} - q_{[v]}) \frac{p}{2}}} \left(\frac{|\tilde{S}_{[v]r}|}{|\tilde{S}_{*r}|} \right)^{\frac{rn+l-k}{2}} + 1 + \right. \\ &\quad \left. + \sum_{[l] \in [D] \setminus \{[v]\}} g_r([l], [v]) \frac{r^{n_{[l]} + (n_{[l]} - q_{[l]}) \frac{p}{2}}}{r^{n_{[v]} + (n_{[v]} - q_{[v]}) \frac{p}{2}}} \left(\frac{|\tilde{S}_{[v]r}|}{|\tilde{S}_{[l]r}|} \right)^{\frac{rn+l-k}{2}} \right]^{-1} \end{aligned}$$

und der Eigenschaft $\bar{S}_{[v]} = \bar{S}_* + H_{[v]}$ für eine geeignet gewählte positiv semidefinite Matrix $H_{[v]}$ (vgl. Beweis des Lemmas 6.26) folgt, wobei $\bar{S}_* = n\bar{\Sigma}_*$ nach Voraussetzung positiv definit ist. Insbesondere gilt $\lim_{r \rightarrow \infty} \tilde{k}_{[v]r} = \bar{k}_{[v]} = 0$ für $|\bar{S}_{[v]}| > |\bar{S}_*|$, so dass der zweite Teil der Behauptung folgt. \square

Satz 6.23 *Es sei $X \in \mathbb{R}^{n \times p}$ mit $|X^t P_N(Z)X| > 0$ gegeben. Dann sind die selbstinformativen Grenzwerte \bar{B} , $\bar{\Sigma}$, \bar{P}_i der Bayesschen Schätzungen der Parameter B , Σ und der Verteilung P_i für $i = 1, \dots, n$:*

$$\begin{aligned}\bar{B} &= (Z^t Z)^{-1} Z^t X, \\ \bar{\Sigma} &= \frac{1}{n} X^t (I_n - Z(Z^t Z)^{-1} Z^t) X, \\ \bar{P}_i &= \frac{1}{n} \sum_{j=1}^n \delta_{X_j - (Z_j - Z_i)\bar{B}}.\end{aligned}$$

BEWEIS : Die Bayesschen Schätzungen für B , Σ und P_i , $i = 1, \dots, n$, sind (vgl. Satz 6.12)

$$\tilde{B}_r = \sum_{v_r \in D_r} k_{v_r} \tilde{B}_{v_r}, \quad \tilde{\Sigma}_r = \sum_{v_r \in D_r} k_{v_r} \tilde{\Sigma}_{v_r}, \quad \tilde{P}_{ir} = \sum_{v_r \in D_r} k_{v_r} \tilde{P}_{iv_r}.$$

Dann existiert eine von r unabhängige Menge $[D]$ (vgl. Lemma 6.15 und 6.17), so dass für die Bayesschen Schätzungen mit dem speziellen Wert $\tilde{X}^r = \mathbf{1}_r \otimes X$

$$\tilde{B}_{(r)} = k_{*r} \tilde{B}_{*r} + \sum_{[v] \in [D]} \tilde{k}_{[v]r} \tilde{B}_{[v]r}, \quad \tilde{\Sigma}_{(r)} = k_{*r} \tilde{\Sigma}_{*r} + \sum_{[v] \in [D]} \tilde{k}_{[v]r} \tilde{\Sigma}_{[v]r}, \quad \tilde{P}_{i(r)} = k_{*r} \tilde{P}_{i*} + \sum_{[v] \in [D]} \tilde{k}_{[v]r} \tilde{P}_{i[v]r}$$

gilt. Weiterhin existieren wegen Lemma 6.19 - 6.22 für $r \rightarrow \infty$ die Grenzwerte der Folgen $(k_{*r})_{r \in \mathbb{N}}$, $(\tilde{B}_{*r})_{r \in \mathbb{N}}$, $(\tilde{\Sigma}_{*r})_{r \in \mathbb{N}}$, $(\tilde{P}_{i*r})_{r \in \mathbb{N}}$ und $(\tilde{k}_{[v]r})_{r \in \mathbb{N}}$, $(\tilde{B}_{[v]r})_{r \in \mathbb{N}}$, $(\tilde{\Sigma}_{[v]r})_{r \in \mathbb{N}}$, $(\tilde{P}_{i[v]r})_{r \in \mathbb{N}}$ für alle $[v] \in [D]$ und es folgt

$$\begin{aligned}\lim_{r \rightarrow \infty} \tilde{B}_r &= \lim_{r \rightarrow \infty} \sum_{v \in D_r} k_{v_r} \tilde{B}_{v_r} = \bar{k}_* \bar{B}_* + \sum_{[v] \in [D]} \bar{k}_{[v]} \bar{B}_{[v]}, \\ \lim_{r \rightarrow \infty} \tilde{\Sigma}_r &= \lim_{r \rightarrow \infty} \sum_{v \in D_r} k_{v_r} \tilde{\Sigma}_{v_r} = \bar{k}_* \bar{\Sigma}_* + \sum_{[v] \in [D]} \bar{k}_{[v]} \bar{\Sigma}_{[v]}, \\ \lim_{r \rightarrow \infty} \tilde{P}_{ir} &= \lim_{r \rightarrow \infty} \sum_{v \in D_r} k_{v_r} \tilde{P}_{iv_r} = \bar{k}_* \hat{F}_{\bar{B}_*, X}^i + \sum_{[v] \in [D]} \bar{k}_{[v]} \hat{F}_{\bar{B}_{[v]}, X}^i.\end{aligned}$$

Insbesondere gilt für alle $[v] \in [D]$ nun $\bar{B}_{[v]} = \bar{B}_{[*]}$ für $\bar{S}_{[v]} = \bar{S}_*$ (vgl. Lemma 6.26). Falls $\bar{k}_{[v]}$ positiv ist, folgt wegen Lemma 6.22 dann $\bar{B}_{[v]} = \bar{B}_{[*]}$, $\bar{\Sigma}_{[v]} = \bar{\Sigma}_{[*]}$, $\hat{F}_{\bar{B}_{[v]}, X}^i = \hat{F}_{\bar{B}_{[*]}, X}^i$ und somit die Behauptung. \square

Bemerkung 6.24 Die selbstinformativen Grenzwerte für die Parameter B und Σ entsprechen somit den MLS'en bzw. Kleinste Quadrate Lösungen in einem multivariaten normalen linearen Modell (vgl. Abschnitt 4.6.3). Damit sind die selbstinformativen Grenzwerte für B und Σ in dem semiparametrischen linearen Modell unabhängig von der Wahl der a priori Parameter der Normal-Wishart-Verteilung. Dagegen wählten wir als einen a priori Parameter für den Dirichlet Prozess eine Normalverteilung, was für eine Abhängigkeit von der a priori Verteilung des Dirichlet Prozesses spricht. Andererseits ist der selbstinformativ Grenzwert der unbekannt Randverteilung immer eine vMLS nach Gill (vgl. Bemerkung 6.8) und in einigen Fällen auch eine vMLS nach Kiefer und Wolfowitz (vgl. Bemerkung 6.5), so dass wir keine Aussage zum Einfluss der Wahl der a priori Parameter für den Dirichlet Prozess geben können. Im nächsten Abschnitt betrachten wir abschließend das semiparametrische Lokationsmodell, in dem die drei vorgestellten Ansätze eine identische Lösung liefern. \square

Algebraische Berechnungen

Lemma 6.25 Für alle $r \in \mathbb{N}$ seien (vgl. Annahme A.6.2)

$$\begin{aligned}\tilde{S}_{*r} &= S_0 + (X_r - Z_r A_0)^t \Omega_{*r}^{-1} (X_r - Z_r A_0), \\ \tilde{A}_{*r} &= A_0 + M_0 Z_r^t \Omega_{*r}^{-1} (X_r - Z_r A_0), \\ \tilde{\Delta}_{*r} &= M_0 - M_0 Z_r^t \Omega_{*r}^{-1} Z_r M_0, \\ \tilde{\rho}_{*r} &= sp[\tilde{\Delta}_{*r} (\frac{1}{rn} Z_r^t Z_r - \frac{rn}{\alpha + rn} \overline{Z}_r^t \overline{Z}_r)].\end{aligned}$$

Dann existieren die Grenzwerte

$$\begin{aligned}\overline{S}_* &:= \lim_{r \rightarrow \infty} r^{-1} \tilde{S}_{*r} = X^t P_{N(Z)} X, \\ \overline{A}_* &:= \lim_{r \rightarrow \infty} \tilde{A}_{*r} = Z^+ X, \\ \overline{\Delta}_* &:= \lim_{r \rightarrow \infty} r \tilde{\Delta}_{*r} = (Z^t Z)^{-1}, \\ \overline{\rho}_* &:= \lim_{r \rightarrow \infty} r \tilde{\rho}_{*r} = sp[\overline{\Delta}_* (\frac{1}{n} Z^t Z - \overline{Z}^t \overline{Z})]\end{aligned}$$

und es gilt $\overline{Z \overline{A}_*} = \overline{X}$ sowie $\overline{S}_* = (X - Z \overline{A}_*)^t (X - Z \overline{A}_*)$. Weiterhin ist $\overline{\Delta}_*$ eine positiv definite Matrix.

BEWEIS : Für die Matrix $\Omega_{*r} := Z_r M_0 Z_r^t + I_{rn}$ gilt mit $X_r = \mathbb{1}_r \otimes X$ und $Z_r = \mathbb{1}_r \otimes Z$

$$\begin{aligned}X_r^t \Omega_{*r}^{-1} X_r &= X_r^t [I_{rn} - Z_r (M_0^{-1} + Z_r^t Z_r)^{-1} Z_r^t] X_r \\ &= r (X^t X - X^t Z (r^{-1} M_0^{-1} + Z^t Z)^{-1} Z^t X).\end{aligned}$$

Insbesondere folgt dann mit $P_{N(Z)} := I_n - Z (Z^t Z)^{-1} Z^t$ und $Z^+ = (Z^t Z)^{-1} Z^t$

$$\begin{aligned}X^t P_{N(Z)} X &= \lim_{r \rightarrow \infty} r^{-1} X_r^t \Omega_{*r}^{-1} X_r, \\ M_0^{-1} Z^+ X &= \lim_{r \rightarrow \infty} Z_r^t \Omega_{*r}^{-1} X_r, \\ M_0^{-1} &= \lim_{r \rightarrow \infty} Z_r^t \Omega_{*r}^{-1} Z_r, \\ (Z^t Z)^{-1} &= \lim_{r \rightarrow \infty} r [M_0 - M_0 Z_r^t \Omega_{*r}^{-1} Z_r M_0].\end{aligned}$$

Zusammen mit

$$\begin{aligned}r^{-1} \tilde{S}_{*r} &= r^{-1} S_0 + r^{-1} (X_r - Z_r A_0)^t \Omega_{*r}^{-1} (X_r - Z_r A_0) \\ &= r^{-1} S_0 + r^{-1} A_0^t Z_r^t \Omega_{*r}^{-1} Z_r A_0 - r^{-1} X_r^t \Omega_{*r}^{-1} Z_r A_0 - r^{-1} A_0^t Z_r^t \Omega_{*r}^{-1} X_r + r^{-1} X_r^t \Omega_{*r}^{-1} X_r, \\ \tilde{A}_{*r} &= A_0 + M_0 Z_r^t \Omega_{*r}^{-1} (X_r - Z_r A_0) = A_0 - M_0 Z_r^t \Omega_{*r}^{-1} Z_r A_0 + M_0 Z_r^t \Omega_{*r}^{-1} X_r, \\ \tilde{\Delta}_{*r} &= M_0 - M_0 Z_r^t \Omega_{*r}^{-1} Z_r M_0, \\ r \tilde{\rho}_{*r} &= sp[r \tilde{\Delta}_{*r} (\frac{1}{rn} Z_r^t Z_r - \frac{rn}{\alpha + rn} \overline{Z}_r^t \overline{Z}_r)]\end{aligned}$$

folgt nun der erste Teil der Behauptung. Unter der Voraussetzung $\text{Rang}[Z] = k$ ist die Matrix $\overline{\Delta}_* = (Z^t Z)^{-1}$ positiv definit. Weiterhin folgt wegen $\mathbb{1}_n \in \mathcal{R}(Z)$

$$n \overline{Z \overline{A}_*} = \mathbb{1}_n^t Z \overline{A}_* = \mathbb{1}_n^t Z Z^+ X = \mathbb{1}_n^t X = n \overline{X},$$

und aus der Eigenschaft $P_{N(Z)} = P_{N(Z)} P_{N(Z)}$ der Projektionsmatrix $P_{N(Z)}$

$$\overline{S}_* = X^t P_{N(Z)} X = X^t P_{N(Z)} P_{N(Z)} X = (X - Z \overline{A}_*)^t (X - Z \overline{A}_*)$$

der zweite Teil der Behauptung. □

Lemma 6.26 *Es sei $[v] \in [D]$ beliebig fixiert und für alle $r \in \mathbb{N}$ sei v_r ein Element aus der Äquivalenzklasse mit dem Repräsentanten $[v]$. Weiterhin seien (vgl. Annahme A.6.2)*

$$\begin{aligned}\tilde{S}_{v_r} &= S_0 + (X_r - Z_r A_0)^t \Omega_{v_r}^{-1} (X_r - Z_r A_0), \\ \tilde{A}_{v_r} &= A_0 + M_0 Z_r^t \Omega_{v_r}^{-1} (X_r - Z_r A_0), \\ \tilde{\Delta}_{v_r} &= M_0 - M_0 Z_r^t \Omega_{v_r}^{-1} Z_r M_0, \\ \tilde{\rho}_{v_r} &= sp[\tilde{\Delta}_{v_r} (\frac{1}{rn} Z_r^t Z_r - \frac{rn}{\alpha + rn} \bar{Z}_r^t \bar{Z}_r)].\end{aligned}$$

Dann existieren die Grenzwerte

$$\begin{aligned}\bar{S}_{[v]} &= \lim_{r \rightarrow \infty} r^{-1} \tilde{S}_{v_r} = X^t P_{\mathcal{N}(Z)} (I_n + C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t) P_{\mathcal{N}(Z)} X, \\ \bar{A}_{[v]} &= \lim_{r \rightarrow \infty} \tilde{A}_{v_r} = Z^+ X + Z^+ C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X, \\ \bar{\Delta}_{[v]} &= \lim_{r \rightarrow \infty} r \tilde{\Delta}_{v_r} = Z^+ (I_n - C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t) (Z^+)^t, \\ \bar{\rho}_{[v]} &= \lim_{r \rightarrow \infty} r \tilde{\rho}_{v_r} = sp[\bar{\Delta}_{[v]} (\frac{1}{n} Z^t Z - \bar{Z}^t \bar{Z})]\end{aligned}$$

mit $D_{[v]} = (C_{[v]}^t Z (Z^t Z)^{-1} Z^t C_{[v]} + B_{[v]} B_{[v]}^t)^{-\frac{1}{2}}$. Weiterhin ist $\bar{\Delta}_{[v]}$ eine positiv semidefinite Matrix und es gilt $\bar{Z} \bar{A}_{[v]} = \bar{X}$, $\bar{S}_{[v]} = (X - Z \bar{A}_{[v]})^t (X - Z \bar{A}_{[v]})$ sowie $\bar{S}_{[v]} = \bar{S}_* \Leftrightarrow \bar{A}_{[v]} = \bar{A}_*$.

BEWEIS : Dazu betrachten wir die Matrix

$$\Omega_{v_r} = Z_r M_0 Z_r^t + K_{v_r} = Z_r M_0 Z_r^t + I_{rn} + C_{v_r} (B_{v_r} B_{v_r}^t - I_{n_{v_r}}) C_{v_r}^t.$$

Es sei nun $W_{v_r} = B_{v_r} B_{v_r}^t - I_{n_{v_r}}$ und J_q die $q \times q$ Eins-Matrix. Dann gilt

$$W_{v_r}^{-1} = \text{Diag}[(n_{1v_r} - 1)^{-1} J_{n_{1v_r}}, \dots, (n_{q_{v_r} v_r} - 1)^{-1} J_{n_{q_{v_r} v_r}}] - I_{n_{v_r}}.$$

Ferner gilt $\Omega_{*r} = Z_r M_0 Z_r^t + I_{rn}$, so dass wegen $(A^{-1} + CB^{-1}C^t)^{-1} = A - AC(C^t AC + B)^{-1}C^t A$ (vgl. Humak (1977), S. 480, A 1.21)

$$\begin{aligned}\Omega_{v_r}^{-1} &= \Omega_{*r}^{-1} - \Omega_{*r}^{-1} C_{v_r} (C_{v_r}^t \Omega_{*r}^{-1} C_{v_r} + W_{v_r}^{-1})^{-1} C_{v_r}^t \Omega_{*r}^{-1} \\ &= \Omega_{*r}^{-1} + \Omega_{*r}^{-1} C_{v_r} H_{v_r}^{-1} C_{v_r}^t \Omega_{*r}^{-1}\end{aligned}$$

folgt, wobei $H_{v_r} = C_{v_r}^t Z_r (M_0^{-1} + Z_r^t Z_r)^{-1} Z_r^t C_{v_r} + B_{v_r} \text{Diag}[\frac{1}{1-n_{1v_r}}, \dots, \frac{1}{1-n_{q_{v_r} v_r}}] B_{v_r}^t$ ist. Insbesondere folgt nun wegen $C_{[v]}^t Z = C_{v_r}^t Z_r$ und $B_{[v]} = B_{v_r}$

$$\begin{aligned}H_{v_r}^{-1} &= r [C_{[v]}^t Z (r^{-1} M_0^{-1} + Z^t Z)^{-1} Z^t C_{[v]} - B_{[v]} \text{Diag}[\frac{r}{n_{1[v]} - 1}, \dots, \frac{r}{n_{q_{[v]} [v]} - 1}] B_{[v]}^t]^{-1} \\ &= r [Q_{[v]r} - B_{[v]} L_{[v]r} B_{[v]}^t]^{-1}\end{aligned}$$

wobei die Matrizen

$$\begin{aligned}Q_{[v]r} &:= C_{[v]}^t Z (\frac{1}{r} M_0^{-1} + Z^t Z)^{-1} Z^t C_{[v]} + B_{[v]} B_{[v]}^t, \\ L_{[v]r} &:= I_q + \text{Diag}[\frac{r}{n_{1[v]} - 1}, \dots, \frac{r}{n_{q_{[v]} [v]} - 1}]\end{aligned}$$

für alle $r \in \mathbb{N}$ und $Q_{[v]} := C_{[v]}^t Z(Z^t Z)^{-1} Z^t C_{[v]} + B_{[v]} B_{[v]}^t$ auf Grund der Eigenschaft $\text{Rang}(C_{[v]}^t Z B_{[v]}) = n_{[v]}$ (vgl. Lemma 6.10) und der Annahme $n_{i[v]} > 1$, $i = 1, \dots, q_{[v]}$, positiv definit sind. Ferner gilt nun $Q_{[v]}^{-1} = \lim_{r \rightarrow \infty} Q_{[v]r}^{-1}$ und $0 = \lim_{r \rightarrow \infty} L_{[v]r}^{-1}$. Damit folgt

$$\begin{aligned} \lim_{r \rightarrow \infty} r^{-1} H_{v_r}^{-1} &= \lim_{r \rightarrow \infty} (Q_{[v]r}^{-1} - Q_{[v]r}^{-1} B_{[v]} [B_{[v]}^t Q_{[v]r}^{-1} B_{[v]} - L_{[v]r}^{-1}]^{-1} B_{[v]} Q_{[v]r}^{-1}) \\ &= Q_{[v]}^{-1} - Q_{[v]}^{-1} B_{[v]} [B_{[v]}^t Q_{[v]}^{-1} B_{[v]}]^{-1} B_{[v]}^t Q_{[v]}^{-1}. \end{aligned}$$

Für $D_{[v]} = Q_{[v]}^{-\frac{1}{2}}$ und $P_{\mathcal{N}(D_{[v]} B_{[v]})} := I - D_{[v]} B_{[v]} (B_{[v]}^t D_{[v]} D_{[v]} B_{[v]})^{-1} B_{[v]}^t D_{[v]}$ gilt somit

$$\lim_{r \rightarrow \infty} r^{-1} H_{v_r}^{-1} = D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]}.$$

Außerdem folgt nun wegen $M_0^{-1} Z^+ C_{[v]} = \lim_{r \rightarrow \infty} r Z_r^t \Omega_{v_r}^{-1} C_{v_r}$ und $C_{[v]}^t P_{\mathcal{N}(Z)} X = \lim_{r \rightarrow \infty} C_{v_r}^t \Omega_{v_r}^{-1} X_r$ (vgl. Lemma 6.25)

$$\begin{aligned} \lim_{r \rightarrow \infty} r^{-1} X_r^t \Omega_{v_r}^{-1} X_r &= X^t P_{\mathcal{N}(Z)} X + X^t P_{\mathcal{N}(Z)} C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X, \\ \lim_{r \rightarrow \infty} Z_r^t \Omega_{v_r}^{-1} X_r &= M_0^{-1} Z^+ X + M_0^{-1} Z^+ C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X, \\ \lim_{r \rightarrow \infty} Z_r^t \Omega_{v_r}^{-1} Z_r &= M_0^{-1}, \\ \lim_{r \rightarrow \infty} r [M_0 - M_0 Z_r^t \Omega_{v_r}^{-1} Z_r M_0] &= (Z^t Z)^{-1} - Z^+ C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t (Z^+)^t. \end{aligned}$$

Zusammen mit

$$\begin{aligned} r^{-1} \tilde{S}_{v_r} &= r^{-1} S_0 + r^{-1} A_0^t Z_r^t \Omega_{v_r}^{-1} Z_r A_0 - r^{-1} X_r^t \Omega_{v_r}^{-1} Z_r A_0 - r^{-1} A_0^t Z_r^t \Omega_{v_r}^{-1} X_r + r^{-1} X_r^t \Omega_{v_r}^{-1} X_r, \\ \tilde{A}_{v_r} &= A_0 + M_0 Z_r^t \Omega_{v_r}^{-1} (X_r - Z_r A_0) = A_0 - M_0 Z_r^t \Omega_{v_r}^{-1} Z_r A_0 + M_0 Z_r^t \Omega_{v_r}^{-1} X_r, \\ \tilde{\Delta}_{v_r} &= M_0 - M_0 Z_r^t \Omega_{v_r}^{-1} Z_r M_0, \\ r \tilde{\rho}_{v_r} &= sp[r \tilde{\Delta}_{v_r} (\frac{1}{rn} Z_r^t Z_r - \frac{rn}{\alpha + rn} \bar{Z}_r^t \bar{Z}_r)] \end{aligned}$$

folgt dann der erste Teil der Behauptung. Weiterhin ist $\bar{\Delta}_{[v]}$ positiv semidefinit, da für alle $r \in \mathbb{N}$ die Matrizen $\tilde{\Delta}_{[v]r}$ positiv semidefinit sind. Unter der Voraussetzung $\mathbb{1}_n \in \mathcal{R}(Z)$ folgt

$$\begin{aligned} n \bar{Z} \bar{A}_{[v]} &= \mathbb{1}_n^t Z \bar{A}_{[v]} = \mathbb{1}_n^t Z Z^+ X + \mathbb{1}_n^t Z Z^+ C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X \\ &= \mathbb{1}_n^t X + \mathbb{1}_n^t C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X \\ &= \mathbb{1}_n^t X + \mathbb{1}_{q[v]}^t B_{[v]}^t D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X = \mathbb{1}_n^t X = n \bar{X}, \end{aligned}$$

wobei die zweite Identität wegen $\mathbb{1}_{n[v]} = C_v^t \mathbb{1}_n = B_v \mathbb{1}_{q[v]}$ gilt. Weiterhin folgt

$$\begin{aligned} \bar{S}_{[v]} &= X^t P_{\mathcal{N}(Z)} (I_n + C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t) P_{\mathcal{N}(Z)} X \\ &= X^t P_{\mathcal{N}(Z)} (I_n + C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} D_{[v]}^{-2} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t) P_{\mathcal{N}(Z)} X \\ &= X^t P_{\mathcal{N}(Z)} (I_n + C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} (D_{[v]}^{-2} - B_{[v]} B_{[v]}^t) D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t) P_{\mathcal{N}(Z)} X \\ &= X^t P_{\mathcal{N}(Z)} (I_n + C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{R}(Z)} C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t) P_{\mathcal{N}(Z)} X \\ &= (X - Z \bar{A}_{[v]})^t (X - Z \bar{A}_{[v]}), \end{aligned}$$

wobei die zweite Identität aus der Eigenschaft $P_{\mathcal{N}(D_{[v]} B_{[v]})} = P_{\mathcal{N}(D_{[v]} B_{[v]})} P_{\mathcal{N}(D_{[v]} B_{[v]})}$ der Projektionsmatrix $P_{\mathcal{N}(D_{[v]} B_{[v]})}$, die dritte Identität wegen $D_{[v]} B_{[v]} \in \mathcal{N}(D_{[v]} B_{[v]})$, die vierte Identität wegen

$D_v^{-2} = C_v^t P_{\mathcal{R}(Z)} C_v + B_v B_v^t$ mit $P_{\mathcal{R}(Z)} = ZZ^+$ und die fünfte Identität wegen $P_{\mathcal{R}(Z)} = P_{\mathcal{R}(Z)} P_{\mathcal{R}(Z)}$ und $P_{\mathcal{R}(Z)} P_{\mathcal{N}(Z)} = 0$ folgt. Insbesondere folgt

$$\begin{aligned}
\bar{S}_{[v]} &= \bar{S}_* \\
&\Leftrightarrow X^t P_{\mathcal{N}(Z)} C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X = 0 \\
&\Leftrightarrow X^t P_{\mathcal{N}(Z)} C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{R}(Z)} P_{\mathcal{R}(Z)} C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X = 0 \\
&\Leftrightarrow P_{\mathcal{R}(Z)} C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X = 0 \\
&\Leftrightarrow Z^+ C_{[v]} D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]} C_{[v]}^t P_{\mathcal{N}(Z)} X = 0 \\
&\Leftrightarrow \bar{A}_{[v]} = \bar{A}_*,
\end{aligned}$$

dabei gilt die erste und letzte Äquivalenz wegen Lemma 6.25 sowie die zweite Äquivalenz in Analogie zu der vorherigen Betrachtung. Die Matrix in der dritten Zeile ist positiv semidefinit, so dass die dritte Äquivalenz folgt. Die vierte Äquivalenz gilt nun mit $\text{Rang}(Z) = k$. \square

Lemma 6.27 *Es sei Z eine $n \times k$ Matrix mit $r[Z] = k \leq n$ und M eine positiv definite $k \times k$ Matrix. Dann gilt für alle $r \in \mathbb{N}$ mit $Z_r = \mathbb{1}_r \otimes Z$*

$$|I_{rn} + Z_r M Z_r| = |I_n + r Z M Z^t|.$$

BEWEIS : Der Beweis erfolgt mittels vollständiger Induktion nach r , wobei

$$\begin{aligned}
&|I_{(r+1)n} + Z_{r+1} M Z_{r+1}^t| \\
&= |I_{rn} + Z_r M Z_r^t| |I_n + Z M Z^t| - Z M Z_r^t (I_{rn} + Z_r M Z_r^t)^{-1} Z_r M Z^t| \\
&= |I_{rn} + Z_r M Z_r^t| |I_n + Z(M^{-1} + r Z^t Z)^{-1} Z^t| \\
&= |I_n + r Z M Z^t| |I_n + Z(M^{-1} + r Z^t Z)^{-1} Z^t| \\
&= |(I_n + Z(M^{-1} + r Z^t Z)^{-1} Z^t)(I_n + r Z M Z^t)| \\
&= |I_n + r Z M Z^t + Z M M^{-1} (M^{-1} + r Z^t Z)^{-1} Z^t + Z M r Z^t Z (M^{-1} + r Z^t Z)^{-1} Z^t| \\
&= |I_n + r Z M Z^t + Z M Z^t| = |I_n + (r+1) Z M Z^t|.
\end{aligned}$$

der Induktionsschritt von r nach $r+1$ ist. \square

Lemma 6.28 *Es sei $[v] \in [D]$ beliebig fixiert und für alle $r \in \mathbb{N}$ sei v_r ein Element aus der Äquivalenzklasse mit dem Repräsentanten $[v]$. Weiterhin seien (vgl. Annahme A.6.2)*

$$\Omega_{*r} = Z_r M_0 Z_r^t + I_{rn} \quad \text{und} \quad \Omega_{v_r} = Z_r M_0 Z_r^t + K_{v_r}.$$

Dann existieren konvergente positive Folgen $(h_{*r})_{r \in \mathbb{N}}$, $(h_{[v]r})_{r \in \mathbb{N}}$ mit den positiven Grenzwerten h_* bzw. $h_{[v]}$ und

$$|\Omega_{*r}| = h_{*r} r^k \quad \text{sowie} \quad |\Omega_{v_r}| = h_{[v]r} r^{k-(n_{[v]}-q_{[v]})}.$$

BEWEIS : Es seien $\lambda_1, \dots, \lambda_k$ die positiven Eigenwerte der positiv semidefiniten Matrix $Z M_0 Z^t$. Dann folgt der erste Teil der Behauptung wegen Lemma 6.27

$$|\Omega_{*r}| = |I_n + r Z M_0 Z^t| = \prod_{i=1}^k (1 + r \lambda_i) = r^k \prod_{i=1}^k \left(\frac{1}{r} + \lambda_i\right).$$

Ferner gilt für alle $r \geq k$ mit $Z_l = \mathbb{1}_l \otimes Z$ für alle $l \in \mathbb{N}$ (vgl. (6.16))

$$|\Omega_{v_r}| = \left| \begin{pmatrix} Z_k M_0 Z_k^t + K_{v_k} & Z_k M_0 Z_{r-k}^t \\ Z_{r-k} M_0 Z_k^t & Z_{r-k} M_0 Z_{r-k}^t + I_{(r-k)n} \end{pmatrix} \right|.$$

Insbesondere folgt dann

$$|\Omega_{v_r}| = |I_{(r-k)n} + Z_{r-k} M_0 Z_{r-k}^t| \\ |K_{v_k} + Z_k M_0 Z_k^t - Z_k M_0 Z_{r-k}^t (I_{(r-k)n} + Z_{r-k} M_0 Z_{r-k}^t)^{-1} Z_{r-k} M_0 Z_k^t|.$$

Es sei nun $M_r = M_0^{-1} + (r-k)Z^t Z$ mit $\lim_{r \rightarrow \infty} r M_r^{-1} = (Z^t Z)^{-1}$, dann folgt

$$|K_{v_k} + Z_k M_r^{-1} Z_k^t| = |K_k + Z_k M_0 Z_k^t - Z_k M_0 Z_{r-k}^t (I_{(r-k)n} + Z_{r-k} M_0 Z_{r-k}^t)^{-1} Z_{r-k} M_0 Z_k^t|$$

und somit $|\Omega_{v_r}| = |\Omega_{*_{r-k}}| |K_{v_k} + Z_k M_r^{-1} Z_k^t|$. Weiterhin gilt

$$|K_{v_k} + Z_k M_r^{-1} Z_k^t| = \left| \begin{pmatrix} C_{v_k}^t Z_k M_r^{-1} Z_k^t C_{v_k} + B_{v_k} B_{v_k}^t & C_{v_k}^t Z_k M_r^{-1} Z_k^t \bar{C}_{v_k} \\ \bar{C}_{v_k}^t Z_k M_r^{-1} Z_k^t C_{v_k} & \bar{C}_{v_k}^t Z_k M_r^{-1} Z_k^t \bar{C}_{v_k} + I_{n_k - n_{[v]}} \end{pmatrix} \right|,$$

wobei $\bar{C}_{v_k}^t \bar{C}_{v_k} = I_{k n} - C_{v_k} C_{v_k}^t$ und $\bar{C}_{v_k}^t \bar{C}_{v_k} = I_{k n - n_{[v]}}$ ist. Ferner seien nun

$$Q_{[v]r} = C_{[v]}^t Z_r M_r^{-1} Z^t C_{[v]} + B_{[v]} B_{[v]}^t, \\ H_{[v]r} = C_{[v]}^t Z M_r^{-1} Z^t C_{[v]} + B_{[v]} B_{[v]}^t = r^{-1} (Q_{[v]r} + (r-1) B_{[v]} B_{[v]}^t), \\ G_{[v]r} = I_{k n - q_{[v]}} + \bar{C}_{v_k}^t Z_k M_r^{-1} Z_k^t \bar{C}_{v_k} - (\bar{C}_{v_k}^t Z_k M_r^{-1} Z_k^t C_{v_k}) H_{[v]r}^{-1} (C_{v_k}^t Z_k M_r^{-1} Z_k^t \bar{C}_{v_k}).$$

Dann folgt

$$|\Omega_{v_r}| = |\Omega_{*_{r-k}}| |H_{[v]r}| |G_{[v]r}|,$$

wobei in Analogie zum Beweis des Lemmas 6.26 die Matrizen $Q_{[v]r}$ für alle $r \in \mathbb{N}$ positiv definit sind und $\lim_{r \rightarrow \infty} r^{-1} Q_{[v]r}^{-1} = Q_{[v]}^{-1}$ sowie $\lim_{r \rightarrow \infty} r^{-1} H_{[v]r}^{-1} = D_{[v]} P_{\mathcal{N}(D_{[v]} B_{[v]})} D_{[v]}$ mit $D_{[v]} = Q_{[v]}^{-\frac{1}{2}}$ gilt. Weiterhin folgt damit $\lim_{r \rightarrow \infty} G_{[v]r} = I_{k n - n_{[v]}}$. Es seien nun $\lambda_1^r \dots \lambda_{q_{[v]}}^r$ die positiven Eigenwerte der positiv semidefiniten Matrix $Q_{[v]r}^{-\frac{1}{2}} B_{[v]} B_{[v]}^t Q_{[v]r}^{-\frac{1}{2}}$. Dann folgt

$$|H_{[v]r}| = r^{-n_{[v]}} |Q_{[v]r}| |I_p + (r-1) Q_{[v]r}^{-\frac{1}{2}} B_{[v]} B_{[v]}^t Q_{[v]r}^{-\frac{1}{2}}|, \\ = r^{q_{[v]} - n_{[v]}} |Q_{[v]r}| \prod_{i=1}^{q_{[v]}} \left(\frac{1}{r} + \frac{r-1}{r} \lambda_i^r \right).$$

Zusammen mit $\lim_{r \rightarrow \infty} G_{[v]r} = I_{k n - n_{[v]}}$ und $\Omega_{*_{r-k}} = r^k \prod_{i=1}^k \left(\frac{1}{r} + \frac{r-k}{r} \lambda_i \right)$ für die positiven Eigenwerte $\lambda_1, \dots, \lambda_k$ der positiv semidefiniten Matrix $Z M_0 Z^t$, folgt nun wegen

$$|\Omega_{v_r}| = r^{k+q_{[v]}-n_{[v]}} \prod_{i=1}^k \left(\frac{1}{r} + \frac{r-k}{r} \lambda_i \right) \prod_{i=1}^{q_{[v]}} \left(\frac{1}{r} + \frac{r-1}{r} \lambda_i^r \right) |Q_{[v]r}| |G_{[v]r}|$$

der zweite Teil der Behauptung. \square

6.5 Semiparametrisches Lokationsmodell

Es seien $\mathbf{X}_1, \dots, \mathbf{X}_n$ unabhängige Zufallsvariablen mit Werten in \mathbb{R}^p und identischer unbekannter Randverteilung P über $(\mathbb{R}^p, \mathfrak{L}^p)$. Des weiteren nehmen wir an, gelte

$$\mathbb{E} \mathbf{X}_i = \mu, \quad \text{Cov} \mathbf{X}_i = \Sigma, \quad i = 1, \dots, n.$$

Der Vektor $\mu \in \mathbb{R}^p$ sowie die $p \times p$ Matrix Σ sind unbekannt. Insbesondere ist dies nun der Spezialfall $Z = \mathbb{1}_n$ des multivariaten semiparametrischen linearen Modell (vgl. Abschnitt 6.1).

Satz 6.29 *Es sei eine Beobachtung $\mathbf{X} = x$ mit $\mathbf{X} \sim P_\theta^n$ und $\theta \in \Theta$ gegeben. Die eindeutig festgelegte verallgemeinerte Maximum-Likelihood-Schätzung $\hat{\theta}(x) = (\hat{\mu}(x), \hat{G}(x))$ nach Kiefer und Wolfowitz ist gegeben durch*

$$\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad \hat{G}(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i - \hat{\mu}(x)} \quad \text{für } x^t = (x_1^t, \dots, x_n^t).$$

BEWEIS : $\hat{\mu}(x)$ ist die einzige Lösung der Gleichung $\bar{x} = \bar{Z}\mu$ für $\mu \in \mathbb{R}^p$ und $Z = \mathbb{1}_n$ (vgl. Bedingung 1 im Satz 6.3). Somit folgt mit Satz 6.3 die Behauptung. \square

Satz 6.30 *Es sei eine Beobachtung $\mathbf{X} = x$ mit $\mathbf{X} \sim P_\theta^n$ und $\theta \in \Theta$ gegeben. Die eindeutig festgelegte verallgemeinerte Maximum-Likelihood-Schätzung $\hat{\theta}(x) = (\hat{\mu}(x), \hat{G}(x))$ nach Gill ist gegeben durch*

$$\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad \hat{G}(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i - \hat{\mu}(x)} \quad \text{für } x^t = (x_1^t, \dots, x_n^t).$$

BEWEIS : $\hat{\mu}(x)$ ist die einzige Lösung der Gleichung $\bar{x} = \bar{Z}\mu$ für $\mu \in \mathbb{R}^p$ und $Z = \mathbb{1}_n$ (vgl. Bedingung 1 im Satz 6.7). Somit folgt mit Satz 6.7 die Behauptung. \square

Bemerkung 6.31 In dem vorgestellten Lokationsmodell ist die gemeinsame Verteilung P_θ^n des Vektors $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ das Produkt identischer Randverteilungen P_θ . Wählen wir hier als Parametrisierung, dass die Randverteilung P einer Zufallsvariable \mathbf{X}_i unbekannt ist und in der Menge aller Wahrscheinlichkeitsverteilungen \mathcal{P}_+ mit endlichem zweiten Moment variiert, so beschreiben wir offensichtlich auch das vorgestellte Lokationsmodell. Insbesondere ist nun die Menge \mathcal{P}_+ eine Teilmenge der Menge \mathcal{P} aller Wahrscheinlichkeitsverteilungen. Für jede Beobachtung $x^n = (x_1^t, \dots, x_n^t)^t \in \mathbb{R}^{n \times p}$ ist die eindeutig bestimmte vMLS in der Menge \mathcal{P} nach Kiefer und Wolfowitz sowie Gill die empirische Verteilung (vgl. Kapitel 5)

$$\hat{F}_{x^n} = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}.$$

Damit ist \hat{F}_{x^n} auch in \mathcal{P}_+ die eindeutig festgelegte vMLS nach Kiefer und Wolfowitz oder Gill, da \hat{F}_{x^n} auch in der Menge \mathcal{P}_+ liegt. Weiterhin sind dann (μ, Σ) abgeleitete Parameter $(\mu, \Sigma) = \varphi(P)$ und die vMLS $(\hat{\mu}, \hat{\Sigma}) = \varphi(\hat{F}_{x^n})$ im Sinne von Kiefer und Wolfowitz sowie Gill entspricht dann den bekannten Kleinste Quadrate Lösungen. \square

Satz 6.32 *Es sei $X \in \mathbb{R}^{n \times p}$ mit $|X^t P_{\mathcal{N}(\mathbf{1}_n)} X| > 0$ gegeben. Dann sind die selbstinformativen Grenzwerte $\bar{\mu}$, $\bar{\Sigma}$, \bar{P}_i der Bayesschen Schätzungen der Parameter μ , Σ und der Verteilung P_i für $i = 1, \dots, n$:*

$$\bar{\mu} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{\Sigma} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{\mu})^t (X_i - \bar{\mu}), \quad \bar{P} = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

BEWEIS : Die Behauptung folgt direkt aus dem Satz 6.23 und entspricht dem Ergebnis von [Bunke \(2002\)](#). □

Kapitel 7

Ausblick

Im Mittelpunkt dieser Arbeit steht der selbstinformativ Grenzwert als eine Möglichkeit der Konstruktion sensibler Schätzungen auf der Grundlage Bayesscher Schätzungen in Situationen ohne Vorliegen einer Vorinformation sowie als eine Verallgemeinerung des Maximum-Likelihood-Prinzips. Wir haben gezeigt, dass unter Regularitätsbedingungen in einer Vielzahl gewöhnlich betrachteter Modelle für diskrete oder absolut stetige Zufallsvariablen mit stetiger Dichte im unbekanntem Parameter der selbstinformativ Grenzwert unabhängig von der Wahl der a priori Verteilung ist und einer MLS, falls diese existiert, entspricht. Des Weiteren konnten wir erste Einblicke in das Verhalten des selbstinformativ Grenzwertes in Modellen mit nicht dominierter Verteilungsfamilie gewinnen. Jedoch ist bisher eine Allgemeinheit der Aussagen in Analogie zu dominierten Verteilungsfamilien sowohl in der Frage der Abhängigkeit von der a priori Verteilung als auch des Zusammenhanges zu bekannten verallgemeinerten Maximum-Likelihood-Schätzungen nicht erreicht. So basieren die vorgestellten selbstinformativ Grenzwerte auf der a priori Annahme eines Dirichlet Prozesses als zufälliges Wahrscheinlichkeitsmaß, wobei die Abhängigkeit von dieser a priori Annahme z.B. in dem vorgestellten semiparametrischen linearen Modell nicht geklärt ist. Ebenso ist die Existenz eines nicht trivialen selbstinformativ a posteriori Trägers für die Folge der a posteriori Verteilungen unter der a priori Annahme eines Dirichlet Prozesses ungeklärt. Alternativ zum Dirichlet Prozess erscheint die Betrachtung einer a priori Gewichtung auf dem messbaren Raum $(\mathcal{P}, \mathfrak{B}_{\mathcal{P}})$ vielversprechend (vgl. Bemerkung 5.5). Da unter der Annahme, dass \mathcal{X} polnisch ist, auch der Raum \mathcal{P} versehen mit dem Prohorovabstand $d_{\mathcal{P}}$ polnisch ist. Insbesondere ist dann die Existenz einer a posteriori Verteilung auf $\mathfrak{B}_{\mathcal{P}}$ gesichert. Weiterhin basieren die betrachteten Modelle auf der Annahme, dass die zugrunde liegenden Zufallsvariablen unabhängig sind. Dies ist aber in vielen Situationen nicht gegeben. Häufig ist davon auszugehen, dass eine Abhängigkeit zwischen den einzelnen Zufallsvariablen vorliegt, was zum Beispiel mit Hilfe der Zeitreihentheorie beschrieben werden kann. Es stellt sich natürlich die Frage, in wie weit die Methode des selbstinformativ Grenzwertes als auch die vorgestellten verallgemeinerten Maximum-Likelihood-Methoden übertragen werden können.

Literaturverzeichnis

- ANDERSON, T. W. (1984). *An introduction to multivariate statistical analysis*. Probability and Mathematical Statistics. John Wiley & Sons, New York, 2. Auflage.
- ANTONIAK, C. E. (1974). „Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems.“, *Ann. of Statist.*, 2, 1152–1174.
- ASH, R. B. (1972). *Measure, integration, and functional analysis*. Academic Press, New York.
- BAUER, H. (1991). *Wahrscheinlichkeitstheorie*. De Gruyter Lehrbuch. Walter de Gruyter, Berlin, 4. Auflage.
- BELLACH, J., P. FRANKEN, E. WARMUTH UND W. WARMUTH (1978). *Maß, Integral und bedingter Erwartungswert*. Wissenschaftliche Taschenbücher. Akademie-Verlag, Berlin.
- BERNARDO, J. M. UND A. F. M. SMITH (1994). *Bayesian theory*. Wiley Series in Probability and Statistics. Wiley, Chichester.
- BILLINGSLEY, P. (1968). *Convergence of probability measures*. John Wiley & Sons, New York.
- BLACKWELL, D. UND J. B. MACQUEEN (1973). „Ferguson distributions via Polya urn schemes.“, *Ann. of Statist.*, 1, 353–355.
- BOX, G. E. P. UND G. C. TIAO (1992). *Bayesian inference in statistical analysis*. Wiley Classics Library. Wiley, New York.
- BUNKE, H., C. HENNIG UND W. H. SCHMIDT (1976). „Weighted combination of prior and sample information and parameter estimation.“, *Math. Operationsforsch. Statistik*, 7, 665–672.
- BUNKE, O. (2002). „Bayes estimates in multivariate semiparametric linear models.“, Discussion paper 58, Sonderforschungsbereich 373, Humboldt Universität, Berlin.
- BUNKE, O. UND J. JOHANNES (2002). „Selfinformative limits of Bayes estimates and generalized maximum likelihood.“, Manuskript, Sonderforschungsbereich 373, Humboldt Universität, Berlin.
- DA PRATO, G. UND J. ZABCZYK (1992). *Stochastic equations in infinite dimensions*. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge.
- DEGROOT, M. H. (1970). *Optimal statistical decisions*. McGraw-Hill Series in Probability and Statistics. McGraw-Hill Book Company, New York.
- DOKSUM, K. (1974). „Tailfree and neutral random probabilities and their posterior distributions.“, *Ann. of Probab.*, 2, 183–201.

- DUBINS, L. E. UND D. A. FREEDMAN (1963). „Random distribution functions.“, *Bull. Am. Math. Soc.*, 69, 548–551.
- DUBINS, L. E. UND D. A. FREEDMAN (1967). „Random distribution functions.“, in *Proc. 5th Berkeley Symp. Math. Stat. Probab., Univ. Calif. 1965/66*, S. 183–214.
- DUDLEY, R. M. (1989). *Real analysis and probability*. Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA.
- FABIUS, J. (1973a). „Neutrality and Dirichlet distributions.“, in *Inform. Theory, statist. Decision Funct., Random Processes; Transact. 6th Prague Conf. 1971*, S. 175–181.
- FABIUS, J. (1973b). „Two characterizations of the Dirichlet distribution.“, *Ann. of Statist.*, 1, 583–587.
- FERGUSON, T. S. (1973). „A Bayesian analysis of some nonparametric problems.“, *Ann. of Statist.*, 1, 209–230.
- GEMAN, S. UND C.-R. HWANG (1982). „Nonparametric maximum likelihood estimation by the method of sieves.“, *Ann. of Statist.*, 10, 401–414.
- GHORAI, J. UND H. RUBIN (1982). „Bayes risk consistency of nonparametric Bayes density estimates.“, *Aust. J. Stat.*, 24, 51–66.
- GILL, R. D. (1989). „Non- and semi-parametric maximum likelihood estimators and the von Mises method. I.“, *Scand. J. Stat.*, 16, 97–128.
- GRENANDER, U. (1981). *Abstract inference*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.
- HAHN, H. UND A. ROSENTHAL (1948). *Set functions*. The University of New Mexico Press, Albuquerque, New Mexico.
- HARTIGAN, J. (1983). *Bayes theory*. Springer Series in Statistics. Springer-Verlag, New York.
- HUMAK, K. (1977). *Statistische Methoden der Modellbildung Band I. Statistische Inferenz für lineare Parameter*. Mathematische Lehrbücher und Monographien, II Abt. Akademie-Verlag, Berlin.
- IBRAGIMOV, I. UND R. KHAS’MINSKIJ (1981). *Statistical estimation. Asymptotic theory. Transl. from the Russian by Samuel Kotz*. Applications of Mathematics. Springer-Verlag, New York.
- KIEFER, J. UND J. WOLFOWITZ (1956). „Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters.“, *Ann. Math. Stat.*, 27, 887–906.
- KOLMOGOROFF, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Nr. 3 in Erg. d. Math. u. ihrer Grenzgebiete. Julius Springer, Berlin.
- KORWAR, R. M. UND M. HOLLANDER (1973). „Contributions to the theory of Dirichlet processes.“, *Ann. of Probab.*, 1, 705–711.
- KRAFT, C. (1964). „A class of distribution function processes with have derivatives“, *J. Appl. Probab.*, 1, 385–388.
- KRAFT, C. UND C. VAN EEDEN (1964). „Bayesian bio-assay“, *Ann. Math. Stat.*, 35, 886–890.
- LO, A. Y. (1984). „On a class of Bayesian nonparametric estimates: I. Density estimates.“, *Ann. of Statist.*, 12, 351–357.

- MÜLLER, P. H. , (ed.) (1991). *Wahrscheinlichkeitsrechnung und mathematische Statistik: Lexikon der Stochastik*. Akademie-Verlag, Berlin, 5. Auflage.
- OWEN, A. B. (2001). *Empirical likelihood*. Monographs on Statistics and Applied Probability. Chapman & Hall/ CRC, Boca Raton, FL.
- PFANZAGL, J. (1969). „On the measurability and consistency of minimum contrast estimates“, *Metrika*, 14, 249–272.
- PFANZAGL, J. (1979). „Conditional distributions as derivatives.“, *Ann. of Probab.*, S. 1046–1049.
- PROHOROV, Y. V. (1956). „Convergence of random processes and limit theorems in probability theory.“, *Theor. Probability Appls.*, 1, 157–214.
- ROLIN, J.-M. (1992). „Some useful properties of the dirichlet process“, Discussion paper 9202, Institut de statistique, Universite Catholique de Louvain.
- SCHMETTERER, L. (1966). *Einführung in die mathematische Statistik*. Springer-Verlag, Wien, 2. Auflage.
- SCHOLZ, F. (1980). „Towards a unified definition of maximum likelihood.“, *Can. J. Stat.*, 8, 193–203.
- SIMAR, L. (1984). „A survey of Bayesian approaches to nonparametric statistics.“, *Math. Operationsforsch. Stat., Ser. Stat.*, 15, 121–142.
- STRASSER, H. (1985). *Mathematical theory of statistics. Statistical experiments and asymptotic decision theory*. De Gruyter Studies in Mathematics. Walter de Gruyter, Berlin.
- WITTING, H. (1985). *Mathematische Statistik I: Parametrische Verfahren bei festem Stichprobenumfang*. B. G. Teubner, Stuttgart.
- ZELLNER, A. (1971). *An introduction to Bayesian inference in econometrics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.

Lebenslauf

Zur Person

Jan Johannes
geboren am 23. September 1973 in Berlin
ledig
wohnhaft in Lincolnstraße 57, 10315 Berlin

Derzeitige Tätigkeit

1999 - dato Wissenschaftlicher Mitarbeiter, Institut für Mathematik,
Humboldt-Universität zu Berlin

Wissenschaftliche Ausbildung

1993 - 1999 Mathematikstudium an der Freien Universität Berlin
Spezialisierung : Mathematische Statistik
Nebenfach : Informatik
April 1995 Vordiplom Mathematik
November 1999 Abschluss : Diplom-Mathematiker
Titel der Diplomarbeit : „Selbstinformative“ Grenzwerte
von Bayesschen Schätzungen in multivariaten Modellen
1995-1999 Betriebswirtschaftsstudium an der Freien Universität Berlin
Oktober 1997 Vordiplom Betriebswirtschaftslehre

Schulische Ausbildung

1990 - 1992 Andreas Gymnasium Berlin-Friedrichshain
Abschluss : Abitur
1980 - 1990 Zille Schule Berlin-Friedrichshain

Berlin, 06. September 2002

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfasst zu haben und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Jan Johannes
06. September 2002