

Asymptotik von Regressionsmodellen

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik der
Universität Karlsruhe
genehmigte

DISSERTATION

von

Dipl.-Math. oec. Guido Bader
aus Offenburg

Tag der mündlichen Prüfung:

18.07.2001

Referent:

Prof. Dr. W. Bischoff

Korreferenten:

Prof. Dr. N. Henze

Prof. Dr. H. Luschgy

Karlsruhe

2001

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Angestellter am Institut für Mathematische Stochastik der Universität Karlsruhe. Mein herzlicher Dank geht an all diejenigen, die mich bei ihrer Anfertigung unterstützt, ermutigt und beraten haben.

- Vor allen danke ich Herrn Prof. Dr. Wolfgang Bischoff für die wissenschaftliche Betreuung der vorliegenden Arbeit. Er hat mir das Thema aufgezeigt und in zahlreichen Gesprächen stets wertvolle Anregungen gegeben.
- Ferner danke ich den Herren Prof. Dr. Norbert Henze und Prof. Dr. Harald Luschgy für die freundliche Übernahme des Korreferats.
- Mein weiterer Dank geht an meinen Kollegen Dr. Nora Gürtler, Dr. Bernhard Klar und Dipl.-Math. Frank Miller für die vielen angeregten Diskussionen und Ratschläge, die diese Arbeit während ihrer Entstehung begleitet haben; Frau Dr. Gürtler danke ich überdies für geduldiges Korrekturlesen und Herrn Miller für seine Hilfe bei den Simulationen.
- Zudem danke ich allen Mitgliedern des Instituts für mathematische Stochastik der Universität Karlsruhe, in deren angenehmer und produktiver Mitte ich die vorliegende Arbeit anfertigen durfte.
- Zuletzt danke ich meiner Mutter von ganzem Herzen. Sie stand mir auf meinem gesamten bisherigen Lebensweg zur Seite und hat mich stets unterstützt und beraten.

Karlsruhe, den 18.07.2001

Inhaltsverzeichnis

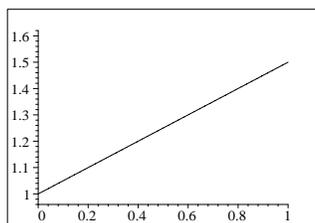
Einleitung	1
1 Grundlagen der LeCam-Theorie	9
1.1 Entscheidungstheorie - Bezeichnungen	9
1.2 Abstände von Experimenten	12
1.3 Lineare Modelle	15
1.4 Gaußexperimente	18
1.5 Ein Satz von Luschgy für Gaußexperimente mit Beobachtungsraum $C[0, 1]$. . .	22
1.6 Schwache Konvergenz von Experimenten	24
1.7 Schwache Konvergenz gegen einen Standard-Gaußshift auf einem Hilbertraum .	25
1.8 Die Hájek-LeCam Minimax-Schranke und ihre Anwendungen	27
1.8.1 Begriffe und Bezeichnungen beim asymptotischen Testen	27
1.8.2 Die Minimax-Schranke	29
1.8.3 Eine Anwendung der Minimax-Schranke für das asymptotische Testen .	30
1.9 Der Satz von Lindae	34
1.10 Asymptotische Äquivalenz	35
2 Asymptotik von Regressionsmodellen bei unabhängigen Beobachtungen	37
2.1 Modellbeschreibung	38
2.2 Konvergenzklassen bei gegebenen Designfolgen	41
2.3 Asymptotik mit dem Satz von Donsker	43
2.4 Lokale Theorie - Gaußsche Grenzwerte	45
2.4.1 Das Prinzip der Lokalisierung	46
2.4.2 Schwache Konvergenz Linearer Modelle mit Lokalisierungsrate $\frac{1}{\sqrt{n}}$. . .	52
2.4.3 Schwache Konvergenz allgemeiner Regressionsmodelle mit Lokalisierungsrate $\frac{1}{\sqrt{n}}$	55
2.4.4 Δ -Konvergenz Linearer Regressionsmodelle mit Lokalisierungsrate $\frac{1}{\sqrt{n}}$.	60
2.4.5 Δ -Konvergenz allgemeiner Regressionsmodelle mit Lokalisierungsrate $\frac{1}{\sqrt{n}}$	64
2.5 Globale Theorie - Asymptotische Äquivalenz unter Normalverteilungsannahme	69
2.5.1 Ein asymptotisches Äquivalent zum Linearen Regressionsmodell unter Normalverteilungsannahme	70
2.5.2 Ausblick	71

3	Anwendungen in der Testtheorie	73
3.1	Tests für das Lineare Regressionsmodell	74
3.1.1	Ein asymptotisch optimaler Test φ_n^*	74
3.1.2	Vergleich von φ_n^* mit dem Likelihood-Quotienten-Test bei normalverteiltem Fehler	75
3.1.3	Asymptotik des Likelihood-Quotienten-Tests bei unbekanntem Fehler	77
3.1.4	Vergleich von φ_n^* mit dem Likelihood-Quotienten-Test bei unbekannter Fehlerverteilung	80
3.2	Tests für das allgemeine Regressionsmodell	83
4	Anwendungen in der Designtheorie	87
4.1	Designtheorie bei fest vorgegebenem Stichprobenumfang	88
4.1.1	Problemstellung	88
4.1.2	Approximative Designtheorie für das Lineare Regressionsmodell - klassische Ansätze	89
4.1.3	Entscheidungstheoretische Ansätze in der Designtheorie	90
4.2	Asymptotische Designtheorie für Lineare Regressionsmodelle mit normalverteilten Fehlern	92
4.3	Lokale asymptotische Designtheorie für Regressionsmodelle	97
4.3.1	Lokalisierte Lineare Regressionsmodelle	97
4.3.2	Lokalisierte allgemeine Regressionsmodelle	102
4.4	Entscheidungstheoretische Interpretation der E -Optimalität	105
5	Lineare Regressionsmodelle mit unbekannter Fehlerverteilung	111
5.1	Modellbeschreibung	112
5.2	Adaptive Schätzverfahren	113
5.2.1	Effiziente und asymptotisch effiziente Schätzer	113
5.2.2	Adaptive Schätzer	117
5.2.3	Die Schätzer von Koul & Susarla	120
5.3	Adaptive Testverfahren	128
5.3.1	Definition und Konstruktion mittels adaptiver Schätzer	128
5.3.2	Tests mit den Schätzern von Koul & Susarla	131
5.4	Simulationen	133
A	Technische Hilfsmittel	137
A.1	Variationsabstand, Hellingerabstand, Affinität und Hellingertransformierte	137
A.2	Hilbertraumwertige Zufallsvariablen	140
A.3	Ein Transformationssatz	141
A.4	Sätze über Funktionen von beschränkter Variation	142
A.5	Stochastische Integrale nichtzufälliger Funktionen	144
A.6	Kernreproduzierende Hilberträume	146
A.7	Absolutstetige Funktionen	148
	Index	150
	Symbolverzeichnis	152
	Literaturverzeichnis	153

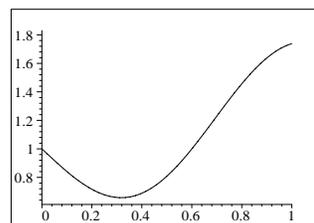
Einleitung

In der mathematischen Statistik werden Modelle entworfen, um damit Fragestellungen aus den verschiedensten Anwendungsbereichen wie der Biologie, der Ökonomie, der Medizin, der Technik u.v.m., zu beantworten. Die vorliegende Arbeit befasst sich mit den vielleicht wichtigsten dieser statistischen Modelle, den Regressionsmodellen. Zunächst wollen wir anhand eines Beispiels aus dem Bereich der Technik verdeutlichen, was wir unter einem Regressionsmodell verstehen.

Bei der maschinellen Produktion hängen Qualitätseigenschaften der produzierten Waren oftmals von der Geschwindigkeit (auch Takt genannt) ab, mit der die Maschine läuft. Der Takt sei stufenlos regelbar zwischen 0, d.h. die Maschine steht oder läuft mit minimaler Geschwindigkeit, und 1, d.h. die Maschine läuft mit Volllast, also maximaler Geschwindigkeit. Eine mögliche Qualitätseigenschaft ist die Präzision, mit der die Maschine arbeitet. Diese könnte z.B. das akkurate Aufkleben von Adressen auf Briefumschläge sein, das genaue Festlöten von Bestandteilen auf Mikrochips oder das exakte Ausführen von Bohrungen in Holz- oder Metallteile. Die Abweichung vom „Idealzustand“ lässt sich beobachten. Wir gehen davon aus, dass wir diesen „Produktionsfehler“ eindimensional messen. (Bei größeren Objekten kann man z.B. auch zwischen vertikaler und horizontaler Abweichung unterscheiden. Dies entspräche dann einem zweidimensionalen Produktionsfehler.) Zwischen dem Takt t und dem mittleren Produktionsfehler $f(t)$ bestehe ein funktionaler Zusammenhang $t \mapsto f(t)$, der z.B. eine der in Abbildung 1 angegebenen Gestalten haben könnte.



Linearer Zusammenhang



Nicht linearer Zusammenhang

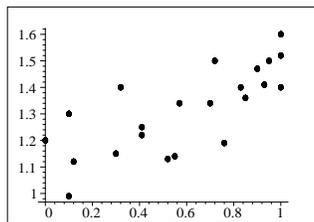
Abbildung 1: Funktionaler Zusammenhang zw. Takt und mittlerem Produktionsfehler

Der funktionale Zusammenhang $t \mapsto f(t)$ zwischen Takt und mittlerem Produktionsfehler ist *unbekannt*, denn wollte man den Zusammenhang berechnen, so müsste man den mittleren Produktionsfehler für jeden Takt $t \in [0, 1]$ bestimmen. Das ist nicht möglich. Darüber hinaus ist es selbst für einen festen Takt t nicht möglich, den Wert $f(t)$ exakt anzugeben, denn bei festem Takt ist nicht jedes produzierte Teil identisch. Die gemessene Abweichung eines Produktes vom Idealzustand hängt von einer *zufälligen Komponente* ab. Die beschriebene Problematik lässt sich wie folgt mathematisch modellieren: Bezeichnet Y_t die tatsächliche Abweichung eines produzierten Gutes bei Takt t , so gilt

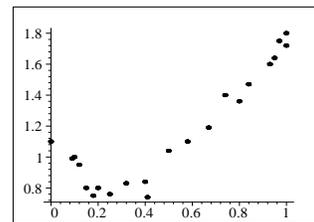
$$Y_t = f(t) + \epsilon_t,$$

wobei ϵ_t eine Zufallsvariable ist, die die zufällige Differenz zwischen tatsächlichem Produktionsfehler Y_t und mittlerem Produktionsfehler $f(t)$ beschreibt. Diese Zufallsgröße kann auch bewirken, dass der tatsächliche Produktionsfehler Y_t kleiner ist, als im Mittel zu erwarten wäre.

Auch wenn sich der funktionale Zusammenhang nicht berechnen lässt, so möchte man sich doch ein gutes Bild von der Funktion f machen. Dazu wird man für verschiedene Takte $0 \leq t_1 \leq \dots \leq t_n$ jeweils ein Gut produzieren und den Produktionsfehler messen. Insgesamt werden somit n „Versuchsprodukte“ bei nicht notwendig verschiedenen Takten t_i hergestellt. Entspricht der wahre funktionale Zusammenhang einer der Funktionen aus Abbildung 1, so könnte das Ergebnis eines solchen Versuches das in Abbildung 2 angegebene Aussehen haben.



Linearer Zusammenhang



Nicht linearer Zusammenhang

Abbildung 2: Beobachtungen vom Umfang $n = 22$

Mathematisch lässt sich ein solcher *Versuch* wie folgt beschreiben: Zu den *Versuchspunkten* t_1, \dots, t_n gehören die Beobachtungen Y_i mit

$$Y_i = f(t_i) + \epsilon_i, \quad i = 1, \dots, n.$$

Wir sprechen hierbei von einem **Regressionsmodell**, und um dieses untersuchen zu können, müssen Voraussetzungen an das mögliche Aussehen der Funktion f sowie an die Fehler ϵ_i gemacht werden.

Der *unbekannten Regressionsfunktion* f unterstellt man üblicherweise, dass sie aus einer *vorgegebenen Klasse* \mathcal{F} von Funktionen stammt. Nimmt man an, dass f eine *Linearkombination* bekannter Funktionen f_1, \dots, f_k ist, also $\mathcal{F} = [f_1, \dots, f_k]$, so sprechen wir von einem **Linearen Regressionsmodell**. Wählt man als \mathcal{F} eine große, „unendlichdimensionale“ Klasse, z.B. die der stetigen Funktionen oder die der Funktionen von beschränkter Variation auf $[0, 1]$ bzw. unendlichdimensionale Teilmengen dieser Funktionenklassen, so sprechen wir von einem **allgemeinen** oder auch „**nichtparametrischen**“ **Regressionsmodell**. Die Bezeichnung „nichtparametrisch“ ist insofern etwas unglücklich, als man von f stets als dem *unbekannten Parameter*, an dem man interessiert ist, und von \mathcal{F} als *Parameterraum* spricht. Wir beschränken uns deshalb in der vorliegenden Arbeit auf die Bezeichnung „allgemeines Regressionsmodell“. Lineare Regressionsmodelle sind somit Spezialfälle der allgemeinen Regressionsmodelle.

Die *Fehler* ϵ_i sind üblicherweise *stochastisch unabhängig* und besitzen *dieselbe Verteilung*. Diese Annahme wird auch in der vorliegenden Arbeit getroffen. Untersuchungen bei Abhängigkeiten der Fehler oder bei nicht identisch verteilten Fehlern würden den Rahmen dieser Arbeit sprengen und bleiben Gegenstand weiterer Forschung. Entscheidend ist jedoch die Frage, ob die Fehlerverteilung *bekannt oder unbekannt* ist. Beide Situationen werden hier untersucht.

Wir werden obige Regressionsmodelle nun noch in zweierlei Hinsicht verallgemeinern. Zum einen lösen wir uns doch ein Stück weit von der Voraussetzung identisch verteilter Fehler. Dazu führen wir eine *bekannte Funktion* $\sigma(t) > 0$, $t \in [0, 1]$, ein und betrachten an Stelle der unabhängigen und identisch verteilten ϵ_i jetzt Fehler der Form $\sigma(t_i)\epsilon_i$. Damit bringen wir die Abhängigkeit der Fehlervarianz - falls existent - vom Versuchspunkt zum Ausdruck. Ist σ konstant, so spricht man von einem *homoskedastischen Regressionsmodell*. In diesem Fall gelten viele Resultate der vorliegenden Arbeit auch dann noch, wenn diese Konstante unbekannt ist und in den jeweiligen Sätzen durch einen konsistenten Schätzer ersetzt wird. Ist σ nicht konstant, so spricht man von einem *heteroskedastischen Regressionsmodell*.

Zum anderen - zunächst nur eine schreibtechnische Veränderung - werden wir die *Versuchspunkte* mit t_{n1}, \dots, t_{nn} bezeichnen. Dies deutet bereits an, dass wir Folgen von Experimenten mit wachsenden Beobachtungsumfängen untersuchen werden.

Diese beiden Verallgemeinerungen führen uns zu den Modellen, die Gegenstand der vorliegenden Arbeit sind, nämlich

$$Y_{ni} = f(t_{ni}) + \sigma(t_{ni})\epsilon_{ni}, \quad i = 1, \dots, n, \quad (1)$$

mit unabhängigen und identisch verteilten Fehlern ϵ_{ni} , einer bekannten Funktion $\sigma > 0$ und einer unbekanntem Regressionsfunktion f , die aus einer fest vorgegebenen Funktionenklasse \mathcal{F} stammen soll. Die Versuchspunkte t_{ni} , $i = 1, \dots, n$, seien Elemente eines abgeschlossenen Intervalls.

Im Zusammenhang mit den Regressionsmodellen (1) ist man nun an den verschiedensten Fragestellungen bezüglich der unbekanntes Regressionsfunktion f interessiert:

1. Die Funktion f soll *geschätzt* werden.
2. Oftmals interessieren aber nur spezielle Eigenschaften der Funktion f . Nimmt man z.B. wie im ersten Schaubild von Abbildung 1 einen linearen Zusammenhang zwischen Takt und mittlerem Produktionsfehler an, so könnte man sich fragen, ob die Steigung der Geraden positiv ist. Dies führt zu *Testproblemen*, die oftmals eine größere Rolle als das Schätzen der Funktion spielen.
3. Im Rahmen der *Entscheidungstheorie* werden noch sehr viel allgemeinere Fragestellungen bezüglich der unbekanntes Regressionsfunktion f untersucht. Beispielsweise könnte man den Produktionsvorgang mit Kosten bewerten (z.B. für schadhafte Produkte, für den Umsatzausfall bei kleinem Takt, für Garantieleistungen bei schlechten Produkten, usw.) und sich nach einer „kostenoptimalen“ Taktfrequenz fragen.
4. Bei allen Fragestellungen hängt der Informationsgehalt der Regressionsmodelle (1) von den Versuchspunkten t_{n1}, \dots, t_{nn} ab. Interessiert man sich z.B. für die Steigung einer Geraden, so ist es offensichtlich sinnlos, alle Versuche in einem einzigen Punkt zu machen, also $t_{n1} = \dots = t_{nn}$ zu wählen. Wie sollen aber die Versuchspunkte gewählt werden (falls die Möglichkeit der Wahl besteht), um zu einer gegebenen Fragestellung die bestmögliche Information zu erhalten? Diese Problematik wird als *Versuchsplanungsproblem* bezeichnet und ist in vielen Anwendungsbereichen von großer Bedeutung, da die einzelnen Versuche oft sehr kostspielig sind (Kosten der Fehlermessung, Kosten schadhafter Produkte, usw.). Steht ein festes Budget zur Beantwortung einer Frage zur Verfügung, so soll ein Experiment durchgeführt werden, das möglichst viel Information liefert.

Zu den obigen Fragestellungen sucht man optimale Lösungen. Dies ist aber oft problematisch. Teils existieren keine optimalen Prozeduren, teils sind sie nicht oder nur sehr schwer berechenbar. Daher geht man zu *asymptotischen Betrachtungen* über, d.h. es werden die Regressionsmodelle (1) für wachsende Beobachtungsumfänge n untersucht.

Die Asymptotik der Regressionsmodelle (1) für wachsende Beobachtungsumfänge n wird in der vorliegenden Arbeit im Rahmen von **LeCam's Theorie der Experimente** untersucht. Zwar gibt es in der LeCam-Theorie bereits zahlreiche Arbeiten, die sich mit der Asymptotik von Linearen (Regressions-)Modellen beschäftigen, doch ist all diesen Arbeiten ein Manko gemeinsam: Sobald der eindimensionale Rahmen der Konstantenregression (das entspricht dem Einstichprobenproblem) verlassen wird, existieren Grenzexperimente nur noch unter sehr abstrakten Voraussetzungen (meist Noether-Bedingungen genannt, die die Existenz von „Grenzstrukturen“ fordern) und können nicht mehr explizit angegeben werden. Allen bisherigen Arbeiten fehlen also Anwendungsbeispiele. In

der vorliegenden Arbeit wird aber ein solches untersucht: Unsere praxisrelevanten Modelle (1) erfüllen nämlich unter geeigneten Voraussetzungen die abstrakten Bedingungen der schon bekannten Sätze und ermöglichen eine explizite Darstellung der Grenzexperimente, was weitreichende Anwendungen zulässt. In dieser Arbeit gehen wir auf die Anwendungen in der Test- und Designtheorie ein. Im Folgenden werden wir den Aufbau der Arbeit erläutern.

In *Kapitel 1* stellen wir die LeCam-Theorie vor, welche technische Grundlage für weite Teile dieser Arbeit ist. Die Darstellung orientiert sich an dem Buch von Strasser (1985). Auf Grund der oben genannten Fragestellungen haben wir den entscheidungstheoretischen Zugang zu LeCam's Δ -Metrik, die den „Informationsabstand“ zweier statistischer Experimente mit demselben Parameterraum misst, gewählt. Neben dem zur Δ -Metrik gehörigen starken Konvergenzbegriff spielt die schwache Konvergenz von Experimenten eine entscheidende Rolle, da wir in Kapitel 2 vor allem schwache Grenzexperimente für unsere Regressionsmodelle angeben. Im Mittelpunkt schwacher Konvergenz steht die Minimax-Schranke von Hájek-LeCam. Wir erläutern deren Auswirkung auf einfache, lineare Testprobleme, wie sie in Kapitel 3 betrachtet werden. Desweiteren untersuchen wir Gaußexperimente und speziell Gaußshifts auf Hilberträumen, die in Kapitel 2 als Grenzexperimente unserer Regressionsmodelle auftreten, und betrachten die Anwendung der LeCam-Theorie auf Lineare Modelle sowie auf Gaußexperimente mit Beobachtungsraum $C[0, 1]$. Dabei orientieren wir uns an dem Buch Torgersen (1991), sowie an den Arbeiten Luschgy (1991a) und Luschgy (1992).

Kapitel 2 ist das Zentrum der vorliegenden Arbeit. Wir betrachten hier die Asymptotik der oben beschriebenen Linearen und allgemeinen Regressionsmodelle im Rahmen der LeCam-Theorie. Als Grenzexperimente im Sinne LeCam's treten Gaußexperimente auf: Im Fall der Linearen Regressionsmodelle (mit Dimension k) sind das Gaußshifts auf dem \mathbb{R}^k , versehen mit einem Innenprodukt, das von der Fisher-Information der Fehlerverteilung sowie von den Regressoren und der Folge der Versuchspunkte abhängt. Im Fall der allgemeinen Regression erhalten wir als Grenzexperimente Gaußexperimente mit Beobachtungsraum $C[0, 1]$.

Wichtigster Bestandteil dieses Kapitels ist der Nachweis der schwachen Konvergenz unserer Regressionsmodelle. Aus den resultierenden Grenzwertsätzen schließen wir dann mit Hilfe des Satzes von Lindae unter restriktiveren Voraussetzungen auf Δ -Konvergenz. Diese stärkeren Grenzwertsätze spielen im weiteren Verlauf der Arbeit nur eine untergeordnete Rolle und können bei einem ersten Lesen übergangen werden.

Den Abschluss von Kapitel 2 bildet eine Aussage über starke asymptotische Äquivalenz für Lineare Regressionsmodelle unter Normalverteilungsannahme. Sie ist für die Designtheorie von besonderer Bedeutung und als *kleines* weiteres Mosaik im Rahmen der neuesten Forschungsergebnisse von Michael Nussbaum zu sehen.

Wir wollen kurz die Voraussetzungen der Grenzwertsätze aus Kapitel 2 andeuten: Grundvoraussetzung für die Existenz von Grenzexperimenten ist eine Lokalisierung im Parameterraum. Die passende Lokalisierungsrate ist in unserem Fall die klassische $1/\sqrt{n}$ -Rate. Die Fehlerverteilung wird in diesem Kapitel als bekannt vorausgesetzt und

muss eine Dichte bezüglich des Lebesgue-Maßes besitzen sowie eine endliche, positive Fisher-Information. Fasst man die Versuchspunkte einer Stufe n als eine Verteilungsfunktion H_n auf, so müssen diese H_n schwach oder gleichmäßig gegen eine weitere Verteilungsfunktion (wir sprechen von einem Grenzdesign) konvergieren. Als Klasse möglicher Regressoren \mathcal{F} kommen im Wesentlichen stetige Funktionen und vor allem Funktionen von beschränkter Variation in Betracht.

In *Kapitel 3* werden wir die schwachen Konvergenzaussagen aus Kapitel 2 auf einfache Testprobleme anwenden. Wir betrachten dabei stetige lineare Abbildungen $\ell : \mathcal{F} \rightarrow \mathbb{R}$ und Testprobleme der Form

$$H_\ell = \{f \in \mathcal{F} \mid \ell(f) \leq 0\}, \quad K_\ell = \{f \in \mathcal{F} \mid \ell(f) > 0\}.$$

Dies sind besonders einfache eindimensionale Testprobleme, die asymptotisch optimale Lösungen (im Sinne der Definitionen aus Kapitel 1) besitzen. Es ist allerdings nicht Ziel der Arbeit, Lösungen für viele verschiedene Testprobleme zu finden, sondern zu demonstrieren, welche weitreichenden Möglichkeiten die Konvergenzaussagen aus Kapitel 2 eröffnen. Eine Ausweitung der Ergebnisse von Kapitel 3 auf andere Testprobleme ist nicht schwer.

Zunächst werden Testprobleme für das Lineare Regressionsmodell untersucht, für die wir eine asymptotisch optimale Testfolge angeben. Diese Tests bauen stark auf der Dichte der Fehlerverteilung auf, so dass deren Kenntnis essentiell ist. Im Vergleich dazu betrachten wir Likelihood-Quotienten-Tests. Darunter verstehen wir Tests mit der Teststatistik des üblichen LQ-Tests unter Normalverteilungsannahme, die wir auch dann anwenden, wenn wir die Fehlerverteilung nicht kennen, sondern nur übliche Bedingungen an deren Momente voraussetzen. Wir untersuchen die Asymptotik dieser LQ-Tests und vergleichen sie mit der asymptotisch optimalen Testfolge bezüglich der Asymptotisch Relativen Effizienz.

Zum Abschluss von Kapitel 3 betrachten wir obige Testprobleme für das allgemeine Regressionsmodell und geben eine asymptotisch optimale Testfolge an.

In *Kapitel 4* analysieren wir die Auswirkungen der Grenzwertsätze aus Kapitel 2 auf die Designtheorie. Zunächst werden die Ansätze der klassischen Designtheorie für das Lineare Regressionsmodell bei festem Beobachtungsumfang n_0 skizziert und die Prinzipien der approximativen Designtheorie geschildert, anschließend gehen wir auf einen entscheidungstheoretischen Ansatz der Designtheorie ein: Im Kontext dieser Arbeit liegt es nahe, verschiedene Designs dadurch zu vergleichen, dass man die zugehörigen Experimente im Sinne der LeCam'schen Halbordnung vergleicht.

Ausgehend von diesem entscheidungstheoretischen Ansatz beschäftigen wir uns im Weiteren mit asymptotischer Designtheorie sowohl für Lineare als auch für allgemeine Regressionsmodelle und suchen asymptotisch optimale Designs, d.h. Designs die ab hinreichend großem Beobachtungsumfang „gut“ sind. Ausgangspunkt für diese Überlegungen sind die schwachen sowie starken Grenzwertsätze aus Kapitel 2. Zum Abschluss von Kapitel 4 geben wir entscheidungstheoretische Interpretationen der E -Optimalität an.

Wesentlich für alle Überlegungen von Kapitel 4 ist, dass sie unabhängig von der Fehlerverteilung sind. War diese in Kapitel 2 noch als bekannt vorausgesetzt und trat in den Grenzexperimenten in Form ihrer Fisher-Information auf, so spielt sie in der asymptotischen Designtheorie keine Rolle mehr.

In *Kapitel 5* beschäftigen wir uns ausschließlich mit Linearen Regressionsmodellen, bei denen im Gegensatz zu den Ausführungen der Kapitel 2 und 3 die Fehlerverteilung unbekannt ist. Ziel ist es, zu den in Kapitel 3 betrachteten Testproblemen eine asymptotisch optimale Testfolge zu konstruieren, die nicht auf der Kenntnis der Fehlerverteilung beruht. Um ein solches „adaptives“ Testverfahren angeben zu können, müssen wir jedoch die Klasse der möglichen Fehlerverteilungen einschränken. Üblicherweise beschränkt man sich auf diejenigen Fehlerverteilungen aus Kapitel 2, die eine zur y -Achse symmetrische Dichte besitzen.

Zu Beginn des Kapitels wiederholen wir (eingebettet in den Kontext der LeCam-Theorie) klassische Begriffe der Schätztheorie: effiziente Schätzer, asymptotisch effiziente Schätzer und adaptive Schätzer. Dann werden Schätzer aus der Arbeit Koul und Susarla (1983) zitiert, mit denen wir einen asymptotisch optimalen und für symmetrische Fehler adaptiven Test konstruieren. Außerdem betrachten wir ein anderes sich aus dieser Arbeit ergebende Testverfahren, das für alle Fehlerverteilungen aus Kapitel 2 „funktioniert“ und das wir mit der Folge der LQ-Tests aus Kapitel 3 vergleichen. Den Abschluss dieses Kapitels bilden einige Simulationen, die dem Leser einen Eindruck über das Verhalten der verschiedenen Testverfahren bei festen Stichprobenumfängen geben sollen.

In *Anhang A* werden noch einige Definitionen und Sätze angegeben, die demjenigen Leser, der nicht mit allen auftretenden Begriffen vertraut ist, das Lesen der Arbeit erleichtern sollen.

Kapitel 1

Grundlagen der LeCam-Theorie

In diesem Kapitel werden wir die Grundzüge der Theorie LeCam's präsentieren. Eine ausführliche Darstellung dieser Ergebnisse findet sich zum Beispiel in den Büchern LeCam (1986), LeCam und Yang (1990) oder Strasser (1985).

1.1 Entscheidungstheorie - Bezeichnungen

Sind $\Theta \neq \emptyset$ eine beliebige fest vorgegebene Menge, $(\mathcal{X}, \mathcal{A})$ ein Messraum und $\{P_\theta \mid \theta \in \Theta\}$ eine Menge von Wahrscheinlichkeitsmaßen auf $(\mathcal{X}, \mathcal{A})$, so nennen wir das Tripel

$$E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$$

ein **statistisches Experiment** mit **Beobachtungsraum** $(\mathcal{X}, \mathcal{A})$ und **Parameterraum** Θ . Mit $\mathcal{E}(\Theta)$ bezeichnen wir die Menge aller Experimente mit Parameterraum Θ . Ein **Entscheidungsraum** D ist ein topologischer Raum, versehen mit der Borelschen σ -Algebra $\mathcal{B}(D)$. Unter einer **randomisierten Entscheidungsfunktion** (kurz: **Entscheidung**) verstehen wir einen Markoff-Kern $\rho : \mathcal{X} \times \mathcal{B}(D) \rightarrow [0, 1]$. Mit $\mathcal{R}(E, D)$ bezeichnen wir die Menge aller randomisierten Entscheidungsfunktionen bei gegebenem Beobachtungsraum $(\mathcal{X}, \mathcal{A})$ und Entscheidungsraum $(D, \mathcal{B}(D))$. Eine **Verlustfunktion** ist eine Familie $W := (W_\theta)_{\theta \in \Theta}$ nach unten beschränkter $\mathcal{B}(D)$ -messbarer Funktionen $W_\theta : D \rightarrow \mathbb{R}$. Wenn wir im Folgenden von einer Eigenschaft von W sprechen, so bedeutet dies, dass alle Funktionen W_θ diese Eigenschaft besitzen. Das Tripel (E, D, W) nennen wir ein **(statistisches) Entscheidungsproblem**. Die Funktion

$$r_\rho : \Theta \rightarrow \mathbb{R}, \quad r_\rho(\theta) = \int_{\mathcal{X}} \int_D W_\theta(a) \rho(x, da) P_\theta(dx) \quad (1.1)$$

heißt die zu dem statistischen Entscheidungsproblem (E, D, W) und der Entscheidung ρ gehörige **Risikofunktion**. Eine Entscheidung $\rho_1 \in \mathcal{R}(E, D)$ heißt **nicht schlechter**

als eine andere Entscheidung $\rho_2 \in \mathcal{R}(E, D)$, falls die zu ρ_1 gehörige Risikofunktion r_{ρ_1} gleichmäßig kleiner oder gleich r_{ρ_2} ist, falls also gilt $r_{\rho_1}(\theta) \leq r_{\rho_2}(\theta)$ für alle $\theta \in \Theta$.

Zu jedem Testproblem gibt es ein korrespondierendes Entscheidungsproblem. Wir stellen kurz den Zusammenhang dar:

Seien $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ ein statistisches Experiment und (H, K) eine Partition des Parameterraums Θ . Wir wählen dann den zweielementigen Entscheidungsraum $D = \{H, K\}$, versehen mit der Potenzmenge $\mathcal{P}(D)$ als σ -Algebra. Jedem randomisierten Test $\varphi : \mathcal{X} \rightarrow [0, 1]$ mit Gütefunktion $g_\varphi : \Theta \rightarrow [0, 1]$, $g_\varphi(\theta) = \int \varphi dP_\theta$ ordnen wir eine Entscheidungsfunktion $\rho_\varphi : \mathcal{X} \times \mathcal{P}(D)$, $\rho_\varphi(x, K) := \varphi(x)$, $\rho_\varphi(x, H) := 1 - \varphi(x)$ zu. Desweiteren betrachten wir die Verlustfunktionen $W_\theta : D \rightarrow \mathbb{R}$, $W_\theta(H) = 1_K(\theta)$, $W_\theta(K) = 1_H(\theta)$. Da D endlich ist, ist $W = (W_\theta)_{\theta \in \Theta}$ eine stetige und beschränkte Verlustfunktion. Jetzt gilt für die Gütefunktion g_φ des Tests und die Risikofunktion der zugehörigen Entscheidungsfunktion der Zusammenhang

$$g_\varphi(\theta) = \begin{cases} r_{\rho_\varphi}(\theta), & \theta \in H, \\ 1 - r_{\rho_\varphi}(\theta), & \theta \in K. \end{cases}$$

Ein Test φ_1 heißt genau dann nicht schlechter als ein Test φ_2 , wenn die zugehörige Entscheidungsfunktion ρ_{φ_1} nicht schlechter ist als ρ_{φ_2} . Umgekehrt kann man jeder Entscheidungsfunktion ρ für obiges Entscheidungsproblem einen „zugehörigen“ Test φ_ρ zuordnen mittels $\varphi_\rho(x) = \rho(x, K)$.

Wir werden jetzt den Begriff der Entscheidungsfunktion auf natürliche Art und Weise verallgemeinern und erhalten damit die funktionalanalytische Grundlage für die im nächsten Abschnitt folgenden Definitionen und Aussagen. Die Mengen $\{P_\theta \mid \theta \in \Theta\}$ und $\mathcal{R}(E, D)$ werden dabei in größere Räume eingebettet, die für die LeCam-Theorie wesentliche strukturelle Eigenschaften besitzen. Sei dazu ein Experiment $E \in \mathcal{E}(\Theta)$ gegeben. Die Menge $\{P_\theta \mid \theta \in \Theta\}$ ist eine Teilmenge von

$$L(E) := \{\mu \in \mathcal{M}(\mathcal{X}) \mid \sigma \in \mathcal{M}(\mathcal{X}), \sigma \text{ singular zu allen } P_\theta \implies \sigma \text{ singular zu } \mu\},$$

dem sogenannten **L-Raum** des Experiments E , wobei $\mathcal{M}(\mathcal{X})$ den Vektorraum aller endlichen signierten Maße auf $(\mathcal{X}, \mathcal{A})$ bezeichnet. Versieht man den L-Raum mit der Variationsnorm $\|\cdot\|_V$, vgl. Definition A.1, so erhält man einen Banachraum und sogar einen Banach-Verband, vgl. Strasser (1985), S. 228.

Weiter sei ein Entscheidungsproblem (E, D, W) mit stetiger und beschränkter Verlustfunktion W gegeben. Schreiben wir die zur Entscheidungsfunktion ρ gehörige Risikofunktion

$$r_\rho(\theta) = \int_{\mathcal{X}} \int_D W_\theta(a) \rho(x, da) P_\theta(dx) =: \beta_\rho(W_\theta, P_\theta), \quad \theta \in \Theta,$$

in Abhängigkeit von W_θ und P_θ , so ist β_ρ linear in beiden Argumenten und die folgende Verallgemeinerung naheliegend: Eine bilineare Funktion $\beta : C_b(D) \times L(E) \rightarrow \mathbb{R}$ auf der

Menge $C_b(D)$ der stetigen und beschränkten Funktionen von D nach \mathbb{R} und auf dem L -Raum $L(E)$ mit

1. $|\beta(f, \mu)| \leq \|f\|_\infty \|\mu\|_V$ für alle $f \in C_b(D)$ und alle $\mu \in L(E)$,
2. $\beta(f, \mu) \geq 0$ für alle $f \in C_b(D)$, $f \geq 0$ und alle $\mu \in L(E)$, $\mu \geq 0$,
3. $\beta(1, \mu) = \mu(\mathcal{X})$ für alle $\mu \in L(E)$

heißt **verallgemeinerte Entscheidungsfunktion** (für E und D). Mit $\mathcal{R}_a(E, D)$ bezeichnen wir die Menge aller verallgemeinerten Entscheidungsfunktionen und die Funktion $r_\beta : \Theta \rightarrow \mathbb{R}$, $r_\beta(\theta) = \beta(W_\theta, P_\theta)$ nennen wir **Risikofunktion** zum Problem (E, D, W) bei gegebener verallgemeinerter Entscheidungsfunktion β .

Der Begriff der verallgemeinerten Entscheidungsfunktion lässt sich mittels

$$\beta(g, \mu) := \sup \{ \beta(f, \mu) \mid f \in C_b(D), f \leq g \}$$

auf natürliche Weise von $C_b(D)$ auf die Menge aller von unten halbstetigen und nach unten beschränkten Funktionen g erweitern. Dies ist allerdings nur sinnvoll für den Fall, dass μ ein Maß ist. Man vergleiche dazu auch Strasser (1985).

Wie in der Motivation bereits gesehen, existiert zu jeder Entscheidungsfunktion eine korrespondierende verallgemeinerte Entscheidungsfunktion. Ist bei gegebenem Entscheidungsproblem (E, D, W) eine Entscheidungsfunktion ρ_1 nicht schlechter als eine Entscheidungsfunktion ρ_2 , so gilt für die korrespondierenden verallgemeinerten Entscheidungsfunktionen

$$\beta_{\rho_1}(W_\theta, P_\theta) \leq \beta_{\rho_2}(W_\theta, P_\theta), \quad \theta \in \Theta.$$

Eine umgekehrte Korrespondenz ist nur in speziellen Fällen möglich. Dazu benötigen wir einen weiteren Begriff: Ein Experiment $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ heißt **dominiert**, falls ein σ -endliches Maß ν auf \mathcal{A} existiert, bezüglich dessen alle Maße P_θ absolutstetig sind, wenn also $P_\theta \ll \nu$ für alle $\theta \in \Theta$ gilt.

Satz 1.1 (vgl. Farrell (1967)) *Seien $E \in \mathcal{E}(\Theta)$ ein dominiertes Experiment und der Entscheidungsraum D ein lokal kompakter Raum mit abzählbarer Basis. Dann existiert zu jeder verallgemeinerten Entscheidungsfunktion $\beta \in \mathcal{R}_a(E, D)$ eine Entscheidungsfunktion $\rho_\beta \in \mathcal{R}(E, D)$ mit*

$$\beta(f, \mu) = \int_{\mathcal{X}} \int_D f(a) \rho_\beta(x, da) \mu(dx) \quad \text{für alle } \mu \in L(E) \text{ und } f \in C_c(D),$$

wobei $C_c(D)$ die Menge der stetigen Funktionen von D nach \mathbb{R} mit kompaktem Träger bezeichnet.

1.2 Abstände von Experimenten

In diesem Abschnitt wollen wir versuchen, den unterschiedlichen Informationsgehalt verschiedener Experimente $E = (\mathcal{X}_1, \mathcal{A}_1, \{P_\theta \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ und $F = (\mathcal{X}_2, \mathcal{A}_2, \{Q_\theta \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ mit gemeinsamem Parameterraum $\Theta \neq \emptyset$ zu messen. Um diese Experimente E und F vergleichen zu können, geben wir uns zunächst einen festen Entscheidungsraum D und eine feste Verlustfunktion $W = (W_\theta)_{\theta \in \Theta}$ vor und nennen (D, W) ein **Entscheidungspaar**. Die beiden statistischen Entscheidungsprobleme (E, D, W) und (F, D, W) werden nun auf die folgende Art und Weise miteinander verglichen.

Definition 1.2 Seien (D, W) ein beliebiges Entscheidungspaar mit von unten halbstetiger und nach unten beschränkter Verlustfunktion W und $\epsilon \geq 0$. Das Experiment E heißt **ϵ -defizient** oder **höchstens ϵ -unzureichend** bezüglich F (für das Paar (D, W)), in Zeichen: $E \underset{(D, W)}{\overset{\epsilon}{\geq}} F$, falls zu jeder verallgemeinerten Entscheidungsfunktion $\beta_2 \in \mathcal{R}_a(F, D)$ eine verallgemeinerte Entscheidungsfunktion $\beta_1 \in \mathcal{R}_a(E, D)$ mit

$$\beta_1(W_\theta, P_\theta) \leq \beta_2(W_\theta, Q_\theta) + \epsilon \|W_\theta\|_\infty \quad \text{für alle } \theta \in \Theta \quad (1.2)$$

existiert.

Gilt (1.2) für $\epsilon = 0$, so heißt E **mindestens so informativ** wie F (oder F **nicht informativer** als E) für das Paar (D, W) , in Zeichen: $E \underset{(D, W)}{\geq} F$.

Im Fall $\epsilon > 0$ ist obige Definition nur sinnvoll, wenn die Verlustfunktion auch nach oben beschränkt ist, vgl. Strasser (1985), S.257.

Lemma 1.3 Ist die Verlustfunktion in Definition 1.2 beschränkt, so gilt

$$E \underset{(D, W)}{\geq} F \quad \iff \quad E \underset{(D, W)}{\overset{\epsilon}{\geq}} F \quad \text{für alle } \epsilon > 0.$$

Bisher haben wir ein festes Entscheidungspaar (D, W) betrachtet. In der folgenden Definition werden Experimente bezüglich aller Entscheidungspaare (D, W) mit festem Parameterraum Θ , beliebigem topologischem Entscheidungsraum D sowie beliebiger auf D definierter und von unten halbstetiger Verlustfunktion W verglichen.

Definition 1.4 Seien $\epsilon \geq 0$ und der Parameterraum Θ beliebig fest. Dann heißt das Experiment E **ϵ -defizient** oder **höchstens ϵ -unzureichend** bezüglich F , in Zeichen: $E \overset{\epsilon}{\geq} F$, falls gilt

$$E \underset{(D, W)}{\overset{\epsilon}{\geq}} F \quad \text{für alle Entscheidungspaare } (D, W) \text{ mit} \quad (1.3)$$

von unten halbstetiger Verlustfunktion W .

Gilt (1.3) für $\epsilon = 0$, so heißt E **mindestens so informativ** wie F (oder F **nicht informativer** als E), in Zeichen: $E \supseteq F$.

Gelten $E \supseteq F$ und $F \supseteq E$, so heißen E und F **äquivalent** oder **gleich informativ**, und wir schreiben $E \sim F$.

Die Relation \sim ist eine *Äquivalenzrelation*, und wir können äquivalente Experimente zu einer *Äquivalenzklasse* zusammenfassen. Mit $\mathcal{E}(\Theta)/\sim$ bezeichnen wir den Raum dieser Äquivalenzklassen aller Experimente mit gemeinsamem Parameterraum Θ . Jetzt sind wir in der Lage, Abstände zwischen Experimenten einzuführen.

Definition 1.5 Seien $E, F \in \mathcal{E}(\Theta)$ zwei Experimente mit gemeinsamem Parameterraum Θ . Dann definieren wir

$$\delta(E, F) := \inf\{\epsilon > 0 \mid E \stackrel{\epsilon}{\supseteq} F\}$$

und

$$\Delta(E, F) := \max\{\delta(E, F), \delta(F, E)\}.$$

$\Delta(E, F)$ heißt **Defizienz** oder **Informationsabstand** der Experimente E und F .

Wir wollen zunächst einige wichtige Aussagen über den Informationsabstand Δ zusammenfassen. Einen Beweis für 1.-5. findet sich in Strasser (1985), S. 296 ff., Aussage 6. folgt direkt aus 5. mittels der Formeln (A.1) und (A.3) im Anhang.

Lemma 1.6 Sei $\Theta \neq \emptyset$ ein beliebiger Parameterraum. Dann ist

1. Δ eine Pseudometrik auf $\mathcal{E}(\Theta)$,
2. Δ eine Metrik auf $\mathcal{E}(\Theta)/\sim$,
3. der metrische Raum $(\mathcal{E}(\Theta)/\sim, \Delta)$ vollständig,
4. der metrische Raum $(\mathcal{E}(\Theta)/\sim, \Delta)$ kompakt, falls Θ endlich ist.

Sind $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$, $F = (\mathcal{X}, \mathcal{A}, \{Q_\theta \mid \theta \in \Theta\})$ Experimente mit gemeinsamem Grundraum $(\mathcal{X}, \mathcal{A})$, und bezeichnet d_H den Hellingerabstand, vgl. Definition A.2, so gilt weiter

5. $\Delta(E, F) \leq \sup\{\|P_\theta - Q_\theta\|_V \mid \theta \in \Theta\}$,
6. $\Delta(E, F) \leq \sqrt{2} \sup\{d_H(P_\theta, Q_\theta) \mid \theta \in \Theta\}$.

Die durch Δ auf $\mathcal{E}(\Theta)/\sim$ induzierte Topologie wird auch als **starke Topologie** bezeichnet.

Satz 1.7 (vgl. Strasser (1985), S. 272) Seien $E, F \in \mathcal{E}(\Theta)$ zwei dominierte Experimente mit gemeinsamem Parameterraum Θ . Dann sind die folgenden Aussagen äquivalent:

1. $\Delta(E, F) = 0$, d.h. die beiden Experimente sind äquivalent.
2. $H(E_\alpha) = H(F_\alpha)$ für alle $\alpha \in A(\Theta)$.

Dabei bezeichnen $A(\Theta)$ das System aller endlichen Teilmengen von Θ , E_α, F_α die auf den Parameterraum α eingeschränkten Experimente E, F und H die Hellingertransformierte, vgl. Definition A.5.

Für eine weitere Darstellung der Δ -Metrik benötigen wir zusätzliche Voraussetzungen an die Experimente.

Definition 1.8 Ein Experiment $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ heißt **polnisch**, falls der Beobachtungsraum \mathcal{X} ein polnischer Raum ist und \mathcal{A} die zugehörige Borelsche σ -Algebra.

Satz 1.9 Seien $E = (\mathcal{X}_1, \mathcal{A}_1, \{P_\theta \mid \theta \in \Theta\})$ dominiert und $F = (\mathcal{X}_2, \mathcal{A}_2, \{Q_\theta \mid \theta \in \Theta\})$ polnisch. Ist $M : \mathcal{X}_1 \times \mathcal{A}_2 \rightarrow [0, 1]$ ein Markoff-Kern, so definieren wir für $\theta \in \Theta$ das Wahrscheinlichkeitsmaß MP_θ auf \mathcal{A}_2 durch

$$MP_\theta(A_2) := \int_{\mathcal{X}_1} M(x_1, A_2) P_\theta(dx_1), \quad A_2 \in \mathcal{A}_2.$$

Dann gilt für den einseitigen δ -Abstand von E zu F

$$\delta(E, F) = \inf_M \sup_\theta \|MP_\theta - Q_\theta\|_V.$$

Dabei wird das Infimum über alle Markoff-Kerne $M : \mathcal{X}_1 \times \mathcal{A}_2 \rightarrow [0, 1]$ gebildet. Sind E und F beide dominiert und polnisch, so gilt

$$\Delta(E, F) = \max \left\{ \inf_{M_1} \sup_\theta \|M_1 P_\theta - Q_\theta\|_V, \inf_{M_2} \sup_\theta \|M_2 Q_\theta - P_\theta\|_V \right\},$$

wobei die Infima über alle Markoff-Kerne $M_1 : \mathcal{X}_1 \times \mathcal{A}_2 \rightarrow [0, 1]$ bzw. $M_2 : \mathcal{X}_2 \times \mathcal{A}_1 \rightarrow [0, 1]$ gebildet werden.

Beweis. Die Aussage folgt aus Nussbaum (1996), Proposition 9.2, und Strasser (1985), Lemma 59.5. \square

Mit $\mathcal{L}(X \mid P)$ bezeichnen wir künftig die Verteilung einer Zufallsvariablen X bezüglich eines Maßes P . Eine weitere nützliche Aussage beinhaltet der folgende Satz, vgl. Brown und Low (1996), S.2388.

Satz 1.10 Seien \mathcal{X}, \mathcal{Y} separable vollständige metrische Räume, $\Theta \neq \emptyset$ ein Parameterraum, $E = (\mathcal{X}, \mathcal{B}(\mathcal{X}), \{P_\theta \mid \theta \in \Theta\})$ ein Experiment und $S : \mathcal{X} \rightarrow \mathcal{Y}$ eine suffiziente Statistik für E . Dann sind die Experimente E und $F := (\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \{\mathcal{L}(S \mid P_\theta) \mid \theta \in \Theta\})$ äquivalent, d.h. es gilt

$$\Delta(E, F) = 0.$$

1.3 Lineare Modelle

Wir wollen in diesem Abschnitt einige Anwendungen der LeCam-Theorie für das Lineare Modell skizzieren. Dazu müssen wir zunächst klären, was wir unter einem solchen verstehen. Seien

- $k, n \in \mathbb{N}$ fest vorgegeben,
- der *Parameterraum* H ein endlichdimensionaler Hilbertraum mit Dimension $\dim(H) = k$,
- der *Beobachtungsraum* \mathcal{X} ein endlichdimensionaler Hilbertraum mit Dimension $\dim(\mathcal{X}) = n$, versehen mit der Borelschen σ -Algebra $\mathcal{B}(\mathcal{X})$,
- $A : \mathcal{X} \rightarrow H$ ein bekannter linearer Operator und A^* der zu A konjugierte Operator,
- $C : \mathcal{X} \rightarrow \mathcal{X}$ ein bekannter, symmetrischer positiv definiten linearer Operator.

Wir beobachten eine \mathcal{X} -wertige Zufallsvariable X mit Erwartungswert

$$EX = A^*h$$

und existenter Covarianz

$$\text{Cov}X = C.$$

Zur Definition von Erwartungswert und Covarianz vergleiche man Definition A.14. Sind Basen von H und \mathcal{X} gegeben, so können wir A mit einer $k \times n$ -Matrix und C mit einer $n \times n$ -Matrix identifizieren. A^* wird dann auch **Modellmatrix** genannt und C **Covarianzmatrix**. Für den Fall, dass C nur positiv semidefinit ist, gelten die Aussagen dieses Abschnitts analog, wenn man die Inverse C^{-1} jeweils durch die Moore-Penrose-Inverse ersetzt.

Seien \mathcal{P}_C die Menge aller Verteilungen auf $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ mit Erwartungswert 0 und existierender positiv definiten Covarianz C und $S_h : \mathcal{X} \rightarrow \mathcal{X}$ gegeben durch $x \mapsto x + A^*h$. Die Verteilung von S_h unter einem $P \in \mathcal{P}_C$ hat dann den Erwartungswert A^*h und dieselbe Covarianz C . Dann bezeichnen wir

$$\mathcal{L}(A, C) := (\mathcal{X}, \mathcal{B}(\mathcal{X}), \{\mathcal{L}(S_h \mid P) \mid h \in H, P \in \mathcal{P}_C\}) \quad (1.4)$$

als **Lineares Modell** für den unbekannt Parameter $h \in H$ und die unbekannt Verteilung $P \in \mathcal{P}_C$. Sind die Zufallsvariablen normalverteilt, so heißt

$$\mathcal{N}(A, C) := (\mathcal{X}, \mathcal{B}(\mathcal{X}), \{N(A^*h, C) \mid h \in H\}) \quad (1.5)$$

Lineares Modell für den unbekannt Parameter $h \in H$ mit normalverteilt Fehler. Dabei sei $N(A^*h, C)$ eine Normalverteilung auf dem Hilbertraum \mathcal{X} mit Erwartungswert A^*h und Covarianz C . Auf den Begriff der Normalverteilung auf einem Hilbertraum gehen wir in Beispiel 1.26 näher ein.

Aussagen über den Informationsabstand von Linearen Modellen, vgl. Definition 1.5, insbesondere unter Normalverteilungsannahme, finden sich in Torgersen (1991). Die anschließenden Sätze werden sich im Folgenden als nützlich erweisen.

Satz 1.11 (vgl. Torgersen (1991), S. 481) *Seien $\mathcal{X}_1, \mathcal{X}_2$ Hilberträume mit den Dimensionen n_1, n_2 und seien $A : \mathcal{X}_1 \rightarrow H, B : \mathcal{X}_2 \rightarrow H$ lineare Operatoren. Dann gilt*

$$\begin{aligned} \frac{1}{2} \|N(0, AA^*) - N(0, BB^*)\|_V &\leq \Delta(\mathcal{N}(A, I_{\mathcal{X}_1}), \mathcal{N}(B, I_{\mathcal{X}_2})) \\ &\leq \|N(0, AA^*) - N(0, BB^*)\|_V. \end{aligned}$$

Dabei bezeichne $I_{\mathcal{X}_j}$ die Identität auf \mathcal{X}_j , $j = 1, 2$, und 0 das Nullelement von H .

Satz 1.12 (vgl. Torgersen (1991), S. 436) *Seien $\mathcal{N}(A, C_1), \mathcal{N}(B, C_2)$ Lineare Modelle mit normalverteilten Fehlern und positiv definiten Covarianzen C_1, C_2 . Beide Experimente mögen den gemeinsamen Parameterraum H haben. Dann sind die folgenden Aussagen äquivalent:*

1. *Das Experiment $\mathcal{N}(B, C_2)$ ist im Sinne LeCam's informativer, vgl. Definition 1.4, als $\mathcal{N}(A, C_1)$, also*

$$\mathcal{N}(A, C_1) \subseteq \mathcal{N}(B, C_2).$$

2. *Bezeichnet \leq_L die Löwner-Ordnung, so gilt*

$$AC_1^{-1}A^* \leq_L BC_2^{-1}B^*.$$

(Für zwei symmetrische Operatoren $D_1, D_2 : H \rightarrow H$ gilt $D_1 \leq_L D_2$ genau dann, wenn $D_2 - D_1$ positiv semidefinit ist. Die Löwner-Ordnung ist keine Totalordnung.)

Satz 1.13 (vgl. Torgersen (1991), S. 481) *Seien $\mathcal{N}(A, I_{\mathcal{X}_1})$ und $\mathcal{N}(B, I_{\mathcal{X}_2})$ Lineare Modelle mit normalverteilten Fehlern und gemeinsamem Parameterraum H . Dann gilt nach Satz 1.12*

$$\mathcal{N}(A, I_{\mathcal{X}_1}) \subseteq \mathcal{N}(B, I_{\mathcal{X}_2}) \iff AA^* \leq_L BB^*.$$

In diesem Fall gilt für den Informationsabstand der Experimente

$$\Delta(\mathcal{N}(A, I_{\mathcal{X}_1}), \mathcal{N}(B, I_{\mathcal{X}_2})) = \delta(\mathcal{N}(A, I_{\mathcal{X}_1}), \mathcal{N}(B, I_{\mathcal{X}_2})) = \|N(0, AA^*) - N(0, BB^*)\|_V.$$

Eine andere naheliegende Möglichkeit, den Informationsgehalt von Linearen Modellen zu vergleichen, bietet die *lineare, erwartungstreue Schätzung bei quadratischem Schaden*. Seien hierzu $\mathcal{L}(A, C_1)$, $\mathcal{L}(B, C_2)$ Lineare Modelle mit gemeinsamem Parameterraum H . Wir nennen das Experiment $\mathcal{L}(A, C_1)$ bezüglich des linearen erwartungstreuen Schätzens bei quadratischem Schaden **nicht informativer** als das Experiment $\mathcal{L}(B, C_2)$ (oder $\mathcal{L}(B, C_2)$ **mindestens so informativ** wie $\mathcal{L}(A, C_1)$), falls es zu jeder Linearform $l : H \rightarrow \mathbb{R}$ und jedem linearen erwartungstreuen Schätzer T_1 für $l(h)$ im Modell $\mathcal{L}(A, C_1)$ einen linearen erwartungstreuen Schätzer T_2 für $l(h)$ im Modell $\mathcal{L}(B, C_2)$ gibt mit nicht größerer Varianz, also

$$\text{Var}(T_2) \leq \text{Var}(T_1).$$

In diesem Fall schreiben wir

$$\mathcal{L}(A, C_1) \subseteq_q \mathcal{L}(B, C_2). \quad (1.6)$$

Die Varianzen der Schätzer T_1 bzw. T_2 sind auf Grund der linearen Struktur konstant für jede Verteilung aus $\mathcal{L}(A, C_1)$ bzw. $\mathcal{L}(B, C_2)$.

Aus der Theorie der Linearen Modelle (Gauß-Markoff-Theorem) und Satz 1.12 folgt unmittelbar

Satz 1.14 (vgl. Torgersen (1991), S. 465) *Seien $\mathcal{L}(A, C_1)$, $\mathcal{L}(B, C_2)$, $\mathcal{N}(A, C_1)$, $\mathcal{N}(B, C_2)$ Lineare Modelle mit positiv definiten Covarianzoperatoren C_1, C_2 . Alle Experimente mögen den gemeinsamen Parameterraum H haben. Dann sind die folgenden Aussagen äquivalent:*

1. $\mathcal{N}(A, C_1) \subseteq \mathcal{N}(B, C_2)$.
2. $\mathcal{L}(A, C_1) \subseteq_q \mathcal{L}(B, C_2)$.

Beispiel 1.15 (vgl. Torgersen (1991), S. 431)

Sei $E = \mathcal{N}(A, C) = (\mathcal{X}, \mathcal{B}(\mathcal{X}), \{N(A^*h, C) \mid h \in H\})$ ein Lineares Modell mit Modell-Operator A und positiv definiten Covarianz C . Seien $\alpha := \{h_1, \dots, h_r\}$, $r \in \mathbb{N}$, beliebig und $z = (z_1, \dots, z_r)^\top \in S_\alpha$ beliebig, wobei S_α der Standard-Simplex des $\mathbb{R}^{|\alpha|}$ ist, vgl. Definition A.5. Dann gilt

$$\begin{aligned} H(E_\alpha)(z) &= \exp \left\{ \frac{1}{2} \left(\sum_{i,j=1}^r z_i z_j \langle h_i, AC^{-1}A^*h_j \rangle - \sum_{i=1}^r z_i \langle h_i, AC^{-1}A^*h_i \rangle \right) \right\} \\ &= \exp \left\{ -\frac{1}{4} \sum_{i,j=1}^r z_i z_j \langle h_i - h_j, AC^{-1}A^*(h_i - h_j) \rangle \right\}. \end{aligned}$$

Beispiel 1.16 (Hellingerabstand von $N(A^*h_1, I_{\mathcal{X}})$ und $N(A^*h_2, I_{\mathcal{X}})$)

Sind $N(A^*h_1, I_{\mathcal{X}})$ und $N(A^*h_2, I_{\mathcal{X}})$ zwei Verteilungen des Linearen Modells $\mathcal{N}(A, I_{\mathcal{X}})$,

und bezeichnet a die Affinität zweier Maße, vgl. Definition A.4, so gilt

$$\begin{aligned} d_H^2(N(A^*h_1, I_{\mathcal{X}}), N(A^*h_2, I_{\mathcal{X}})) &= 1 - a(N(A^*h_1, I_{\mathcal{X}}), N(A^*h_2, I_{\mathcal{X}})) \\ &= 1 - \exp \left\{ -\frac{1}{8} \langle A^*(h_1 - h_2), A^*(h_1 - h_2) \rangle \right\}. \end{aligned}$$

Beweis. Folgt direkt aus Lemma A.11 und Beispiel 1.15. \square

Bemerkung 1.17 Definiert man den linearen Unterraum $\tilde{H} := A^*(H) \subseteq \mathcal{X}$, so kann man das Experiment E aus Beispiel 1.15 auch in der Form $E = (\mathcal{X}, \mathcal{B}(\mathcal{X}), \{N(\tilde{h}, C) \mid \tilde{h} \in \tilde{H}\})$ schreiben. Der (Modell-) Operator A tritt also gar nicht mehr explizit auf, sondern nur noch der Bildraum \tilde{H} seiner Konjugierten A^* . Für beliebiges $\alpha := \{\tilde{h}_1, \dots, \tilde{h}_r\}$, $r \in \mathbb{N}$, und $z = (z_1, \dots, z_r)^\top \in S_\alpha$ erhält man dann die Hellingertransformierte in der Form

$$H(E_\alpha)(z) = \exp \left\{ -\frac{1}{4} \sum_{i,j=1}^r z_i z_j \langle \tilde{h}_i - \tilde{h}_j, C^{-1}(\tilde{h}_i - \tilde{h}_j) \rangle \right\}.$$

Entsprechend erhält man in Beispiel 1.16 für $\tilde{h}_1, \tilde{h}_2 \in \tilde{H}$ den Hellingerabstand

$$d_H^2(N(\tilde{h}_1, I_{\mathcal{X}}), N(\tilde{h}_2, I_{\mathcal{X}})) = 1 - \exp \left\{ -\frac{1}{8} \|\tilde{h}_1 - \tilde{h}_2\|^2 \right\}.$$

1.4 Gaußexperimente

In diesem Abschnitt werden wir die sogenannten Gaußexperimente einführen. Sie sind die wichtigsten Experimente der Statistik und spielen bei vielen asymptotischen Überlegungen eine entscheidende Rolle. Sei dazu $\Theta \neq \emptyset$ ein beliebiger Parameterraum.

Definition 1.18 (Gaußprozess) Ein stochastischer Prozess $(\Omega, \mathcal{A}, P, \{X_\theta \mid \theta \in \Theta\})$ mit reellwertigen Zufallsvariablen X_θ heißt **Gaußprozess**, wenn alle endlichdimensionalen Verteilungen (evtl. entartete) Normalverteilungen sind.

Ein Experiment $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ mit Parameterraum Θ heißt **homogen**, falls alle Maße P_θ , $\theta \in \Theta$, äquivalent sind.

Definition 1.19 (Gaußexperiment) Ein homogenes Experiment $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ heißt **Gaußexperiment**, wenn ein $\theta_0 \in \Theta$ existiert, so dass der Log-Likelihood-Prozess unter P_{θ_0} :

$$\left(\mathcal{X}, \mathcal{A}, P_{\theta_0}, \left\{ \log \left(\frac{dP_\theta}{dP_{\theta_0}} \right) \mid \theta \in \Theta \right\} \right)$$

ein Gaußprozess ist. Wir nennen den Covarianzkern dieses Prozesses dann **Covarianzkern von E zur Basis θ_0** .

Ist Θ Teilmenge eines linearen Raumes, ist $0 \in \Theta$ und ist der Covarianzkern von E zur Basis 0 bilinear, so heißt das Experiment E ein **Gaußshift**.

An einem einfachen Beispiel wollen wir verdeutlichen, dass diese abstrakte Definition eines Gaußexperimentes mit der üblichen Vorstellung von einem Gaußexperiment korrespondiert.

Beispiel 1.20 Es sei $\sigma^2 > 0$ fest vorgegeben. Dann ist das durch die eindimensionalen Normalverteilungen $N(\theta, \sigma^2)$ gegebene Experiment

$$\mathcal{N}(1, \sigma^2) = (\mathbb{R}, \mathcal{B}, \{N(\theta, \sigma^2) \mid \theta \in \mathbb{R}\})$$

ein Gaußexperiment mit Covarianzkern $R_{\theta_0}(\theta_1, \theta_2) = \frac{(\theta_1 - \theta_0)(\theta_2 - \theta_0)}{\sigma^2}$, $\theta_1, \theta_2 \in \mathbb{R}$, zur Basis $\theta_0 \in \mathbb{R}$. Es ist sogar ein Gaußshift. Für den Log-Likelihood-Prozess gilt

$$\mathcal{L} \left(\log \left(\frac{dP_\theta}{dP_{\theta_0}} \right) \mid P_{\theta_0} \right) = N \left(-\frac{1}{2} \cdot \frac{(\theta - \theta_0)^2}{\sigma^2}, \frac{(\theta - \theta_0)^2}{\sigma^2} \right).$$

Man beachte, dass die Erwartungswerte dieser Normalverteilungen genau das $-\frac{1}{2}$ -fache der jeweiligen Varianzen sind. Der Trend des Log-Likelihood-Prozesses ist somit direkt aus dem Covarianzkern ablesbar. Dagegen ist das Experiment

$$(\mathbb{R}, \mathcal{B}, \{N(\theta, \sigma^2) \mid \theta \in \mathbb{R}, \sigma^2 > 0\})$$

kein Gaußexperiment im Sinne von Definition 1.19.

Die folgenden Aussagen sollen die besonders einfachen Strukturen von Gaußexperimenten verdeutlichen.

Bemerkung 1.21 Ist E ein Gaußexperiment, so ist für jedes $\theta_0 \in \Theta$ der *Log-Likelihood-Prozess* unter P_{θ_0} ein Gaußprozess. Für die Covarianzkern R_{θ_0} bzw. $R_{\theta'_0}$ zu den Basen θ_0 bzw. θ'_0 gilt

$$R_{\theta_0}(\theta_1, \theta_2) - \frac{1}{2} (R_{\theta_0}(\theta_1, \theta_1) + R_{\theta_0}(\theta_2, \theta_2)) = R_{\theta'_0}(\theta_1, \theta_2) - \frac{1}{2} (R_{\theta'_0}(\theta_1, \theta_1) + R_{\theta'_0}(\theta_2, \theta_2)) \quad (1.7)$$

für alle $\theta_1, \theta_2 \in \Theta$, vgl. Janssen, Milbrodt und Strasser (1985), S. 21. Der Zusammenhang zwischen den Covarianzkernen zu verschiedenen Basen gilt nicht allgemein für Log-Likelihood-Prozesse anderer homogener Experimente.

Lemma 1.22 (Hellingertransformierte eines Gaußexperimentes)

Seien E ein Gaußexperiment, $\theta_0 \in \Theta$ beliebig und R_{θ_0} der Covarianzkern von E zur Basis θ_0 . Seien $\alpha := \{\theta_1, \dots, \theta_r\} \in A(\Theta)$ und $z = (z_1, \dots, z_r)^\top \in S_\alpha$ beliebig. Dann gilt

$$H(E_\alpha)(z) = \exp \left\{ \frac{1}{2} \left(\sum_{i,j=1}^r z_i z_j R_{\theta_0}(\theta_i, \theta_j) - \sum_{i=1}^r z_i R_{\theta_0}(\theta_i, \theta_i) \right) \right\}.$$

Die Hellingertransformierte ist unabhängig von der Basis θ_0 .

Beweis. Vgl. Becker (1983), S. 25, oder Janssen et al. (1985), S. 20.

Bemerkung 1.23 Der Covarianzkern eines Gaußprozesses ist symmetrisch und positiv semidefinit. Umgekehrt existiert zu jedem symmetrischen positiv semidefiniten Kern $R : \Theta \times \Theta \rightarrow \mathbb{R}$ ein Gaußprozess mit Covarianzkern R , vgl. Strasser (1985), S. 340. Ebenso existiert ein Gaußexperiment mit der Hellingertransformierten

$$H(E_\alpha)(z) = \exp \left\{ \frac{1}{2} \left(\sum_{i,j=1}^r z_i z_j R(\theta_i, \theta_j) - \sum_{i=1}^r z_i R(\theta_i, \theta_i) \right) \right\}$$

für $\alpha := \{\theta_1, \dots, \theta_r\} \in A(\Theta)$, $z = (z_1, \dots, z_r)^\top \in S_\alpha$, vgl. Becker (1983), S.25.

Ist $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum, so ist das Innenprodukt $\langle \cdot, \cdot \rangle$ ein symmetrischer positiv definit Kern auf $H \times H$. Das führt zur folgenden Definition.

Definition 1.24 (Standard-Gaußshift auf Hilbertraum)

Der Parameterraum $\Theta = (H, \langle \cdot, \cdot \rangle)$ sei ein separabler Hilbertraum. Wir nennen ein Gaußexperiment $E \in \mathcal{E}(H)$, dessen Covarianzkern R_0 zur Basis 0 gleich dem Innenprodukt auf H ist, also

$$R_0(h_1, h_2) = \langle h_1, h_2 \rangle, \quad h_1, h_2 \in H,$$

einen **Standard-Gaußshift auf dem Hilbertraum H** . Für einen solchen Standard-Gaußshift schreiben wir künftig auch G_H .

Bemerkung 1.25

1. Die Existenz eines Standard-Gaußshifts ist für jeden separablen Hilbertraum durch Bemerkung 1.23 gewährleistet. Alle Standard-Gaußshifts auf demselben Hilbertraum sind im Sinne LeCam's äquivalent. Man beachte dazu Satz 1.7 und die Tatsache, dass mit dem Covarianzkern die Hellingertransformierte eindeutig festgelegt ist.
2. Die Separabilität des Hilbertraums gewährleistet die Messbarkeit des Innenproduktes.
3. Für die Hellingertransformierte eines Standard-Gaußshifts G_H gilt nach Lemma 1.22

$$H(E_\alpha)(z) = \exp \left\{ \frac{1}{2} \left(\sum_{i,j=1}^r z_i z_j \langle h_i, h_j \rangle - \sum_{i=1}^r z_i \langle h_i, h_i \rangle \right) \right\} \quad (1.8)$$

$$= \exp \left\{ -\frac{1}{4} \sum_{i,j=1}^r z_i z_j \langle h_i - h_j, h_i - h_j \rangle \right\} \quad (1.9)$$

für $\alpha := \{h_1, \dots, h_r\} \in A(H)$, $z = (z_1, \dots, z_r)^\top \in S_\alpha$. Dabei erhält man Gleichung (1.9) durch elementare Berechnungen aus (1.8). Man beachte dafür: Für $z \in S_\alpha$ ist $z_1 + \dots + z_r = 1$.

Wir wollen jetzt einige Standard-Gaußshifts auf Hilberträumen explizit angeben. Diese werden in den asymptotischen Überlegungen zu Regressionsmodellen in Kapitel 2 als Grenzexperimente auftreten.

Beispiel 1.26 (Standard-Gaußshift auf einem Hilbertraum)

Eine Zufallsvariable X mit Werten in einem separablen Hilbertraum $(H, \langle \cdot, \cdot \rangle)$ heißt **normalverteilt**, falls $\langle X, h \rangle$ für alle $h \in H$ eine reelle Normalverteilung besitzt. In diesem Fall existieren der Erwartungswert γ und die Covarianz S von X , vgl. Anhang A.2. Man nennt die Verteilung von X eine **Normalverteilung mit Erwartungswert γ und Covarianz S** und schreibt $N(\gamma, S)$.

Ist H ein *endlichdimensionaler* Hilbertraum und I_H die Identität auf H , so existiert das Wahrscheinlichkeitsmaß $N(0, I_H)$ und wird **Standardnormalverteilung** genannt. Leicht sieht man, dass das Experiment

$$G_H = (H, \mathcal{B}(H), \{N(h, I_H) \mid h \in H\})$$

der *Standard-Gaußshift* auf H ist, vgl. Strasser (1985), S 345.

Ist H ein *unendlichdimensionaler* Hilbertraum, so ist die Identität I_H kein kompakter Operator mehr. Da die Covarianz einer Normalverteilung aber ein kompakter Operator sein muss, vgl. Kuo (1975), S. 29, existiert keine Standardnormalverteilung $N(0, I_H)$ auf H . Wir können den Standard-Gaußshift auf einem unendlichdimensionalen Hilbertraum also nicht in ähnlich einfacher Weise angeben wie im endlichdimensionalen Fall. Diese Problematik führt zur Theorie der Abstrakten Wiener-Räume, vgl. Strasser (1985), S. 350 ff..

Beispiel 1.27 (Standard-Gaußshift mit Beobachtungsraum $C[0, 1]$)

Dieses Beispiel baut auf der in Anhang A.6 dargestellten Theorie kernreproduzierender Hilberträume auf. Sei dazu $Z(s)$, $s \in [0, 1]$, ein zentrierter Gaußprozess mit stetigen Pfaden und Covarianzkern $R : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$, und sei $(H(R), \langle \cdot, \cdot \rangle_{H(R)})$ der zugehörige kernreproduzierende Hilbertraum. Für $h \in H(R) \subseteq C[0, 1]$ ist

$$Z_h(s) := h(s) + Z(s), \quad s \in [0, 1],$$

wiederum ein Gaußprozess mit stetigen Pfaden. P^Z, P^{Z_h} seien die zu den Prozessen Z bzw. Z_h gehörigen Verteilungen auf $(C[0, 1], \mathcal{B}(C[0, 1]))$. Dann ist das statistische Experiment

$$G_{H(R)} = (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{Z_h} \mid h \in H(R)\})$$

der Standard-Gaußshift auf $H(R)$. Nach den Formeln (A.14) und (A.18) in Anhang A.6 hat $G_{H(R)}$ nämlich den Covarianzkern

$$\text{Cov} \left(\log \left(\frac{dP^{Z_{h_1}}}{dP^Z} \right), \log \left(\frac{dP^{Z_{h_2}}}{dP^Z} \right) \right) = \langle h_1, h_2 \rangle_{H(R)},$$

$h_1, h_2 \in H(R)$, zur Basis 0. Für die Hellingertransformierte von $G_{H(R)}$ gilt nach Formel (1.8) bzw. Formel (1.9)

$$\begin{aligned} H(E_\alpha)(z) &= \exp \left\{ \frac{1}{2} \left(\sum_{i,j=1}^r z_i z_j \langle h_i, h_j \rangle_{H(R)} - \sum_{i=1}^r z_i \|h_i\|_{H(R)}^2 \right) \right\} \\ &= \exp \left\{ -\frac{1}{4} \sum_{i,j=1}^r z_i z_j \|h_i - h_j\|_{H(R)}^2 \right\} \end{aligned}$$

für $\alpha := \{h_1, \dots, h_r\} \in A(H(R))$, $z = (z_1, \dots, z_r)^\top \in S_\alpha$.

Zum Abschluss wollen wir eine Verbindung zu den Linearen Modellen aus Abschnitt 1.3 herstellen.

Beispiel 1.28 Es sei $A \in \mathbb{R}^{k \times k}$ eine symmetrische positiv definite Matrix, durch die mittels $\langle x, y \rangle_A := x^\top A y$, $x, y \in \mathbb{R}^k$, ein Innenprodukt auf \mathbb{R}^k definiert wird. H_A sei der Hilbertraum \mathbb{R}^k , versehen mit diesem Innenprodukt. Mit $A^{\frac{1}{2}}$ bezeichnen wir die eindeutig bestimmte symmetrische und positiv definite Wurzel von A . Dann sind der Standard-Gaußshift G_{H_A} und das Lineare Modell $\mathcal{N}(A^{\frac{1}{2}}, I_k)$ (mit Beobachtungs- und Parameterraum \mathbb{R}^k) im Sinne LeCam's äquivalent, also $G_{H_A} \sim \mathcal{N}(A^{\frac{1}{2}}, I_k)$, oder anders gesagt: $\mathcal{N}(A^{\frac{1}{2}}, I_k)$ ist der Standard-Gaußshift auf H_A .

Beweis. Nach Satz 1.7 genügt es, die Hellingertransformierten von G_{H_A} und $\mathcal{N}(A^{\frac{1}{2}}, I_k)$ zu vergleichen. Diese stimmen aber nach Beispiel 1.15 und Formel (1.8) überein. \square

1.5 Ein Satz von Luschgy für Gaußexperimente mit Beobachtungsraum $\mathbf{C}[0, 1]$

Im Abschnitt 1.3 haben wir untersucht, wann ein Lineares Modell mit normalverteilten Fehlern im Sinne LeCam's mindestens so informativ ist wie ein anderes derartiges Modell. Zentrum dieser Theorie, die auf die Arbeiten des norwegischen Mathematikers Erik Torgersen zurückgeht, ist Satz 1.12, der einen Zusammenhang zur Löwner-Ordnung von Matrizen herstellt, oder anders gesagt, zur positiv Semidefinitheit der Differenz zweier Bilinearformen.

Während die Arbeiten von Torgersen einen endlichdimensionalen Beobachtungsraum zu Grunde legen, untersucht Harald Luschgy in seinen Arbeiten Luschgy (1991a) und Luschgy (1992) die Bedeutung der LeCam'schen Informationshalbordnung „ \subseteq “ für Gaußexperimente mit Beobachtungsraum $C[0, 1]$. Ein solches Experiment haben wir bereits in Beispiel 1.27 kennengelernt. Im Folgenden wollen wir einen speziellen Satz von Luschgy anführen und schildern zunächst die zu Grunde liegende Situation.

Seien Z_1, Z_2 zwei Gaußprozesse mit stetigen Pfaden auf dem Intervall $[0, 1]$. Mit R_i , $i = 1, 2$, bezeichnen wir ihre Kovarianzkerne, und $(H(R_i), \langle \cdot, \cdot \rangle_{H(R_i)})$ seien die zugehörigen kernreproduzierenden Hilberträume. Weiter seien $\Theta \neq \emptyset$ ein beliebiger Parameterraum und $S_i : \Theta \times [0, 1] \rightarrow \mathbb{R}$ bekannte Funktionen mit

$$S_i(\theta, \cdot) \in H(R_i) \quad \text{für alle } \theta \in \Theta \text{ und } i = 1, 2. \quad (1.10)$$

Wir beobachten die stochastischen Prozesse

$$Z_{\theta,i}(t) := S_i(\theta, t) + Z_i(t), \quad t \in [0, 1], \theta \in \Theta.$$

Da die Prozesse Z_i stetige Pfade haben, gilt $H(R_i) \subseteq C[0, 1]$ nach Abschnitt A.6, und wegen (1.10) haben dann auch die Prozesse $Z_{\theta,i}$ stetige Pfade. Mit $P^{\theta,i}$ bezeichnen wir die Verteilungen der Prozesse $Z_{\theta,i}$ auf $(C[0, 1], \mathcal{B}(C[0, 1]))$. Im Folgenden werden wir die Experimente

$$E_i := (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{\theta,i} \mid \theta \in \Theta\}), \quad i = 1, 2,$$

betrachten. Wie in Beispiel 1.27 sieht man, dass die Experimente E_i Gaußexperimente mit den Kovarianzkernen

$$K_{\theta_0,i}(\theta_1, \theta_2) = \langle S_i(\theta_1, \cdot) - S_i(\theta_0, \cdot), S_i(\theta_2, \cdot) - S_i(\theta_0, \cdot) \rangle_{H(R_i)}, \quad \theta_1, \theta_2 \in \Theta,$$

zur Basis $\theta_0 \in \Theta$ sind.

Satz 1.29 (vgl. Luschgy (1991a), S. 212) *Unter den oben beschriebenen Voraussetzungen gelten die folgenden Aussagen.*

1. *Ist $K_{\theta_0,1} - K_{\theta_0,2}$ positiv semidefinit für eine beliebige Basis $\theta_0 \in \Theta$ (und damit auch für alle Basen $\theta_0 \in \Theta$), d.h. gilt*

$$\sum_{j,l=1}^n \alpha_j \alpha_l (K_{\theta_0,1} - K_{\theta_0,2})(\theta_j, \theta_l) \geq 0$$

für alle $\theta_1, \dots, \theta_n \in \Theta$, $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ und $n \in \mathbb{N}$, so folgt

$$E_1 \supseteq E_2,$$

d.h. das Experiment E_1 ist mindestens so informativ wie das Experiment E_2 .

2. *Ist $E_1 \supseteq E_2$, so ist $\exp(K_{\theta_0,1}) - \exp(K_{\theta_0,2})$ positiv semidefinit für jede Basis $\theta_0 \in \Theta$. Insbesondere gilt dann $K_{\theta_0,1}(\theta, \theta) \geq K_{\theta_0,2}(\theta, \theta)$ für alle $\theta_0, \theta \in \Theta$.*
3. *$E_1 \sim E_2$ genau dann, wenn für eine Basis $\theta_0 \in \Theta$ (und damit für alle Basen) die Kovarianzkerne identisch sind. Das ist auch äquivalent zu $K_{\theta_0,1}(\theta, \theta) = K_{\theta_0,2}(\theta, \theta)$ für alle $\theta_0, \theta \in \Theta$.*

1.6 Schwache Konvergenz von Experimenten

Der mit der Metrik Δ auf $\mathcal{E}(\Theta)/\sim$ verbundene Konvergenzbegriff für Experimente ist für viele Anwendungen zu stark, so dass wir einen schwächeren Konvergenzbegriff benötigen. Experimente, die wir im Sinne dieses Konvergenzbegriffs als „schwache Grenzwerte“ von Experimentfolgen erhalten, werden nicht mehr die vollständige „asymptotische Information“ der Experimentfolge beinhalten. Dennoch ist „viel“ Information in ihnen enthalten. In Abschnitt 1.8.3 werden wir zum Beispiel mittels schwacher Konvergenz asymptotisch optimale Testfolgen konstruieren.

Definition 1.30 Seien $\Theta \neq \emptyset$ ein Parameterraum und $E_n, E \in \mathcal{E}(\Theta)$. Die Experimentfolge $(E_n)_{n \in \mathbb{N}}$ **konvergiert schwach** gegen das Experiment E , in Zeichen: $E_n \xrightarrow{w} E$, wenn gilt

$$\Delta(E_{n,\alpha}, E_\alpha) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für alle endlichen Teilmengen } \alpha \text{ von } \Theta.$$

Bemerkung 1.31 (vgl. Strasser (1985), S. 298) Die durch den Begriff der schwachen Konvergenz erzeugte Topologie auf $\mathcal{E}(\Theta)/\sim$ wird die **schwache Topologie** genannt. $\mathcal{E}(\Theta)/\sim$ ist bezüglich der schwachen Topologie *kompakt*.

Die folgende Charakterisierung der schwachen Konvergenz findet sich zum Beispiel in Strasser (1985), S. 302.

Satz 1.32 Seien $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,\theta} \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ und $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$. Dann sind die folgenden Aussagen äquivalent:

1. $E_n \xrightarrow{w} E$,
2. $H(E_{n,\alpha}) \rightarrow H(E_\alpha)$ punktweise auf S_α für alle $\alpha \in A(\Theta)$.

Mit Lemma A.11 folgt daraus

Korollar 1.33 Seien E_n, E Experimente wie in Satz 1.32 mit $E_n \xrightarrow{w} E$. Dann gilt für $\theta_1, \theta_2 \in \Theta$

$$\lim_{n \rightarrow \infty} d_H(P_{n,\theta_1}, P_{n,\theta_2}) = d_H(P_{\theta_1}, P_{\theta_2}).$$

Wie wir bereits in Abschnitt 1.4 erwähnt haben, heißt ein Experiment $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ mit Parameterraum Θ **homogen**, falls alle Maße P_θ , $\theta \in \Theta$, äquivalent sind.

Eine Folge von Experimenten $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,\theta} \mid \theta \in \Theta\})$, $n \in \mathbb{N}$, heißt **asymptotisch homogen**, falls für alle $\theta_1, \theta_2 \in \Theta$ gilt

$$\lim_{n \rightarrow \infty} P_{n,\theta_1}(A_n) = 0 \quad \text{für } A_n \in \mathcal{A}_n \quad \implies \quad \lim_{n \rightarrow \infty} P_{n,\theta_2}(A_n) = 0.$$

Ist Θ ein metrischer Raum mit Metrik d , so heißt $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ **stetig**, falls die Abbildung

$$f_E : (\Theta, d) \rightarrow (\mathcal{M}(\mathcal{X}), \|\cdot\|_V), \quad f_E(\theta) := P_\theta$$

stetig ist. Die Experimentfolge E_n heißt **asymptotisch stetig**, falls für alle $\theta_1 \in \Theta$ gilt

$$\forall \epsilon > 0 \exists \delta > 0 : \quad d(\theta_1, \theta_2) < \delta \quad \Longrightarrow \quad \limsup_{n \rightarrow \infty} \|P_{n, \theta_1} - P_{n, \theta_2}\|_V < \epsilon.$$

Bemerkung 1.34 (vgl. Strasser (1985), S. 304 f.) Seien $E_n, E \in \mathcal{E}(\Theta)$, $n \in \mathbb{N}$, mit $E_n \xrightarrow{w} E$. Dann gilt

$$(E_n)_{n \in \mathbb{N}} \text{ asymptotisch homogen} \iff E \text{ homogen.}$$

Ist Θ zudem ein metrischer Raum, so gilt

$$(E_n)_{n \in \mathbb{N}} \text{ asymptotisch stetig} \iff E \text{ stetig.}$$

1.7 Schwache Konvergenz gegen einen Standard-Gaußshft auf einem Hilbertraum

Definition 1.35 (Stochastische Konvergenz) Seien $(\Omega_n, \mathcal{A}_n, P_n)$ Maßräume und $f_n : \Omega_n \rightarrow \overline{\mathbb{R}}$ messbare Funktionen. Dann heißt die Folge $(f_n)_{n \in \mathbb{N}}$ **stochastisch gegen 0 konvergent** (oder **konvergent gegen 0 nach Maßfolge $(P_n)_{n \in \mathbb{N}}$**), wenn gilt

$$\forall \epsilon > 0 : \quad P_n(\{|f_n| \geq \epsilon\}) \xrightarrow{n \rightarrow \infty} 0.$$

In Zeichen: $f_n \xrightarrow{P_n} 0$. In Analogie zur üblichen o_P -Notation schreiben wir $f_n = o_{P_n}(1)$, falls die Folge $(f_n)_{n \in \mathbb{N}}$ stochastisch gegen 0 konvergiert.

Sind P, Q Wahrscheinlichkeitsmaße auf einem Messraum (Ω, \mathcal{A}) , dann existiert nach dem Lebesgueschen Zerlegungssatz eine \mathcal{A} -messbare Funktion $f \geq 0$ und eine Q -Nullmenge $N \in \mathcal{A}$ mit

$$P(A) = \int_A f dQ + P(A \cap N), \quad A \in \mathcal{A}.$$

Die Funktion f und die Nullmenge N sind dabei Q -fast eindeutig bestimmt und damit ebenso der **Likelihood-Quotient**

$$\frac{dP}{dQ} := f.$$

Sei $(H, \langle \cdot, \cdot \rangle)$ ein separabler Hilbertraum. Der folgende Satz enthält die zentrale Charakterisierung einer schwach gegen einen Standard-Gaußshft konvergenten Experimentfolge; ein Beweis findet sich in Strasser (1985), S. 409.

Satz 1.36 Seien $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,h} \mid h \in H\}) \in \mathcal{E}(H)$ eine Experimentfolge und G_H der Standard-Gaußshift auf H . Dann sind die folgenden beiden Aussagen äquivalent:

1. $E_n \xrightarrow{w} G_H$,

2. Für die durch

$$\frac{dP_{n,h}}{dP_{n,0}} = \exp\left(L_{n,h} - \frac{1}{2}\|h\|^2\right) \quad \text{für } h \in H \text{ und } n \in \mathbb{N} \quad (1.11)$$

$P_{n,0}$ -fast eindeutig definierte Folge stochastischer Prozesse

$$(L_n)_{n \in \mathbb{N}} := (\mathcal{X}_n, \mathcal{A}_n, P_{n,0}, \{L_{n,h} \mid h \in H\})_{n \in \mathbb{N}}$$

gilt

(a) $\mathcal{L}(L_{n,h} \mid P_{n,0}) \xrightarrow{\mathcal{D}} N(0, \|h\|^2)$ für alle $h \in H$.

Mit $\xrightarrow{\mathcal{D}}$ bezeichnen wir dabei die übliche Verteilungskonvergenz (schwache Konvergenz) von Wahrscheinlichkeitsmaßen bzw. Zufallsvariablen.

(b) Für alle $\alpha, \beta \in \mathbb{R}$ und $h_1, h_2 \in H$ gilt $\alpha L_{n,h_1} + \beta L_{n,h_2} - L_{n,\alpha h_1 + \beta h_2} \xrightarrow{P_{n,0}} 0$.

Bemerkung 1.37 Durch (1.11) werden $\overline{\mathbb{R}}$ -wertige Zufallsvariablen

$$L_{n,h} = \log\left(\frac{dP_{n,h}}{dP_{n,0}}\right) + \frac{1}{2}\|h\|^2$$

definiert. $L_{n,h}$ nimmt dort den Wert $-\infty$ an, wo $P_{n,h}$ keine Masse bezüglich $P_{n,0}$ besitzt. Man beachte jedoch, dass die Experimentfolge $(E_n)_{n \in \mathbb{N}}$ asymptotisch homogen sein muss. Dies folgt einerseits bei gegebener schwacher Konvergenz aus der Homogenität des Grenzexperimentes, vgl. Bemerkung 1.34, und andererseits aus der Bedingung $\mathcal{L}(L_{n,h} \mid P_{n,0}) \xrightarrow{\mathcal{D}} N(0, \|h\|^2)$ von obigem Satz. Während in Strasser (1985) nicht im Detail auf diese Problematik eingegangen wird, wird der Sachverhalt in Witting und Müller-Funk (1995), S. 308 ff., ausführlich diskutiert.

Besonders einfach ist der Nachweis der schwachen Konvergenz, wenn man die Zufallsvariablen $L_{n,h}$ in der Form

$$L_{n,h} = X_{n,h} + o_{P_{n,0}}(1), \quad n \in \mathbb{N}, h \in H, \quad (1.12)$$

mit linearen Prozessen X_n , $n \in \mathbb{N}$, und stochastisch gegen 0 konvergenten Resten aufspalten kann. Ein Prozess $X = \{X_h \mid h \in H\}$ heißt **linear**, falls gilt

$$\alpha X_{h_1} + \beta X_{h_2} = X_{\alpha h_1 + \beta h_2} \quad \text{für alle } \alpha, \beta \in \mathbb{R} \text{ und } h_1, h_2 \in H.$$

Der Likelihood-Quotient aus (1.11) hat dann die Form

$$\frac{dP_{n,h}}{dP_{n,0}} = \exp\left(X_{n,h} - \frac{1}{2}\|h\|^2 + o_{P_{n,0}}(1)\right) \quad \text{für } h \in H \text{ und } n \in \mathbb{N}.$$

Korollar 1.38 Seien die Experimente E_n, G_H wie in Satz 1.36. Lassen sich die Zufallsvariablen L_n aus (1.11) wie in (1.12) in lineare Prozesse und stochastisch gegen 0 konvergente Reste zerlegen, dann sind die folgenden Aussagen äquivalent:

1. $E_n \xrightarrow{w} G_H$,
2. $\mathcal{L}(X_{n,h} | P_{n,0}) \xrightarrow{\mathcal{D}} N(0, \|h\|^2)$ für alle $h \in H$.

$(X_n)_{n \in \mathbb{N}}$ heißt in diesem Fall eine für $(E_n)_{n \in \mathbb{N}}$ **zentrale Folge**.

Bemerkung 1.39 Ein Standard-Gaußshift auf einem separablen Hilbertraum ist definitionsgemäß homogen. Aus der Darstellung der Hellingertransformierten (1.8), der Ungleichung (A.3) sowie Lemma A.11 folgt, dass der Standard-Gaußshift stetig ist. Damit muss nach Bemerkung 1.34 jede gegen einen solchen Gaußshift schwach konvergente Experimentfolge asymptotisch homogen und asymptotisch stetig sein.

1.8 Die Hájek-LeCam Minimax-Schranke und ihre Anwendungen

1.8.1 Begriffe und Bezeichnungen beim asymptotischen Testen

Seien $E = (\mathcal{X}, \mathcal{A}, \{P_\theta | \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ ein statistisches Experiment mit *metrischem Parameterraum* (Θ, d) und $\varphi : \mathcal{X} \rightarrow [0, 1]$ ein (*randomisierter*) Test mit zugehöriger Gütefunktion $g_\varphi : \Theta \rightarrow [0, 1]$, $g_\varphi(\theta) = \int \varphi dP_\theta$. Man beachte, dass die Gütefunktion auf ganz Θ definiert ist und im Folgenden auch auf ganz Θ betrachtet wird. Wir wollen eine **Hypothese H** gegen eine **Alternative K** testen. Das Testproblem (H, K) ist dabei eine Partition des metrischen Raumes Θ . Den Zusammenhang zur Entscheidungstheorie haben wir bereits in Abschnitt 1.1 erörtert.

Der Test φ heißt **unverfälscht** (für das Problem (H, K)), falls

$$\sup_{\theta \in H} g_\varphi(\theta) \leq \inf_{\theta \in K} g_\varphi(\theta)$$

ist. Jede Zahl $\alpha \in [\sup_{\theta \in H} g_\varphi(\theta), \inf_{\theta \in K} g_\varphi(\theta)]$ heißt dann **Niveau** des unverfälschten Tests φ .

Geben wir uns umgekehrt ein Niveau $\alpha \in [0, 1]$ vor, so heißt jeder Test φ mit

$$\sup_{\theta \in H} g_\varphi(\theta) \leq \alpha$$

ein **Test zum Niveau α** . Ein unverfälschter Test φ heißt **α -ähnlich** (für das Problem (H, K)), falls gilt

$$g_\varphi(\theta) = \alpha \quad \text{für alle } \theta \in \overline{H} \cap \overline{K}.$$

Mit $\mathcal{F}(\mathcal{X}, \mathcal{A})$ bezeichnen wir künftig die Menge aller Tests, mit $\mathcal{F}_\alpha(\mathbf{H}, \mathbf{K})$ die Klasse aller Tests zum Niveau α und mit $\mathcal{F}_\alpha^s(\mathbf{H}, \mathbf{K})$ die Klasse aller α -ähnlichen Tests. Zwei Tests mit derselben Gütefunktion heißen **äquivalent**.

Definition 1.40 Seien $\mathcal{F}_0 \subseteq \mathcal{F}(\mathcal{X}, \mathcal{A})$ eine Klasse von Tests, $\varphi^* \in \mathcal{F}_0$ und (H, K) ein Testproblem. Dann heißt

1. φ^* **zulässig** in \mathcal{F}_0 , falls kein Test $\varphi \in \mathcal{F}_0$ existiert mit

$$\begin{aligned} g_\varphi(\theta) &\leq g_{\varphi^*}(\theta) \quad \text{für alle } \theta \in H \\ g_\varphi(\theta) &\geq g_{\varphi^*}(\theta) \quad \text{für alle } \theta \in K, \end{aligned}$$

der nicht äquivalent zu φ^* ist, falls also kein Test φ aus \mathcal{F}_0 existiert, der gleichmäßig nicht schlechter und an mindestens einer Stelle echt besser als φ^* ist.

2. φ^* **optimal** in \mathcal{F}_0 , falls für alle Tests $\varphi \in \mathcal{F}_0$ gilt

$$\begin{aligned} g_{\varphi^*}(\theta) &\leq g_\varphi(\theta) \quad \text{für alle } \theta \in H \\ g_{\varphi^*}(\theta) &\geq g_\varphi(\theta) \quad \text{für alle } \theta \in K, \end{aligned}$$

d.h. φ^* ist gleichmäßig nicht schlechter als jeder andere Test in \mathcal{F}_0 .

Wir betrachten nun ein Folge von Experimenten $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,\theta} \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$, $n \in \mathbb{N}$, sowie ein Testproblem (H, K) , d.h. eine Partition von Θ . Eine Folge von Tests $(\varphi_n)_{n \in \mathbb{N}}$ mit $\varphi_n \in \mathcal{F}(\mathcal{X}_n, \mathcal{A}_n)$ und Gütefunktionen $g_{\varphi_n}(\theta) = \int_{\mathcal{X}_n} \varphi_n dP_{n,\theta}$, $\theta \in \Theta$, heißt **asymptotisch unverfälscht zum Niveau $\alpha \in [0, 1]$** , falls gilt

$$\begin{aligned} \limsup_{n \rightarrow \infty} g_{\varphi_n}(\theta) &\leq \alpha \quad \text{für alle } \theta \in H, \\ \liminf_{n \rightarrow \infty} g_{\varphi_n}(\theta) &\geq \alpha \quad \text{für alle } \theta \in K. \end{aligned}$$

Eine Folge von Tests $(\varphi_n)_{n \in \mathbb{N}}$ mit $\varphi_n \in \mathcal{F}(\mathcal{X}_n, \mathcal{A}_n)$ heißt **asymptotisch α -ähnlich**, falls gilt

$$\lim_{n \rightarrow \infty} g_{\varphi_n}(\theta) = \alpha \quad \text{für alle } \theta \in \overline{H} \cap \overline{K}.$$

Definition 1.41 Sei \mathcal{F}_0 jetzt eine Klasse von Testfolgen $(\varphi_n)_{n \in \mathbb{N}}$ mit $\varphi_n \in \mathcal{F}(\mathcal{X}_n, \mathcal{A}_n)$. Eine Testfolge $(\varphi_n^*)_{n \in \mathbb{N}} \in \mathcal{F}_0$ heißt **asymptotisch optimal** in der Klasse \mathcal{F}_0 , falls für alle Testfolgen $(\varphi_n)_{n \in \mathbb{N}} \in \mathcal{F}_0$ gilt

$$\begin{aligned} \limsup_{n \rightarrow \infty} g_{\varphi_n^*}(\theta) &\leq \liminf_{n \rightarrow \infty} g_{\varphi_n}(\theta) \quad \text{für alle } \theta \in H, \\ \liminf_{n \rightarrow \infty} g_{\varphi_n^*}(\theta) &\geq \limsup_{n \rightarrow \infty} g_{\varphi_n}(\theta) \quad \text{für alle } \theta \in K. \end{aligned} \tag{1.13}$$

Bemerkung 1.42 Definition 1.41 impliziert die Existenz des Grenzwertes $\lim_{n \rightarrow \infty} g_{\varphi_n^*}(\theta)$ für alle Parameter $\theta \in \Theta$. Bedingung (1.13) ist damit äquivalent zu

$$\begin{aligned} \lim_{n \rightarrow \infty} g_{\varphi_n^*}(\theta) &\leq \liminf_{n \rightarrow \infty} g_{\varphi_n}(\theta) \quad \text{für alle } \theta \in H, \\ \lim_{n \rightarrow \infty} g_{\varphi_n^*}(\theta) &\geq \limsup_{n \rightarrow \infty} g_{\varphi_n}(\theta) \quad \text{für alle } \theta \in K. \end{aligned}$$

1.8.2 Die Minimax-Schranke

Wir betrachten in diesem Abschnitt Experimente $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,\theta} \mid \theta \in \Theta\})$, die schwach gegen ein Experiment $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$ konvergieren, und versehen die Menge aller verallgemeinerten Entscheidungsfunktionen $\mathcal{R}_a(E, D)$ (D sei ein beliebiger topologischer Entscheidungsraum, vgl. Abschnitt 1.1) mit der Topologie der punktweisen Konvergenz auf $C_b(D) \times L(E)$. Bezüglich dieser Topologie ist die Menge $\mathcal{R}_a(E, D)$ konvex und kompakt, vgl. Strasser (1985), S. 231.

Jetzt wollen wir Folgen verallgemeinerter Entscheidungsfunktionen betrachten.

Definition 1.43 *Unter obigen Voraussetzungen an E_n, E seien $\beta_n \in \mathcal{R}_a(E_n, D)$, $n \in \mathbb{N}$, und $\beta \in \mathcal{R}_a(E, D)$ verallgemeinerte Entscheidungsfunktionen. Dann heißt*

1. $(\beta_n)_{n \in \mathbb{N}}$ **verteilungskonvergent** gegen β , falls gilt

$$\lim_{n \rightarrow \infty} \beta_n(f, P_{n,\theta}) = \beta(f, P_\theta) \quad \text{für alle } f \in C_b(D) \text{ und alle } \theta \in \Theta.$$

2. β ein **Häufungspunkt (nach Verteilung)** der Folge $(\beta_n)_{n \in \mathbb{N}}$, falls für jede endliche Menge $\alpha \in A(\Theta)$ und jede endliche Menge $C \subseteq C_b(D)$ eine Teilfolge $(\beta_{n_m})_{m \in \mathbb{N}}$ mit

$$\lim_{m \rightarrow \infty} \beta_{n_m}(f, P_{n_m,\theta}) = \beta(f, P_\theta) \quad \text{für alle } f \in C \text{ und alle } \theta \in \alpha$$

existiert.

Bemerkung 1.44 In LeCam (1979) wird nachgewiesen, dass jede Folge verallgemeinerter Entscheidungsfunktionen auch Häufungspunkte (nach Verteilung) hat.

Bemerkung 1.45 Wir betrachten speziell ein Testproblem (H, K) . Zu randomisierten Tests $\varphi_n \in \mathcal{F}(\mathcal{X}_n, \mathcal{A}_n)$ und $\varphi \in \mathcal{F}(\mathcal{X}, \mathcal{A})$ existieren nach Abschnitt 1.1 korrespondierende verallgemeinerte Entscheidungsfunktionen β_{φ_n} und β_φ . Man erkennt aus deren Konstruktion sofort, dass aus der Verteilungskonvergenz der β_{φ_n} gegen β_φ die punktweise Konvergenz

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X}_n} \varphi_n(x) P_{n,\theta}(dx) = \lim_{n \rightarrow \infty} g_{\varphi_n}(\theta) = g_\varphi(\theta) = \int_{\mathcal{X}} \varphi(x) P_\theta(dx) \quad \text{für } \theta \in \Theta$$

der Gütefunktionen folgt.

Jetzt werden wir die in der Theorie schwacher Konvergenz so wichtige asymptotische Minimax-Schranke von Hájek-LeCam anführen. Die folgende Formulierung findet sich in Strasser (1985), S. 310 f. und ist eine Verallgemeinerung der Formulierung LeCam's.

Theorem 1.46 (Hájek-LeCam Minimax-Schranke) *Wir betrachten für die obigen Experimente $E_n \xrightarrow{w} E$ ein Entscheidungspaar (D, W) mit von unten halbstetiger Verlustfunktion W . Die Folge verallgemeinerter Entscheidungsfunktionen $(\beta_n)_{n \in \mathbb{N}}$ mit $\beta_n \in \mathcal{R}_a(E_n, D)$ besitze die nichtleere (vgl. Bemerkung 1.44) Menge $\mathcal{R}_0 \subseteq \mathcal{R}_a(E, D)$ von Häufungspunkten (nach Verteilung). Mit $\overline{co(\mathcal{R}_0)}$ bezeichnen wir die gemäß obiger Topologie der punktweisen Konvergenz auf $C_b(D) \times L(E)$ abgeschlossene konvexe Hülle von \mathcal{R}_0 . Dann gilt für jede Menge $\Theta_0 \subseteq \Theta$ die Abschätzung*

$$\liminf_{n \rightarrow \infty} \sup_{\theta \in \Theta_0} \beta_n(W_\theta, P_{n,\theta}) \geq \inf_{\beta \in \overline{co(\mathcal{R}_0)}} \sup_{\theta \in \Theta_0} \beta(W_\theta, P_\theta).$$

Bemerkung 1.47 Da $\mathcal{R}_a(E, D)$ wie zu Beginn von 1.8.2 erwähnt bezüglich der Topologie der punktweisen Konvergenz konvex und kompakt ist, gilt $\overline{co(\mathcal{R}_0)} \subseteq \mathcal{R}_a(E, D)$.

Eine etwas schwächere Variante von Theorem 1.46 ist

Korollar 1.48 (zur Hájek-LeCam Minimax-Schranke 1.46)

Sind $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,\theta} \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ und $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ Experimente mit $E_n \xrightarrow{w} E$, und ist (D, W) ein Entscheidungspaar mit von unten halbstetiger Verlustfunktion W , so gilt für alle Folgen verallgemeinerter Entscheidungsfunktionen $\beta_n \in \mathcal{R}_a(E_n, D)$

$$\liminf_{n \rightarrow \infty} \sup_{\theta \in \Theta_0} \beta_n(W_\theta, P_{n,\theta}) \geq \inf_{\beta \in \mathcal{R}_a(E, D)} \sup_{\theta \in \Theta_0} \beta(W_\theta, P_\theta).$$

für jede Menge $\Theta_0 \subseteq \Theta$.

1.8.3 Eine Anwendung der Minimax-Schranke für das asymptotische Testen

Seien $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,h} \mid h \in H\}) \in \mathcal{E}(H)$, $n \in \mathbb{N}$, Experimente, die schwach gegen den Standard-Gaußshift G_H auf einem separablen Hilbertraum $(H, \langle \cdot, \cdot \rangle)$ konvergieren. Für die Folge $(E_n)_{n \in \mathbb{N}}$ möge gemäß Korollar 1.38 eine zentrale Folge linearer Prozesse $(X_n)_{n \in \mathbb{N}}$ existieren. Wir untersuchen das spezielle Testproblem

$$H_\ell = \{h \in H \mid \ell(h) \leq 0\}, \quad K_\ell = \{h \in H \mid \ell(h) > 0\}$$

für eine fest vorgegebene stetige Linearform $\ell : H \rightarrow \mathbb{R}$. Gemäß dem Darstellungssatz von Frechét-Riesz gibt es ein Element $h_\ell \in H$ mit

$$\ell(h) = \langle h, h_\ell \rangle \quad \text{und} \quad \|\ell\| = \|h_\ell\|.$$

Eine der vielen Folgerungen aus der Minimax-Schranke 1.46 bei vorliegender schwacher Konvergenz ist

Satz 1.49 (vgl. **Strasser (1985), S. 431**) *Unter den gerade beschriebenen Voraussetzungen ist die Testfolge $(\varphi_n^*)_{n \in \mathbb{N}}$ mit $\varphi_n^* : \mathcal{X}_n \rightarrow [0, 1]$,*

$$\varphi_n^*(x_n) = \begin{cases} 1, & X_{n,h_\ell}(x_n) > \|\ell\| \cdot \Phi^{-1}(1 - \alpha) \\ 0, & X_{n,h_\ell}(x_n) < \|\ell\| \cdot \Phi^{-1}(1 - \alpha) \end{cases}$$

asymptotisch optimal in der Klasse aller asymptotisch α -ähnlichen Testfolgen im Sinne von Definition 1.41 bzw. Bemerkung 1.42. (Im asymptotisch nicht relevanten Fall $X_{n,h_\ell}(x_n) = \|\ell\| \cdot \Phi^{-1}(1 - \alpha)$ können die randomisierten Tests φ_n^ beliebig definiert werden.) Für die Folge $(g_{\varphi_n^*})_{n \in \mathbb{N}}$ der Gütefunktionen gilt*

$$\lim_{n \rightarrow \infty} g_{\varphi_n^*}(h) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\ell(h)}{\|\ell\|} \right) \quad \text{für alle } h \in H.$$

Diesen Satz wollen wir in Hinblick auf spätere Anwendungen verallgemeinern, in denen keine schwache Konvergenz gegen einen „vollständigen“ Standard-Gaußshift G_H vorliegt, sondern nur die auf eine Teilmenge $\Theta \subseteq H$ eingeschränkten Experimente $E_{n,\Theta}$ schwach gegen die Einschränkung $G_{H,\Theta}$ konvergieren, also

$$E_{n,\Theta} \xrightarrow{w} G_{H,\Theta}$$

gilt. Desweiteren sei $0 \in \Theta$, und es soll ein Analogon zur zentralen Folge existieren, d.h. Zufallsvariablen $X_{n,h} : \mathcal{X}_n \rightarrow \mathbb{R}$, $h \in H$, $n \in \mathbb{N}$ mit

$$\frac{dP_{n,h}}{dP_{n,0}} = \exp \left(X_{n,h} - \frac{1}{2} \|h\|^2 + o_{P_{n,0}}(1) \right), \quad h \in \Theta, \quad (1.14)$$

$$\mathcal{L}(X_{n,h} \mid P_{n,0}) \xrightarrow{\mathcal{D}} N(0, \|h\|^2), \quad h \in \Theta, \quad (1.15)$$

und

$$\alpha X_{n,h_1} + \beta X_{n,h_2} = X_{n,\alpha h_1 + \beta h_2}, \quad \alpha, \beta \in \mathbb{R}, \quad h_1, h_2 \in H. \quad (1.16)$$

Man beachte, dass die Zufallsvariablen $X_{n,h}$ für alle Elemente des Hilbertraums H definiert und die Prozesse $\{X_{n,h} \mid h \in H\}$ linear sein sollen, wohingegen die Konvergenzbedingungen (1.14) und (1.15) nur für die Elemente der Teilmenge Θ vorausgesetzt werden. In Kapitel 2 werden wir sehen, dass diese Voraussetzungen natürlich sind und in praxisrelevanten Fällen auftreten.

Es stellt sich die Frage, ob die Testfolgen aus Satz 1.49 auch in dieser neuen Situation asymptotisch optimal sind. Im Allgemeinen wird dies nicht mehr der Fall sein, sondern wir müssen zusätzliche Voraussetzungen an das Testproblem, also an die Linearform ℓ , stellen, damit die Optimalität erhalten bleibt. Um solche Voraussetzungen angeben zu können, werden wir im Folgenden den Beweis von Satz 1.49 analysieren. Grundlage für die Optimalität der Testfolge aus Satz 1.49 war die Minimax-Schranke von Hájek-LeCam 1.46, mit deren Hilfe auch das folgende, auf unsere neue Situation angepasste Lemma bewiesen wird.

Lemma 1.50 Seien $\Theta \neq \emptyset$ eine beliebige Teilmenge des separablen Hilbertraums $(H, \langle \cdot, \cdot \rangle)$ und $E_{n,\Theta} \in \mathcal{E}(\Theta)$ Experimente mit $E_{n,\Theta} \xrightarrow{w} G_{H,\Theta}$. Weiter seien $\ell : H \rightarrow \mathbb{R}$ eine beliebige stetige Linearform sowie

$$H_\ell = \{h \in \Theta \mid \ell(h) \leq 0\}, \quad K_\ell = \{h \in \Theta \mid \ell(h) > 0\}$$

das damit verbundene Testproblem. Für (H_ℓ, K_ℓ) gelte $\overline{H_\ell} \cap \overline{K_\ell} \neq \emptyset$, damit der Begriff „asymptotisch α -ähnlich“ sinnvoll definiert ist. Dann gilt für beliebiges Niveau $\alpha \in (0, 1)$ und jede asymptotisch α -ähnliche Testfolge $(\psi_n)_{n \in \mathbb{N}}$ mit $\psi_n : \mathcal{X}_n \rightarrow [0, 1]$

$$\begin{aligned} \liminf_{n \rightarrow \infty} g_{\psi_n}(h) &\geq \Phi\left(\Phi^{-1}(\alpha) + \frac{\ell(h)}{\|\ell\|}\right) && \text{für } h \in H_\ell, \\ \limsup_{n \rightarrow \infty} g_{\psi_n}(h) &\leq \Phi\left(\Phi^{-1}(\alpha) + \frac{\ell(h)}{\|\ell\|}\right) && \text{für } h \in K_\ell. \end{aligned}$$

Beweis. Der Beweis verläuft völlig analog zu Lemma 82.6 aus Strasser (1985), wo die Aussage für den Fall bewiesen wird, dass keine Einschränkung des Parameterraums vorliegt. \square

Ist $\ell : H \rightarrow \mathbb{R}$ eine gegebene Linearform mit zugehörigem Hilbertraumelement h_ℓ gemäß dem Satz von Frechét-Riesz, so ist der nächste wesentliche Schritt im Beweis von Satz 1.49 die asymptotische Untersuchung der X_{n,h_ℓ} bezüglich der Maße $P_{n,h}$, $h \in \Theta$. Im Falle $\Theta = H$ kann man allein aus (1.15) und (1.16) wie im folgenden Satz auf die schwache Konvergenz

$$\mathcal{L}(X_{n,h_\ell} \mid P_{n,h}) \xrightarrow{\mathcal{D}} N(\langle h, h_\ell \rangle, \|h_\ell\|^2)$$

schließen. Gilt (1.15) nicht für alle $h \in H$, so benötigen wir zusätzliche Voraussetzungen.

Satz 1.51 Seien $\Theta \neq \emptyset$ eine beliebige Teilmenge des separablen Hilbertraums $(H, \langle \cdot, \cdot \rangle)$ und $E_{n,\Theta} \in \mathcal{E}(\Theta)$ Experimente mit $E_{n,\Theta} \xrightarrow{w} G_{H,\Theta}$. Ferner existiere ein Analogon zur zentralen Folge, also Zufallsvariablen $X_{n,h}$, $h \in H$, die (1.14), (1.15) und (1.16) erfüllen. Unter der zusätzlichen Voraussetzung an die Linearform ℓ mit zugehörigem h_ℓ

$$\mathcal{L}(X_{n,\alpha h_\ell + \beta h} \mid P_{n,0}) \xrightarrow{\mathcal{D}} N(0, \|\alpha h_\ell + \beta h\|^2) \quad \text{für alle } \alpha, \beta \in \mathbb{R}, h \in \Theta, \quad (1.17)$$

gilt dann die schwache Konvergenz

$$\mathcal{L}(X_{n,h_\ell} \mid P_{n,h}) \xrightarrow{\mathcal{D}} N(\langle h, h_\ell \rangle, \|h_\ell\|^2) \quad \text{für alle } h \in \Theta. \quad (1.18)$$

Beweis. Der Beweis orientiert sich an Strasser (1985), Lemma 80.10, Lemma 80.11 und Theorem 80.12. Seien dazu $\alpha, \beta \in \mathbb{R}$ und $h \in \Theta$ beliebig. Dann folgt aus (1.17) mittels der Technik von Cramér-Wold, vgl. Bauer (1991), S. 268, und unter Beachtung der Linearitätseigenschaft (1.16) die schwache Konvergenz

$$\mathcal{L}\left((X_{n,h}, X_{n,h_\ell})^\top \mid P_{n,0}\right) \xrightarrow{\mathcal{D}} N(0, C) \quad \text{mit} \quad C := \begin{pmatrix} \|h\|^2 & \langle h, h_\ell \rangle \\ \langle h, h_\ell \rangle & \|h_\ell\|^2 \end{pmatrix}. \quad (1.19)$$

Wir gehen zunächst davon aus, dass C regulär ist. Sei $u \in C_c(\mathbb{R}^2)$ eine stetige Funktion mit kompaktem Träger. Dann folgt mit (1.14) und (1.19) für $h \in \Theta$

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \int u(X_{n,h}, X_{n,h_\ell}) dP_{n,h} \\
&= \lim_{n \rightarrow \infty} \int u(X_{n,h}, X_{n,h_\ell}) \frac{dP_{n,h}}{dP_{n,0}} dP_{n,0} \\
&\stackrel{(1.14)}{=} \lim_{n \rightarrow \infty} \int u(X_{n,h}, X_{n,h_\ell}) \exp \left\{ X_{n,h} - \frac{1}{2} \|h\|^2 + o_{P_{n,0}}(1) \right\} dP_{n,0} \\
&\stackrel{(1.19)}{=} \int u(x, y) \exp \left\{ x - \frac{1}{2} \|h\|^2 \right\} N(0, C)(dx, dy) \\
&= \int u(x, y) N \left(\begin{pmatrix} \|h\|^2 \\ \langle h, h_\ell \rangle \end{pmatrix}, C \right) (dx, dy), \tag{1.20}
\end{aligned}$$

wobei man (1.20) wegen der Regularität von C leicht mittels Lebesgue-Dichten einsieht. Beachtet man, dass in lokal-kompakten Räumen die vage und die schwache Konvergenz von Wahrscheinlichkeitsmaßen übereinstimmen, vgl. Bauer (1992), Korollar 30.9, so genügt es, Funktionen u mit kompaktem Träger zu betrachten, und es folgt die schwache Konvergenz

$$\mathcal{L}((X_{n,h}, X_{n,h_\ell})^\top | P_{n,h}) \xrightarrow{\mathcal{D}} N \left(\begin{pmatrix} \|h\|^2 \\ \langle h, h_\ell \rangle \end{pmatrix}, C \right)$$

und daraus dann

$$\mathcal{L}(X_{n,h_\ell} | P_{n,h}) \xrightarrow{\mathcal{D}} N(\langle h, h_\ell \rangle, \|h_\ell\|^2).$$

Ist C nicht regulär, d.h. h und h_ℓ sind linear abhängig, so beweist man die schwache Konvergenz (1.18) durch Nachweis von

$$\lim_{n \rightarrow \infty} \int u(X_{n,h_\ell}) dP_{n,h} = \int u(x) N(\langle h, h_\ell \rangle, \|h_\ell\|^2)(dx)$$

für $u \in C_c(\mathbb{R})$, wobei man wie oben vorgeht und zusätzlich die Linearitätseigenschaft (1.16) verwendet. \square

Mit diesem Ergebnis können wir eine Verallgemeinerung von Satz 1.49 beweisen.

Satz 1.52 *Seien $\Theta \neq \emptyset$ eine beliebige Teilmenge des separablen Hilbertraums $(H, \langle \cdot, \cdot \rangle)$ und $E_{n,\Theta} \in \mathcal{E}(\Theta)$ Experimente mit $E_{n,\Theta} \xrightarrow{w} G_{H,\Theta}$. Ferner existiere ein Analogon zur zentralen Folge, also Zufallsvariablen $X_{n,h}$, $h \in H$, die (1.14), (1.15) und (1.16) erfüllen, und für die Linearform ℓ mit zugehörigem h_ℓ gelte*

$$\mathcal{L}(X_{n,\alpha h_\ell + \beta h} | P_{n,0}) \xrightarrow{\mathcal{D}} N(0, \|\alpha h_\ell + \beta h\|^2) \quad \text{für alle } \alpha, \beta \in \mathbb{R}, h \in \Theta. \tag{1.21}$$

Dann ist für das Testproblem

$$H_\ell = \{h \in \Theta \mid \ell(h) \leq 0\}, \quad K_\ell = \{h \in \Theta \mid \ell(h) > 0\}$$

die Testfolge $(\varphi_n^*)_{n \in \mathbb{N}}$ mit $\varphi_n^* : \mathcal{X}_n \rightarrow [0, 1]$,

$$\varphi_n^*(x_n) = \begin{cases} 1, & X_{n, h_\ell}(x_n) > \|\ell\| \cdot \Phi^{-1}(1 - \alpha) \\ 0, & X_{n, h_\ell}(x_n) < \|\ell\| \cdot \Phi^{-1}(1 - \alpha) \end{cases}$$

asymptotisch optimal in der Klasse aller asymptotisch α -ähnlichen Testfolgen im Sinne von Definition 1.41 bzw. Bemerkung 1.42. Dabei setzen wir $\overline{H}_\ell \cap \overline{K}_\ell \neq \emptyset$ voraus. Für die Folge $(g_{\varphi_n^*})_{n \in \mathbb{N}}$ der Gütefunktionen gilt

$$\lim_{n \rightarrow \infty} g_{\varphi_n^*}(h) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\ell(h)}{\|\ell\|} \right) \quad \text{für alle } h \in \Theta.$$

Beweis. Für die Grenzgüte der gegebenen Testfolge gilt nach Satz 1.51

$$\lim_{n \rightarrow \infty} g_{\varphi_n^*}(h) = \lim_{n \rightarrow \infty} P_{n, h} (X_{n, h_\ell} > \|\ell\| \cdot \Phi^{-1}(1 - \alpha)) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\ell(h)}{\|\ell\|} \right)$$

für $h \in \Theta$. Die Optimalität der Testfolge folgt aus Lemma 1.50. \square

1.9 Der Satz von Lindae

Geben wir uns eine schwach konvergente Experimentfolge vor, so stellt sich die Frage, unter welchen zusätzlichen Voraussetzungen diese Experimentfolge auch stark konvergiert, d.h. in der Δ -Metrik. Eine Antwort liefert der Satz von Lindae, vgl. Lindae (1972), LeCam (1986), S. 92, oder Torgersen (1991), S. 396. Zunächst benötigen wir einige topologische Begriffe.

Eine Teilmenge M eines metrischen Raums S heißt **präkompakt** oder **totalbeschränkt**, wenn M für jedes $\epsilon > 0$ durch endlich viele Kugeln vom Radius ϵ überdeckt werden kann. Das ist gleichbedeutend damit, dass der Abschluss von M in der Vervollständigung von S kompakt ist. Für vollständige metrische Räume stimmen die Begriffe *präkompakt* und *relativ kompakt* somit überein.

Entscheidend für den Satz von Lindae ist der L-Raum $L(E)$ eines Experiments E . In Abschnitt 1.1 haben wir gesehen, dass $L(E)$ mit der Variationsnorm zum Banachraum wird.

Satz 1.53 (Lindae)

Seien $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n, \theta} \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ und $E = (\mathcal{X}, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$ Experimente mit $E_n \xrightarrow{w} E$. Weiter sei

1. die Menge $\{P_\theta \mid \theta \in \Theta\}$ totalbeschränkt in $(L(E), \|\cdot\|_V)$, und

2. es gelte die Konvergenz

$$\|P_{n,\theta_1} - P_{n,\theta_2}\|_V \xrightarrow{n \rightarrow \infty} \|P_{\theta_1} - P_{\theta_2}\|_V \quad \text{gleichmäßig in } \theta_1, \theta_2 \in \Theta. \quad (1.22)$$

Dann folgt die starke Konvergenz

$$\Delta(E_n, E) \xrightarrow{n \rightarrow \infty} 0.$$

Bemerkung 1.54 Wegen Ungleichung (A.3) können wir bei der Überprüfung der Voraussetzungen des Satzes von Lindae an Stelle des Variationsabstandes auch den Hellingerabstand verwenden.

Das folgende Lemma zeigt eine Möglichkeit, die Totalbeschränktheitsvoraussetzung im Satz von Lindae nachzuweisen, wenn als Grenzexperiment ein Standard-Gaußshift auf einem Hilbertraum gegeben ist.

Lemma 1.55 (Totalbeschränktheit bei Gaußshifts auf Hilberträumen)

Sei $G_H = (\mathcal{X}, \mathcal{A}, \{P_h \mid h \in H\})$ der Standard-Gaußshift auf einem Hilbertraum $(H, \langle \cdot, \cdot \rangle)$. Seien $H_0 \subset H$ und $G_0 = (\mathcal{X}, \mathcal{A}, \{P_h \mid h \in H_0\})$ die Einschränkung von G_H auf H_0 . Dann gilt

$$\{P_h \mid h \in H_0\} \text{ totalbeschränkt in } (L(G_0), \|\cdot\|_V) \iff H_0 \text{ totalbeschränkt in } H.$$

Beweis. Seien $h_1, h_2 \in H_0$ und $1 > \epsilon > 0$ beliebig. Dann gilt nach Formel (1.9) auf Seite 20 und Lemma A.11 für den Hellingerabstand von P_{h_1} und P_{h_2}

$$d_H^2(P_{h_1}, P_{h_2}) = 1 - \exp \left\{ -\frac{1}{8} \|h_1 - h_2\|^2 \right\},$$

und es ist

$$d_H^2(P_{h_1}, P_{h_2}) < \epsilon \iff \|h_1 - h_2\|^2 < -8 \ln(1 - \epsilon).$$

Jetzt folgt die Behauptung aus der Definition der Totalbeschränktheit zusammen mit Bemerkung 1.54. \square

1.10 Asymptotische Äquivalenz

In vielen Fällen besitzt eine Experimentfolge $(E_n)_{n \in \mathbb{N}}$ kein oder nur ein entartetes Grenzexperiment im Sinne von LeCam's Δ -Metrik. Das gleiche trifft für die schwache Konvergenz im Sinne LeCam's zu. Deshalb ist es mitunter geschickter, *begleitende einfache* Experimentfolgen zu $(E_n)_{n \in \mathbb{N}}$ anzugeben, die den Experimenten E_n für hinreichend großes $n \in \mathbb{N}$ *ähnlich* sind. Wir wollen diese Ähnlichkeit konkretisieren.

Definition 1.56 (Asymptotische Äquivalenz) Es seien $(E_n)_{n \in \mathbb{N}}, (F_n)_{n \in \mathbb{N}} \subset \mathcal{E}(\Theta)$ zwei Experimentfolgen. Die beiden Folgen heißen **(stark) asymptotisch äquivalent**, wenn gilt

$$\lim_{n \rightarrow \infty} \Delta(E_n, F_n) = 0.$$

Definition 1.57 (Schwache asymptotische Äquivalenz) Gegeben seien zwei Folgen von Experimenten $(E_n)_{n \in \mathbb{N}}, (F_n)_{n \in \mathbb{N}} \subset \mathcal{E}(\Theta)$. Die beiden Folgen heißen **schwach asymptotisch äquivalent**, wenn gilt

$$\lim_{n \rightarrow \infty} \Delta(E_{n,\alpha}, F_{n,\alpha}) = 0 \quad \text{für alle } \alpha \in A(\Theta).$$

Die folgende Charakterisierung der schwachen asymptotischen Äquivalenz findet sich in Becker (1983), S. 20.

Satz 1.58 Zwei Experimentfolgen $(E_n)_{n \in \mathbb{N}}, (F_n)_{n \in \mathbb{N}} \subset \mathcal{E}(\Theta)$ sind genau dann schwach asymptotisch äquivalent, wenn die Differenz der Hellingertransformierten punktweise gegen 0 konvergiert:

$$\lim_{n \rightarrow \infty} H(E_{n,\alpha})(z) - H(F_{n,\alpha})(z) = 0 \quad \text{für alle } \alpha \in A(\Theta) \text{ und alle } z \in S_\alpha.$$

Bemerkung 1.59 Seien $(E_n)_{n \in \mathbb{N}}, (F_n)_{n \in \mathbb{N}} \subset \mathcal{E}(\Theta)$ zwei Experimentfolgen und $E \in \mathcal{E}(\Theta)$. Dann folgt aus den Sätzen 1.32 und 1.58 sofort:

1. Sind $(E_n)_{n \in \mathbb{N}}$ und $(F_n)_{n \in \mathbb{N}}$ schwach asymptotisch äquivalent und gilt $E_n \xrightarrow{w} E$, so gilt auch $F_n \xrightarrow{w} E$.
2. Gilt $E_n \xrightarrow{w} E$ und $F_n \xrightarrow{w} E$, so sind $(E_n)_{n \in \mathbb{N}}$ und $(F_n)_{n \in \mathbb{N}}$ schwach asymptotisch äquivalent.

Kapitel 2

Asymptotik von Regressionsmodellen bei unabhängigen Beobachtungen

Dieses Kapitel widmet sich ausschließlich der Asymptotik von Regressionsmodellen. Bei vorgegebener Folge von Versuchsplänen werden wir der zugehörigen Folge von Regressionsmodellen ein geeignetes Grenzmodell zuordnen bzw. eine geeignete Folge von Grenzmodellen. Diese werden wir auf verschiedene Art und Weise erhalten:

1. Durch Einbettung in den Raum stetiger Funktionen mit Hilfe des Partialsummenprozesses. Mit Hilfe des Satzes von Donsker erhält man dann ein Grenzmodell. Man vergleiche dazu vor allem die Arbeiten Bischoff (1998), Bischoff und Miller (2000a) sowie Bischoff und Miller (2000b).
2. Durch schwache Konvergenz im Sinne LeCam's, wobei wir im Parameterraum lokalisieren müssen, um nichttriviale Grenzwerte zu erhalten. Die Wahl eines geeigneten Lokalisierungspunktes hängt von der interessierenden Fragestellung ab. Diese Konvergenzart entspricht im Wesentlichen dem Konzept der "Lokalen Asymptotischen Normalität" (LAN), das in der Literatur eine große Rolle spielt. Stellvertretend für die zahlreiche Literatur zur Theorie und Anwendung dieses Konzeptes seien die Bücher LeCam (1986), Strasser (1985) sowie Witting und Müller-Funk (1995) genannt.
3. Im Sinne der starken Konvergenz bzw. der starken asymptotischen Äquivalenz LeCam's. Hierzu gibt es auf Grund großer technischer Schwierigkeiten erst sehr wenige Publikationen, von denen vor allem die Arbeiten Brown und Low (1996), Nussbaum (1996), Grama und Nussbaum (1998) sowie Milstein und Nussbaum (1998) hervorzuheben sind.

Grob gesprochen kann man sagen: Je stärker die Konvergenzart ist, desto mehr Information enthält das Grenzmodell und desto restriktiver müssen auch die Voraussetzungen sein, um Konvergenz nachweisen zu können.

2.1 Modellbeschreibung

Zunächst soll geklärt werden, was wir in der vorliegenden Arbeit unter einem Regressionsmodell verstehen. Dazu geben wir uns als **Fehlerverteilung** ein *bekanntes Wahrscheinlichkeitsmaß* Q auf $(\mathbb{R}, \mathcal{B})$ vor. Mit

$$Q_s(\cdot) := Q(\cdot - s), \quad s \in \mathbb{R}, \quad (2.1)$$

bezeichnen wir das *um* $s \in \mathbb{R}$ *nach rechts geshiftete* Maß Q und mit

$$E_Q := (\mathbb{R}, \mathcal{B}, \{Q_s \mid s \in \mathbb{R}\}) \quad (2.2)$$

das **zu** Q **gehörige Shiftexperiment**.

Weiter seien der **Versuchsbereich** $\mathcal{V} = [a, b] \subseteq \mathbb{R}$ sowie eine *bekannte Funktion* $\sigma : [a, b] \rightarrow (0, \infty)$ gegeben, wobei σ^2 im Folgenden als **Varianzfunktion** bezeichnet wird. Es seien (t_{ni}) , $i = 1, \dots, n$, $n \in \mathbb{N}$, ein **Dreiecksschema von Versuchspunkten** mit

$$a \leq t_{n1} \leq t_{n2} \leq \dots \leq t_{nn} \leq b$$

und

ϵ_{ni} , $i = 1, \dots, n$, stochastisch unabhängige Zufallsvariablen mit Verteilung Q .

Wir beobachten die Zufallsvariablen

$$Y_{ni} = f(t_{ni}) + \sigma(t_{ni})\epsilon_{ni} \quad i = 1, \dots, n, \quad (2.3)$$

und interessieren uns für die *unbekannte Regressionsfunktion* $f \in \mathcal{F}$, wobei

$$\mathcal{F} \subseteq \{f : [a, b] \rightarrow \mathbb{R}\}$$

die *Klasse der möglichen Regressoren*, also der **Parameterraum** sei. Mit den Abkürzungen

$$\begin{aligned} t_n &:= (t_{n1}, \dots, t_{nn})^\top, \\ f(t_n) &:= (f(t_{n1}), \dots, f(t_{nn}))^\top, \\ \epsilon_n &:= (\epsilon_{n1}, \dots, \epsilon_{nn})^\top, \\ Y_n &:= (Y_{n1}, \dots, Y_{nn})^\top, \\ \Sigma_n &:= \begin{pmatrix} \sigma(t_{n1}) & & 0 \\ & \ddots & \\ 0 & & \sigma(t_{nn}) \end{pmatrix} \end{aligned}$$

lässt sich (2.3) schreiben als

$$Y_n = f(t_n) + \Sigma_n \epsilon_n, \quad (2.4)$$

und es gilt

$$Y_{ni} \sim Q \left(\frac{\cdot - f(t_{ni})}{\sigma(t_{ni})} \right), \quad i = 1, \dots, n.$$

Für festes $i \in \{1, \dots, n\}$ lässt sich das Beobachten der Zufallsvariablen Y_{ni} damit durch das statistische Experiment

$$F_{ni} = \left(\mathbb{R}, \mathcal{B}, \left\{ Q \left(\frac{\cdot - f(t_{ni})}{\sigma(t_{ni})} \right) \mid f \in \mathcal{F} \right\} \right)$$

beschreiben. Da auf Grund der stochastischen Unabhängigkeit der Y_{ni}

$$Y_n \sim \bigotimes_{i=1}^n Q \left(\frac{\cdot - f(t_{ni})}{\sigma(t_{ni})} \right)$$

gilt, lässt sich das Beobachten der Zufallsvariablen Y_n durch das statistische Experiment

$$F_n = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q \left(\frac{\cdot - f(t_{ni})}{\sigma(t_{ni})} \right) \mid f \in \mathcal{F} \right\} \right) \quad (2.5)$$

beschreiben. Somit haben wir ein Dreiecksschema (F_{ni}) , $i = 1, \dots, n$, $n \in \mathbb{N}$, von Experimenten mit

$$F_n =: \bigotimes_{i=1}^n F_{ni}, \quad n \in \mathbb{N},$$

das die Folge der Regressionsmodelle beschreibt.

Im weiteren Verlauf wird es sich als nützlich erweisen, die Varianzfunktion σ^2 suffizient „in die Regressionsfunktion hineinzutransformieren“. Wegen $\sigma > 0$ auf ganz \mathcal{V} sind die Abbildungen

$$T_n : \begin{cases} \mathbb{R}^n & \rightarrow \mathbb{R}^n \\ x & \mapsto \Sigma_n^{-1} x \end{cases}$$

bijektiv und somit nach Witting (1985), Satz 3.10, suffiziente Statistiken für die Experimente F_n . Bei Anwendung dieser Statistiken auf die Y_n erhalten wir

$$Z_n = \Sigma_n^{-1} f(t_n) + \epsilon_n =: (Z_{n1}, \dots, Z_{nn})^\top,$$

und mit der Schreibweise (2.1) gilt für die Verteilung der Zufallsvariablen Z_{ni}

$$Z_{ni} \sim Q \left(\cdot - \frac{f(t_{ni})}{\sigma(t_{ni})} \right) = Q_{\frac{f(t_{ni})}{\sigma(t_{ni})}}, \quad i = 1, \dots, n,$$

was auf die Experimente

$$E_n = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f(t_{ni})}{\sigma(t_{ni})}} \mid f \in \mathcal{F} \right\} \right) \quad (2.6)$$

führt. Nach Satz 1.10 sind die Experimente E_n und F_n im Sinne LeCams äquivalent, d.h. es gilt

$$\Delta(E_n, F_n) = 0, \quad n \in \mathbb{N}.$$

Wenn nicht ausdrücklich darauf hingewiesen wird, werden wir im Folgenden nicht mehr zwischen den Experimenten E_n und F_n unterscheiden.

Die Experimente E_n bezeichnen wir im weiteren Verlauf der vorliegenden Arbeit als **allgemeine Regressionsmodelle**. Einen Spezialfall der allgemeinen Regressionsmodelle stellen die Linearen Regressionsmodelle dar: Sind f_1, \dots, f_k bekannte Funktionen, und ist der Parameterraum \mathcal{F} gleich der linearen Hülle der f_1, \dots, f_k , also

$$\mathcal{F} := [f_1, \dots, f_k],$$

so nennen wir die Experimente E_n aus (2.6) **Lineare Regressionsmodelle**. Mittels der Abkürzung

$$f := (f_1, \dots, f_k)^\top$$

können wir E_n auch in der Form

$$E_n = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})}} \mid \beta \in \mathbb{R}^k \right\} \right) \quad (2.7)$$

schreiben.

Einen wichtigen Vertreter der Linearen Regressionsmodelle führen wir im folgenden Beispiel an.

Beispiel 2.1 Ist $\mathcal{F} = [f_1, \dots, f_k]$, ist also der Parameterraum gleich der linearen Hülle der bekannten Funktionen f_1, \dots, f_k , $k \in \mathbb{N}$, und ist $Q = N(0, 1)$, dann ist

$$F_n = \mathcal{N}(A_n, \Sigma_n^2)$$

mit Modellmatrix

$$A_n^\top = \begin{pmatrix} f_1(t_{n1}) & \cdots & f_k(t_{n1}) \\ \vdots & \ddots & \vdots \\ f_1(t_{nn}) & \cdots & f_k(t_{nn}) \end{pmatrix}, \quad (2.8)$$

Covarianzmatrix

$$\Sigma_n^2 = \begin{pmatrix} \sigma^2(t_{n1}) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(t_{nn}) \end{pmatrix} \quad (2.9)$$

und Parameterraum \mathbb{R}^k (man vergleiche Notation (1.5) auf Seite 16 und wähle dort als Beobachtungsraum $\mathcal{X} = \mathbb{R}^n$ versehen mit dem kanonischen Skalarprodukt). Durch die suffizienten Transformationen T_n gehen die Experimente F_n über in

$$E_n = \mathcal{N}(A_n \Sigma_n^{-1}, I_n).$$

Beispiel 2.2 Setzt man in Beispiel 2.1 die Fehlerverteilung *nicht* als bekannt voraus, sondern lässt alle Verteilungen mit Erwartungswert 0 und Varianz 1 zu, so erhält man die Linearen Modelle

$$\mathcal{L}(A_n, \Sigma_n^2)$$

mit A_n und Σ_n^2 wie in Beispiel 2.1 (man vergleiche Notation (1.4) auf Seite 15 mit $\mathcal{X} = \mathbb{R}^n$). Durch die suffizienten Transformationen T_n gehen die Experimente $\mathcal{L}(A_n, \Sigma_n^2)$ über in

$$\mathcal{L}(A_n \Sigma_n^{-1}, I_n).$$

Solche Transformationen entsprechen auch dem üblichen Vorgehen in der Theorie der Linearen Modelle.

2.2 Konvergenzklassen bei gegebenen Designfolgen

Wir wollen in diesem Kapitel Konvergenzaussagen für die in Abschnitt 2.1 beschriebenen Regressionsmodelle angeben. Eine Grundvoraussetzung hierbei ist, dass sich die Folge der Versuchspunkte t_{n1}, \dots, t_{nn} nicht völlig willkürlich verhält. Es sei

$$H_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{[a,t]}(t_{ni}), \quad t \in [a, b], \quad (2.10)$$

die zu den Versuchspunkten t_{n1}, \dots, t_{nn} gehörige Verteilungsfunktion, wobei wir die Folge der H_n künftig auch als **Designfolge** bezeichnen. Diese muss gegen eine Verteilungsfunktion H auf $[a, b]$ konvergieren, die als **Grenzdesign** bezeichnet und mit dem zugehörigen Wahrscheinlichkeitsmaß identifiziert wird. Die folgenden beiden Konvergenzarten sind für die weiteren Überlegungen von Bedeutung:

1. *Verteilungskonvergenz* (schwache Konvergenz) der zugehörigen Wahrscheinlichkeitsmaße, also

$$H_n \xrightarrow{\mathcal{D}} H, \quad (2.11)$$

was gleichbedeutend ist mit der punktweisen Konvergenz $H_n(t) \rightarrow H(t)$ in allen Stetigkeitspunkten t von H .

2. *Gleichmäßige Konvergenz* der Verteilungsfunktionen, also

$$\sup_{t \in [a, b]} |H_n(t) - H(t)| \rightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (2.12)$$

Ist das Grenzdesign H stetig, so sind die beiden Konvergenzarten nach dem Satz von Pólya identisch.

Beispiel 2.3 Ist H gegeben mit rechtsseitig stetiger Quantilfunktion H^{-1} und gilt für die Versuchspunkte

$$t_{ni} = H^{-1} \left(\frac{i}{n} \right), \quad i = 1, \dots, n,$$

so nennen wir H_n eine **H -äquidistante Designfolge**. In diesem Fall liegt gleichmäßige Konvergenz vor, also

$$\sup_{t \in [a, b]} |H_n(t) - H(t)| \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Sind Designs H_n gegeben, die im Sinne von (2.11) oder (2.12) gegen ein Grenzdesign H konvergieren, so interessieren uns vor allen diejenigen Funktionen $f : [a, b] \rightarrow \mathbb{R}$, für die

$$\frac{1}{n} \sum_{i=1}^n f(t_{ni}) = \int_{[a, b]} f dH_n \xrightarrow{n \rightarrow \infty} \int_{[a, b]} f dH$$

gilt. Genauer gesagt, sind die folgenden Funktionenklassen von Interesse.

Definition 2.4 Sind Designs H_n gegeben, die im Sinne von (2.11) oder (2.12) gegen ein Grenzdesign H konvergieren, so heißt eine Menge $\mathbf{K}(H_n, H) \subset \mathbf{L}^2(H)$ von Funktionen $f : [a, b] \rightarrow \mathbb{R}$ eine **H_n - H -Konvergenzklasse**, wenn für alle $f_1, f_2 \in \mathbf{K}(H_n, H)$ gilt

$$\langle f_1, f_2 \rangle_{L^2(H_n)} = \int_{[a, b]} f_1 f_2 dH_n \xrightarrow{n \rightarrow \infty} \int_{[a, b]} f_1 f_2 dH = \langle f_1, f_2 \rangle_{L^2(H)}.$$

Dabei müssen f_1, f_2 nicht notwendig verschieden sein. Die Innenprodukte $\langle f_1, f_2 \rangle_{L^2(H_n)}$ brauchen erst ab hinreichend großem $n \in \mathbb{N}$ zu existieren.

Bemerkung 2.5

1. Die lineare Hülle (im Vektorraum $L^2(H)$) einer H_n - H -Konvergenzklasse ist wiederum eine H_n - H -Konvergenzklasse.
2. Jede Teilmenge einer H_n - H -Konvergenzklasse ist eine H_n - H -Konvergenzklasse.

Die folgenden Beispiele für H_n - H -Konvergenzklassen sind von großer Bedeutung und decken nahezu alle praxisrelevanten Regressionsprobleme ab. Dabei sind die angegebenen Funktionenklassen jeweils als *Teilmengen von $L^2(H)$* zu verstehen.

Beispiel 2.6 Es seien Designs H_n gegeben, die im Sinne von (2.11) oder (2.12) gegen ein Grenzdesign H konvergieren.

1. Liegt Verteilungskonvergenz der H_n vor, so bildet der *Raum aller stetigen Funktionen* $C[a, b]$ definitionsgemäß eine H_n - H -Konvergenzklasse.

2. Liegt gleichmäßige Konvergenz der Verteilungsfunktionen H_n vor, so bildet nach Satz A.18 in Anhang A.4 der *Raum aller Funktionen von beschränkter Variation* $BV[a, b]$ eine H_n - H -Konvergenzklasse.

Man beachte, dass beide Räume Funktionenalgebren sind.

2.3 Asymptotik mit dem Satz von Donsker

Die Asymptotik von Regressionsmodellen, die man mittels des Satzes von Donsker erhält, ist nicht zentraler Bestandteil dieser Arbeit, sondern wird ausführlich und mit verschiedenen Anwendungen in den zu Beginn dieses Kapitels zitierten Arbeiten untersucht. Dennoch wollen wir hier kurz diesen Weg hin zu einem asymptotischen Regressionsmodell für eine spezielle Situation schildern, um Unterschiede zur Asymptotik der folgenden Abschnitte besser erkennen zu können. Seien dazu

- die Regressorenklasse $\mathcal{F} \subseteq BV[a, b]$,
- \mathcal{Q} die Menge aller Wahrscheinlichkeitsmaße auf \mathcal{B} mit Erwartungswert 0 und Varianz 1,
- $\sigma : [a, b] \rightarrow (0, \infty)$, $\sigma \in BV[a, b]$, eine bekannte Funktion und
- H_n eine H -äquidistante Designfolge zu gegebenem Grenzdesign H , die somit nach Beispiel 2.3 gleichmäßig gegen H konvergiert.

Wir verwenden ab jetzt wieder die Bezeichnungen von Abschnitt 2.1 und wollen hier nur nichttransformierte Experimente, ähnlich zu F_n aus (2.5), untersuchen. Im Gegensatz zu den F_n betrachten wir die Experimente

$$\tilde{F}_n = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q \left(\frac{\cdot - f(t_{ni})}{\sigma(t_{ni})} \right) \mid (f, Q) \in \mathcal{F} \times \mathcal{Q} \right\} \right), \quad (2.13)$$

d.h. der Parameterraum ist $\mathcal{F} \times \mathcal{Q}$. Sowohl die wahre Regressionsfunktion als auch die wahre Fehlerverteilung sind unbekannt. Experimente mit einer solchen Struktur des Parameterraums sind uns bei der Einführung Linearer Modelle in Notation (1.4) auf Seite 15 bereits begegnet. Ist $\mathcal{F} = [f_1, \dots, f_k]$ mit bekannten Funktionen f_1, \dots, f_k wie in Beispiel 2.2, so gilt $\tilde{F}_n = \mathcal{L}(A_n, \Sigma_n^2)$.

Im Folgenden werden wir kurz den Weg hin zu einem Grenzmodell skizzieren. Seien die linearen Operatoren $T_n : \mathbb{R}^n \rightarrow C[0, 1]$ gegeben durch

$$(T_n(a))(s) = \sum_{i=1}^{[ns]} a_i + (ns - [ns])a_{[ns]+1}, \quad s \in [0, 1], \quad (2.14)$$

für $a = (a_1, \dots, a_n)^\top \in \mathbb{R}^n$, $a_{n+1} = 0$, wobei $[\cdot]$ die Gauß-Klammer bezeichne. Dann gilt für die Beobachtungen $Y_n = f(t_n) + \Sigma_n \epsilon_n$ (vgl. (2.4) für $f \in \mathcal{F}$ und $\epsilon_{n1} \sim Q \in \mathcal{Q}$)

$$T_n(Y_n) = T_n(f(t_n)) + T_n(\Sigma_n \epsilon_n).$$

Satz 2.7 Für $n \rightarrow \infty$ gilt

$$\frac{1}{\sqrt{n}} T_n(\Sigma_n \epsilon_n) \xrightarrow{\mathcal{D}} \left(\int_0^s \sigma \circ H^{-1}(x) W(dx) \right)_{s \in [0,1]} \quad \text{in } C[0,1], \quad (2.15)$$

$$\frac{1}{n} T_n(f(t_n))(s) \rightarrow \int_{[0,s]} f \circ H^{-1} d\lambda \quad \text{gleichmäßig in } [0,1], \quad (2.16)$$

wobei W die Brownsche Bewegung mit stetigen Pfaden auf $[0,1]$ bezeichnet.

Wegen

$$\frac{1}{n} T_n(f(t_n)) = \frac{1}{\sqrt{n}} T_n \left(\frac{1}{\sqrt{n}} f(t_n) \right)$$

liegt eine Lokalisierung in \mathcal{F} mit der Rate $\frac{1}{\sqrt{n}}$ vor. Das Konzept der Lokalisierung wird weiter hinten in Abschnitt 2.4.1 genauer erläutert.

Die stochastischen Integrale in Formel (2.15) kann man entweder pfadweise als Riemann-Stieltjes Integrale verstehen oder als „stochastische Integrale nichtzufälliger Funktionen“, wie in Anhang A.5 beschrieben. Diese beiden Integralbegriffe stimmen fast sicher überein. Für Details dazu betrachte man zum Beispiel Wentzell (1981).

Beweis. Gleichung (2.15) ist eine Variante des Satzes von Donsker, deren Beweis wir hier lediglich skizzieren wollen. Dazu definieren wir die Abbildungen

$$h_n : \begin{cases} C[0,1] & \rightarrow C[0,1] \\ g & \mapsto \left(s \mapsto \int_0^s \sigma \circ H_n^{-1}(x) g(dx) \right) \end{cases}$$

für $n \in \mathbb{N}$ und

$$h : \begin{cases} C[0,1] & \rightarrow C[0,1] \\ g & \mapsto \left(s \mapsto \int_0^s \sigma \circ H^{-1}(x) g(dx) \right) \end{cases}.$$

Man beachte, dass diese Abbildungen wohldefiniert sind, da sie nach Satz A.21 in Anhang A.4 eine stetige Funktion g wiederum auf eine stetige Funktion abbilden. Dann kann man durch elementare Abschätzungen nachweisen, dass für alle $g_n, g \in C[0,1]$ mit $\|g_n - g\|_\infty \rightarrow 0$ gilt

$$\|h_n(g_n) - h(g)\|_\infty \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

Nach dem Satz von Donsker, vgl. Billingsley (1968), Theorem 10.1, gilt

$$\frac{1}{\sqrt{n}}T_n(\epsilon_n) \xrightarrow{\mathcal{D}} W,$$

und mit dem Satz von Rubin, vgl. Billingsley (1968), Theorem 5.5, folgt

$$\frac{1}{\sqrt{n}}T_n(\Sigma_n \epsilon_n) = h_n \left(\frac{1}{\sqrt{n}}T_n(\epsilon_n) \right) \xrightarrow{\mathcal{D}} h(W) = \left(\int_0^s \sigma \circ H^{-1}(x) W(dx) \right)_{s \in [0,1]}.$$

Ein Beweis von Gleichung (2.16) findet sich in Bischoff und Miller (2000b). \square

Dies führt uns für festes $f \in \mathcal{F}$ zu einem Grenzprozess für die Folge der „lokalisierten“ Prozesse $\frac{1}{\sqrt{n}}T_n(\frac{1}{\sqrt{n}}f(t_n) + \Sigma_n \epsilon_n)$, nämlich dem Gaußprozess

$$Z_f(s) := \int_{[0,s]} f \circ H^{-1}(x) \lambda(dx) + \int_0^s \sigma \circ H^{-1}(x) W(dx), \quad s \in [0,1],$$

der stetige Pfade hat (vgl. Satz A.21 und beachte, dass wir das stochastische Integral pfadweise auffassen können). Ist P^{Z_f} die Verteilung der Zufallsvariablen Z_f auf dem Messraum $(C[0,1], \mathcal{B}(C[0,1]))$, so können wir den Experimenten \tilde{F}_n das Grenzexperiment

$$\tilde{F} = (C[0,1], \mathcal{B}(C[0,1]), \{P^{Z_f} \mid f \in \mathcal{F}\}) \quad (2.17)$$

gegenüberstellen. Entscheidend hierbei ist, dass jegliche Information über die Fehlerverteilung in den Grenzexperimenten verloren ist, der Parameterraum ist nur noch \mathcal{F} . Die wahre Fehlerverteilung ist meistens jedoch ein reiner Störparameter und aus statistischer Sicht nicht von Interesse. Der statistisch relevante Parameter hingegen ist der Regressor.

Führen wir ähnliche Überlegungen für die transformierten Experimente

$$\left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f(t_{ni})}{\sigma(t_{ni})}} \mid (f, Q) \in \mathcal{F} \times \mathcal{Q} \right\} \right)$$

durch, so gelangen wir zu den Gaußprozessen

$$\bar{Z}_f(s) := \int_{[0,s]} \frac{f \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} \lambda(dx) + W(s), \quad s \in [0,1].$$

2.4 Lokale Theorie - Gaußsche Grenzwerte

Um einen stärkeren asymptotischen Zusammenhang zwischen Regressionsmodellen und Grenzexperimenten von ähnlicher Form wie \tilde{F} , vgl. (2.17), herzustellen, müssen wir in

diesem Kapitel die *Fehlerverteilung* als *bekannt* voraussetzen. Dies ist der wesentliche Unterschied zu dem Modell aus Abschnitt 2.3.

2.4.1 Das Prinzip der Lokalisierung

Wir verwenden im Folgenden die Bezeichnungen von Abschnitt 2.1. Insbesondere betrachten wir ab jetzt nur noch die Experimente

$$E_n = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f(t_{ni})}{\sigma(t_{ni})}} \mid f \in \mathcal{F} \right\} \right)$$

aus (2.6). Wenn wir in den folgenden Abschnitten von Konvergenz von Experimenten sprechen, so ist, wenn nicht ausdrücklich anders vermerkt, immer die *schwache Konvergenz im Sinne LeCam's* gemeint.

In Anbetracht von Formel (A.6) wird der Hellingerabstand zweier zu verschiedenen Parametern f_1, f_2 gehörigen Wahrscheinlichkeitsmaße $\bigotimes_{i=1}^n Q_{\frac{f_1(t_{ni})}{\sigma(t_{ni})}}$, $\bigotimes_{i=1}^n Q_{\frac{f_2(t_{ni})}{\sigma(t_{ni})}}$ im Allgemeinen gegen 1 konvergieren. Somit werden die Regressionsmodelle E_n nach Korollar 1.33 gegen ein entartetes Grenzexperiment, ein vollständig informatives Experiment, konvergieren. Darunter versteht man ein Experiment, bei dem die Wahrscheinlichkeitsmaße zu verschiedenen Parametern orthogonal sind, d.h. Hellingerabstand 1 haben. Bei diesen Experimenten genügt eine einzige Beobachtung, um den wahren Parameter mit Wahrscheinlichkeit 1 zu identifizieren. Um nichtentartete Grenzexperimente zu erhalten, werden wir die Experimente E_n mit einer *Rate* $\delta_n \downarrow 0$ *um einen fest vorgegebenen Parameter* $f_0 \in \mathcal{F}$ *lokalisieren*. Hier stellt sich natürlich die Frage, wie die Rate δ_n in *Abhängigkeit von der Fehlerverteilung* Q zu wählen ist. Konvergiert δ_n zu langsam gegen Null, so erhält man wie bei den nicht lokalisierten Experimenten ein vollständig informatives Grenzexperiment. Konvergiert δ_n zu schnell gegen Null, so erhält man ein vollständig uninformatives Experiment, d.h. ein Experiment, bei dem die Wahrscheinlichkeitsmaße zu verschiedenen Parametern identisch sind. Die Parameter sind dann nicht mehr identifizierbar.

Eine Antwort auf diese Fragestellung kann für den Fall der Linearen Regressionsmodelle aus den Überlegungen in Janssen et al. (1985), S. 42 f., hergeleitet werden, was jedoch etwas mühsam ist und hier nicht ausgeführt werden soll. Man erhält dabei das folgende Resultat:

An die Fehlerverteilung setzten wir voraus, dass Konstanten a, b, c existieren mit $a \geq 1$, $b \geq 0$, $c > 0$ und $(a = 1 \implies b = 0)$ und dass gilt

$$\lim_{\substack{s \rightarrow 0 \\ s \neq 0}} \frac{d_H^2(Q, Q_s)}{|s|^a \left(\log \frac{1}{|s|} \right)^b} = c. \quad (2.18)$$

Desweiteren fordern wir, dass die bekannten Regressoren $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ beschränkt sind. Dann ist die Rate

$$\delta_n := \frac{1}{\sqrt[a]{n \cdot (\log n)^b}}$$

geeignet, d.h. die Folge der um $f_0 \in \mathcal{F}$ lokalisierten Experimente

$$E_n^{loc, f_0} := \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f_0(t_{ni})}{\sigma(t_{ni})} + \delta_n \cdot \frac{f(t_{ni})}{\sigma(t_{ni})}} \mid f \in \mathcal{F} \right\} \right) \quad (2.19)$$

besitzt nur nichtentartete schwache Häufungspunkte.

Man beachte, dass nach Bemerkung 1.31 Häufungspunkte existieren. Man kann diese sogar charakterisieren. Sie sind „direkte Produkte“ von einem Gauß- und einem Poisson-Experiment, vgl. dazu und zur Definition von Poisson-Experimenten Janssen et al. (1985). Man beachte auch, dass dort bei Formel (1) auf Seite 42 das Quadrat beim Hellingerabstand vergessen wurde.

Ist der Lokalisierungspunkt $f_0 \equiv 0$, so werden wir für die um 0 lokalisierten Experimente künftig nur E_n^{loc} schreiben. Ebenso werden wir die Schreibweisen E_n^{loc} und E_n^{loc, f_0} verwenden, wenn der Parameterraum \mathcal{F} nicht von einem Linearen Regressionsmodell herrührt.

Bemerkung 2.8 Fehlerverteilungen Q , die Punktmasse besitzen, können Bedingung (2.18) nicht erfüllen.

Beweis. Sei x_0 ein beliebiger Punkt mit $Q(\{x_0\}) > 0$. Da Q ein endliches Maß ist, existiert ein $\epsilon > 0$ mit $Q(\{x\}) < \frac{1}{2}Q(\{x_0\})$ für alle x mit $|x - x_0| < \epsilon$. Folglich existiert eine Konstante $c > 0$ mit $d_H(Q, Q_s) > c$ für alle $0 < |s| < \epsilon$. Also kann (2.18) nicht erfüllt sein.

2.4.1.1 Lokalisierungsrate $\frac{1}{\sqrt{n}}$

Wir werden jetzt diskrete Fehlerverteilungen für unsere Konvergenzüberlegungen ausschließen und uns einer anderen wichtigen Klasse, einer Teilmenge der Lebesgue-stetigen Verteilungen, zuwenden.

Ist das zu der Fehlerverteilung Q gehörige Shiftexperiment E_Q L^2 -differenzierbar, d.h. ist die Abbildung $s \mapsto \sqrt{\frac{dQ_s}{d\lambda}}$ im Punkt $s = 0$ Fréchet-differenzierbar in $L^2(\lambda)$, und besitzt E_Q eine endliche positive Fisher-Information J , so ist Bedingung (2.18) erfüllt mit $a = 2$, $b = 0$ und $c = \frac{J}{8}$, vgl. Janssen et al. (1985), S. 42. Die passende Lokalisierungsrate ist

demnach $\delta_n = \frac{1}{\sqrt{n}}$. Unter den folgenden Voraussetzungen an Q , die auch als **Hájek-Bedingungen** bezeichnet werden, vgl. Hájek und Šidák (1967), ist diese Differenzierbarkeitseigenschaft erfüllt, vgl. Strasser (1985), S. 393, oder LeCam und Yang (1990), S. 109.

Voraussetzungen 2.9 *Wir setzen an die Fehlerverteilung Q voraus:*

1. Q besitze eine absolutstetige Lebesgue-Dichte q mit Ableitung q' gemäß Satz A.27 in Anhang A.7.
2. Es gelte

$$0 < J := \int \left(\frac{q'}{q}\right)^2 dQ = \int \left(\frac{q'}{q}\right)^2 q d\lambda < \infty, \quad (2.20)$$

wobei hier $0 \cdot \infty := 0$ gesetzt wird. J ist die Fisher-Information des Shiftexperiments E_Q , vgl. (2.2) auf Seite 38; diese ist wegen der Translationsinvarianz des Lebesgue-Maßes λ konstant.

Wir wollen einige Beispiele von Verteilungen anführen, die die Voraussetzungen 2.9 erfüllen. Die notwendigen Berechnungen sind einfach und werden hier nicht durchgeführt.

Beispiel 2.10 (Beispiele für Fehlerverteilungen, die 2.9 erfüllen)

Die Voraussetzungen 2.9 an die Fehlerverteilung Q werden von den folgenden Verteilungen erfüllt:

1. $N(\mu, \tau^2)$ -Verteilung. Die zugehörige Fisher-Information von E_Q ist dann

$$J = \frac{1}{\tau^2}.$$

2. $\Gamma(\alpha, \beta)$ -Verteilung für $\beta > 0$ und $\alpha > 2$. Q hat dann eine Lebesgue-Dichte

$$q(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot e^{-\beta x} \cdot x^{\alpha-1} \cdot 1_{(0,\infty)}(x),$$

und die zugehörige Fisher-Information von E_Q ist

$$J = \frac{\beta^2}{\alpha - 2}.$$

3. Weibull-Verteilung mit Parametern $\gamma > 0$ und $p > 2$. Q hat dann eine Lebesgue-Dichte

$$q(x) = \gamma p \cdot x^{p-1} \cdot e^{-\gamma x^p} \cdot 1_{(0,\infty)}(x),$$

und für $p = 4$ ist die zugehörige Fisher-Information von E_Q

$$J = 9\sqrt{\pi\gamma}.$$

Für andere Werte $p > 2$ ist J nicht mehr elementar berechenbar.

4. Zweiseitige Exponential-Verteilung mit Parameter $r > \frac{1}{2}$. Q hat dann eine Lebesgue-Dichte

$$q(x) = \left(\frac{2}{r} \Gamma \left(\frac{1}{r} \right) \right)^{-1} e^{-|x|^r},$$

und die zugehörige Fisher-Information von E_Q ist

$$J = r^2 \Gamma \left(2 - \frac{1}{r} \right) \frac{1}{\Gamma \left(\frac{1}{r} \right)}.$$

Bemerkenswert ist auch, dass wir an die Fehlerverteilung Q keine Momentenbedingung stellen. So muss z.B. der Erwartungswert des Fehlers nicht existieren, wie das folgende Beispiel illustriert.

Beispiel 2.11 Besitzt Q eine Cauchy-Verteilung mit Parameter $\beta > 0$, d.h. eine Lebesgue-Dichte

$$q(x) = \frac{\beta}{\pi} \cdot \frac{1}{\beta^2 + x^2}, \quad x \in \mathbb{R},$$

so sind die Voraussetzungen 2.9 erfüllt mit

$$J = \frac{1}{2\beta^2},$$

aber der Erwartungswert von Q existiert nicht.

Wir wollen jetzt noch einige Fehlerverteilungen betrachten, die die Voraussetzungen 2.9 nicht erfüllen. Die zugehörigen Shiftexperimente E_Q sind dann im Allgemeinen nicht mehr L^2 -differenzierbar. Dennoch erfüllen diese Fehlerverteilungen die Bedingung (2.18). Die geeignete Lokalisierungsrate δ_n ist i.A. nicht mehr die klassische $\frac{1}{\sqrt{n}}$ -Rate sondern eine „schnellere“ Rate. Für Aussagen über den Zusammenhang zwischen der Lokalisierungsrate und den möglichen schwachen Häufungspunkten der lokalisierten Experimente vergleiche man LeCam (1986). In Janssen et al. (1985), S. 7, wird erwähnt, dass unter „natürlichen Invarianzbedingungen“ $\frac{1}{\sqrt{n}}$ die langsamste sinnvolle Lokalisierungsrate ist und dass die schwachen Häufungspunkte bei dieser Rate nur Gaußexperimente sein können.

Beispiel 2.12 (Beispiele für Fehlerverteilungen, die 2.9 nicht erfüllen)

Die Voraussetzungen 2.9 werden von den folgenden Verteilungen Q nicht erfüllt.

1. Gleichverteilung auf einem Intervall $[\alpha, \beta]$. Dann ist Q_s eine Gleichverteilung auf $[\alpha + s, \beta + s]$, $s \in \mathbb{R}$. Für die Gleichverteilung existiert keine stetige, also erst recht keine absolutstetige Lebesgue-Dichte. Für $|s| < \beta - \alpha$ erhalten wir

$$d_H^2(Q, Q_s) = \frac{1}{2} \int \left(\sqrt{\frac{dQ}{d\lambda}} - \sqrt{\frac{dQ_s}{d\lambda}} \right)^2 d\lambda = \frac{|s|}{\beta - \alpha}.$$

Bedingung (2.18) ist somit erfüllt mit $a = 1, b = 0$ und $c = \frac{1}{\beta - \alpha}$. Als Lokalisierungsrate wählt man dementsprechend $\delta_n = \frac{1}{n}$.

2. $\Gamma(\alpha, \beta)$ -Verteilung mit Parametern $\beta > 0$ und $0 < \alpha \leq 2$. In diesen Fällen ist die Fisher-Information $J = \infty$. Für die $\Gamma(\alpha, 1)$ -Verteilung ist in Janssen et al. (1985), S. 43, die Bedingung (2.18) nachgewiesen. Als geeignete Lokalisierungsraten ergeben sich

$$\delta_n = \begin{cases} \frac{1}{\sqrt{n \log(n)}}, & \text{für } \alpha = 2 \\ \frac{1}{\sqrt[n]{n}}, & \text{für } 0 < \alpha < 2. \end{cases}$$

3. Zweiseitige Exponential-Verteilung mit Parameter $0 < r \leq \frac{1}{2}$ und Lebesgue-Dichte

$$q(x) = \left(\frac{2}{r} \Gamma\left(\frac{1}{r}\right) \right)^{-1} e^{-|x|^r}.$$

Auch hier ist die Fisher-Information $J = \infty$. Bedingung (2.18) ist zum Beispiel in LeCam und Yang (1990) nachgewiesen und führt zu den Lokalisierungsraten

$$\delta_n = \begin{cases} \frac{1}{\sqrt{n \log(n)}}, & \text{für } r = \frac{1}{2} \\ \frac{1}{n^{1+2r}}, & \text{für } 0 < r < \frac{1}{2}. \end{cases}$$

2.4.1.2 Exponentialentwicklung von Dichtequotienten

Betrachtet man Satz 1.36 und speziell Formel (1.11) auf Seite 26, so erkennt man, dass schwache Konvergenz im Sinne LeCam's direkt mit einer Exponentialdarstellung der Dichtequotienten einer Experimentfolge verbunden ist. Um mit Hilfe dieses Satzes schwache Konvergenz der uns interessierenden Regressionsmodelle nachzuweisen, benötigen wir eine solche Exponentialentwicklung. Zentrum dieser Überlegungen ist die L^2 -Differenzierbarkeit von Experimenten - eine Theorie, die auf LeCam selbst zurückgeht. Wir wollen hier nicht weiter auf die theoretischen Zusammenhänge eingehen, sondern lediglich die Konsequenzen für unsere Regressionsmodelle aufzeigen. Der folgende Satz entspricht einer Anwendung von Theorem 79.2 aus Strasser (1985) auf den für uns relevanten Fall und ist dort als Beispiel 79.4 explizit aufgeführt.

Satz 2.13 *Es sei Q eine Verteilung, die den Voraussetzungen 2.9 genügt. Weiter sei s_{ni} , $i = 1, \dots, n$, $n \in \mathbb{N}$, ein Dreiecksschema reeller Zahlen mit*

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} |s_{ni}| = 0 \tag{2.21}$$

und

$$\limsup_{n \rightarrow \infty} \sum_{i=1}^n s_{ni}^2 < \infty. \tag{2.22}$$

Dann gilt für $x_1, \dots, x_n \in \mathbb{R}$, $n \in \mathbb{N}$,

$$\prod_{i=1}^n \frac{dQ_{s_{ni}}}{dQ}(x_i) = \exp \left\{ \sum_{i=1}^n s_{ni} \cdot \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i) - \frac{J}{2} \sum_{i=1}^n s_{ni}^2 + o_{Q^n}(1) \right\}.$$

Bemerkung 2.14 Betrachten wir jetzt die mit der Rate $\delta_n := \frac{1}{\sqrt{n}}$ um den Ursprung $f_0 \equiv 0$ lokalisierten Regressionsmodelle E_n^{loc} aus Formel (2.19), so ist

$$s_{ni} = \frac{f(t_{ni})}{\sqrt{n} \sigma(t_{ni})}, \quad i = 1, \dots, n, \quad n \in \mathbb{N}.$$

Existiert ein Grenzdesign H , gegen das die Designfolge H_n im Sinne von (2.11) oder (2.12) konvergiert, und ist $\frac{1}{\sigma} \mathcal{F} := \{ \frac{f}{\sigma} \mid f \in \mathcal{F} \}$ eine H_n - H -Konvergenzklasse, so ist Bedingung (2.22) von Satz 2.13 für jede Funktion $f \in \mathcal{F}$ automatisch erfüllt, denn es gilt

$$\limsup_{n \rightarrow \infty} \sum_{i=1}^n s_{ni}^2 = \lim_{n \rightarrow \infty} \sum_{i=1}^n \left(\frac{f(t_{ni})}{\sqrt{n} \sigma(t_{ni})} \right)^2 = \lim_{n \rightarrow \infty} \int_{[a,b]} \frac{f^2}{\sigma^2} dH_n = \int_{[a,b]} \frac{f^2}{\sigma^2} dH < \infty.$$

Bedingung (2.21) wird dann zu

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{|f(t_{ni})|}{\sqrt{n} \sigma(t_{ni})} = 0 \tag{2.23}$$

und ist trivialerweise erfüllt, wenn die Funktion $\frac{f}{\sigma}$ beschränkt ist.

Im Kontext lokalisierter Regressionsmodelle wird man aus der Gültigkeit von Bedingung (2.22) oftmals auf die Gültigkeit von (2.23) schließen können. Das folgende Beispiel, in dem (2.22) gilt, aber (2.23) nicht gültig ist, ist völlig unnatürlich und soll hier nur der Vollständigkeit wegen angeführt werden.

Beispiel 2.15 Es seien $\mathcal{V} = [0, 1]$, $H = id_{[0,1]}$ die Verteilungsfunktion des Lebesgue-Maßes auf $[0, 1]$ und $t_{ni} = \frac{i}{n}$ für $i = 2, \dots, n$, sowie $t_{n1} = \frac{\pi}{4n}$. Dann konvergieren die zu den Versuchspunkten gehörigen Verteilungsfunktionen H_n offensichtlich gleichmäßig gegen H . Weiter sei $\sigma \equiv 1$ und f gegeben durch

$$f(x) := \begin{cases} 0 & \text{für } x \in \mathbb{Q}, \\ \sqrt{n} & \text{für } x = \frac{\pi}{4n}, \quad n \in \mathbb{N}, \\ 1 & \text{sonst.} \end{cases}$$

Dann ist (2.22) erfüllt wegen

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \left(\frac{f(t_{ni})}{\sqrt{n} \sigma(t_{ni})} \right)^2 = \lim_{n \rightarrow \infty} \int_{[0,1]} \frac{f^2}{\sigma^2} dH_n = 1 = \int_{[0,1]} \frac{f^2}{\sigma^2} dH,$$

aber es ist $\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{|f(t_{ni})|}{\sqrt{n} \sigma(t_{ni})} = 1$, und somit ist (2.23) nicht erfüllt.

2.4.2 Schwache Konvergenz Linearer Modelle mit Lokalisierungsrate $\frac{1}{\sqrt{n}}$

Bevor wir uns mit der Asymptotik allgemeiner Regressionsmodelle beschäftigen, werden wir den wichtigen Spezialfall der Linearen Regressionsmodelle untersuchen. Die in diesem Abschnitt angeführten Resultate werden wir später für den Beweis des allgemeinen Falls nutzen. Die lokale Asymptotik von Linearen Regressionsmodellen wurde bereits in Witting und Müller-Funk (1995) untersucht. Der folgende Satz ist, genau genommen, in der dortigen Formulierung des 2. LeCam Lemmas (S. 412, dort aber nicht bewiesen) enthalten. Es ist jedoch einfacher, das Ergebnis direkt zu beweisen, als die dortigen Voraussetzungen zu überprüfen und die dortige Aussage geeignet umzuformulieren. Anzumerken ist auch noch, dass die Formulierung des 2. LeCam Lemmas in der abstrakten Version von Witting und Müller-Funk (1995) ein rein theoretisches Ergebnis ist. Erst die folgende Version, Satz 2.16, mit „konvergenten Designfolgen“ erlaubt die Anwendung der Konvergenzaussagen auf praktisch relevante Problemstellungen.

Gegeben seien die Linearen Regressionsmodelle

$$E_n = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})}} \mid \beta \in \mathbb{R}^k \right\} \right)$$

aus (2.7) auf Seite 40 mit bekanntem $f = (f_1, \dots, f_k)^\top$. Einer einfachen Darstellung wegen lokalisieren wir nur um den Ursprung und erhalten die lokalisierten Experimente

$$E_n^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{1}{\sqrt{n}} \cdot \frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})}} \mid \beta \in \mathbb{R}^k \right\} \right), \quad (2.24)$$

für die der folgende Konvergenzsatz gilt.

Satz 2.16 (2. LeCam Lemma) *Es gelten die Voraussetzungen 2.9 an die Fehlerverteilung und die folgenden Voraussetzungen an die Regressoren, die Varianzfunktion und die Designfolge:*

1. *Es existiere ein Grenzdesign H , gegen das die Designfolge H_n im Sinne von (2.11) oder (2.12) konvergiert,*
2. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ bilden eine H_n - H -Konvergenzklasse, und es gelte*

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{|f_j(t_{ni})|}{\sqrt{n}\sigma(t_{ni})} = 0, \quad \text{für } j = 1, \dots, k, \quad (2.25)$$

3. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien linear unabhängig in $L^2(H)$.*

Sei

$$B := \begin{pmatrix} \int_{[a,b]} \frac{f_1^2}{\sigma^2} dH & \cdots & \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH \\ \vdots & \ddots & \vdots \\ \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH & \cdots & \int_{[a,b]} \frac{f_k^2}{\sigma^2} dH \end{pmatrix}. \quad (2.26)$$

Da $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ linear unabhängig in $L^2(H)$ sind, ist B und damit auch JB positiv definit, und durch $\langle \beta, \gamma \rangle_{JB} := J\beta^\top B\gamma$, $\beta, \gamma \in \mathbb{R}^k$, wird ein Innenprodukt auf \mathbb{R}^k definiert. Damit ist $H_{JB} := (\mathbb{R}^k, \langle \cdot, \cdot \rangle_{JB})$ ein Hilbertraum.

Unter obigen Voraussetzungen und mit diesen Bezeichnungen konvergiert die Folge der um den Ursprung lokalisierten Linearen Regressionsmodelle E_n^{loc} schwach gegen den Standard-Gaußshift $G_{H_{JB}}$, vgl. Beispiel 1.26, auf H_{JB} , also

$$E_n^{\text{loc}} \xrightarrow{w} G_{H_{JB}}.$$

$G_{H_{JB}}$ ist nach Beispiel 1.28 äquivalent zu $\mathcal{N}\left(\sqrt{JB}^{\frac{1}{2}}, I_k\right)$, so dass auch die schwache Konvergenz

$$E_n^{\text{loc}} \xrightarrow{w} \mathcal{N}\left(\sqrt{JB}^{\frac{1}{2}}, I_k\right)$$

folgt. Dabei ist $B^{\frac{1}{2}}$ die eindeutig bestimmte symmetrische und positiv definite Wurzel von B . Durch

$$X_{n,\beta}(x) := \sum_{i=1}^n \frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})} \cdot \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i), \quad (2.27)$$

$\beta \in \mathbb{R}^k$, $n \in \mathbb{N}$, $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, ist eine zentrale Folge gegeben.

Beweis. Sei

$$B_n := \begin{pmatrix} \int_{[a,b]} \frac{f_1^2}{\sigma^2} dH_n & \cdots & \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_n \\ \vdots & \ddots & \vdots \\ \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_n & \cdots & \int_{[a,b]} \frac{f_k^2}{\sigma^2} dH_n \end{pmatrix}. \quad (2.28)$$

Da nach Voraussetzung die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ eine H_n - H -Konvergenzklasse bilden, konvergieren die Matrizen B_n elementweise gegen B , und damit folgt für jedes $\beta \in \mathbb{R}^k$

$$\|\beta\|_{JB}^2 - J\beta^\top B_n\beta \xrightarrow{n \rightarrow \infty} 0. \quad (2.29)$$

Auf Grund der Voraussetzungen und Bemerkung 2.14 führt Satz 2.13 zu der folgenden Exponentialentwicklung des Likelihood-Quotienten

$$\frac{d \otimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}}}{dQ^n}(x) = \prod_{i=1}^n \frac{dQ_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}}}{dQ}(x_i)$$

$$\begin{aligned}
&\stackrel{2.13}{=} \exp \left\{ \sum_{i=1}^n \frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})} \cdot \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i) - \frac{J}{2} \sum_{i=1}^n \left(\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})} \right)^2 + o_{Q^n}(1) \right\} \\
&= \exp \left\{ X_{n,\beta}(x) - \frac{J}{2} \beta^\top B_n \beta + o_{Q^n}(1) \right\} \\
&= \exp \left\{ X_{n,\beta}(x) - \frac{1}{2} \|\beta\|_{JB}^2 + \frac{1}{2} \|\beta\|_{JB}^2 - \frac{J}{2} \beta^\top B_n \beta + o_{Q^n}(1) \right\} \\
&\stackrel{(2.29)}{=} \exp \left\{ X_{n,\beta}(x) - \frac{1}{2} \|\beta\|_{JB}^2 + o_{Q^n}(1) \right\}, \quad x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n.
\end{aligned}$$

Um zum Nachweis der schwachen Konvergenz $E_n^{loc} \xrightarrow{w} G_{H_{JB}}$ Korollar 1.38 anzuwenden, müssen wir zeigen, dass für jedes $\beta \in \mathbb{R}^k$ die Verteilung von $X_{n,\beta}$ unter dem Produktmaß Q^n schwach gegen $N(0, \|\beta\|_{JB}^2)$ konvergiert. Seien dazu

$$X_{ni} := \underbrace{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}}_{=: s_{ni}} \cdot \underbrace{\left(-\frac{q'}{q} \right) \cdot 1_{\{q>0\}}}_{=: \tilde{q}}, \quad i = 1, \dots, n.$$

Wir berechnen Erwartungswert und Varianz von \tilde{q} bezüglich Q . Man beachte dabei, dass aus (2.20) die Lebesgue-Integrierbarkeit von q' folgt, denn aus der quadratischen Integrierbarkeit von $\frac{q'}{q}$ bezüglich Q folgt die Integrierbarkeit von $\frac{q'}{q}$ bezüglich Q und damit die Lebesgue-Integrierbarkeit von q' . Mit Lemma A.29, Lemma A.30, Korollar A.28 und Gleichung (2.20) folgt

$$\begin{aligned}
E_Q(\tilde{q}) &= \int \tilde{q} dQ = - \int q' d\lambda = 0, \\
\text{Var}_Q(\tilde{q}) &= \int \tilde{q}^2 dQ = \int \left(\frac{q'}{q} \right)^2 q d\lambda = J.
\end{aligned}$$

Da wir die Verteilung von $X_{n,\beta}$ unter dem Produktmaß Q^n betrachten wollen, sind die X_{ni} stochastisch unabhängig, und es gilt

$$s_n^2 := \text{Var} \left(\sum_{i=1}^n X_{ni} \right) = \sum_{i=1}^n \text{Var}(X_{ni}) = \frac{J}{n} \sum_{i=1}^n \left(\frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})} \right)^2 = J \beta^\top B_n \beta \xrightarrow{n \rightarrow \infty} \|\beta\|_{JB}^2. \tag{2.30}$$

Wegen Voraussetzung (2.25) ist, wie wir in Bemerkung 2.14 gesehen haben, Bedingung (2.21) für die vorliegenden s_{ni} erfüllt. Damit und wegen (2.30) können wir den Zentralen Grenzwertsatz für lineare Statistiken, vgl. Witting und Müller-Funk (1995), S. 112 f., anwenden. Es folgt daraus, dass die Verteilung von $X_{n,\beta}$ unter dem Produktmaß Q^n schwach gegen $N(0, \|\beta\|_{JB}^2)$ konvergiert. Damit haben wir aber

$$E_n^{loc} \xrightarrow{w} G_{H_{JB}}$$

nachgewiesen und auch gezeigt, dass durch (2.27) eine zentrale Folge gegeben ist. \square

2.4.3 Schwache Konvergenz allgemeiner Regressionsmodelle mit Lokalisierungsrate $\frac{1}{\sqrt{n}}$

Wir verlassen nun den endlichdimensionalen parametrischen Rahmen und untersuchen die Asymptotik allgemeiner Regressionsmodelle. Dabei treten als Grenzwerte Gaußexperimente mit Beobachtungsraum $C[0, 1]$ auf. Bevor wir einen Konvergenzsatz für allgemeine Regressionsmodelle beweisen, werden zunächst die zugehörigen Grenzmodelle eingeführt und analysiert.

Es seien H ein beliebiges Grenzdesign auf dem Versuchsbereich $\mathcal{V} = [a, b]$, $J \in (0, \infty)$ und W eine Version der Brownschen Bewegung mit stetigen Pfaden auf $[0, 1]$. Für die Klasse der Regressoren \mathcal{F} gelte

$$\mathcal{F} \subseteq \left\{ f \mid \frac{f}{\sigma} \in L^2(H) \right\} =: \mathcal{F}_\sigma,$$

und für alle $f_1, f_2 \in \mathcal{F}$ mit $f_1 \neq f_2$ sei $\frac{f_1}{\sigma} \neq \frac{f_2}{\sigma}$ in $L^2(H)$. Zu jedem $f \in \mathcal{F}$ definieren wir eine zugehörige Funktion $h_f : [0, 1] \rightarrow \mathbb{R}$ durch

$$h_f(s) := \sqrt{J} \int_{[0, s]} \frac{f \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} \lambda(dx), \quad s \in [0, 1], \quad (2.31)$$

sowie einen stochastischen Prozess auf $[0, 1]$ durch

$$Z_f(s) := h_f(s) + W(s), \quad s \in [0, 1]. \quad (2.32)$$

Wegen $\frac{f}{\sigma} \in L^2(H)$ ist nach dem Transformationssatz A.17 die Funktion

$$h'_f := \sqrt{J} \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} \in L^2(\lambda),$$

und es gilt $h_f \in H(R)$, wobei

$$H(R) = \left\{ h : [0, 1] \rightarrow \mathbb{R} \mid \exists h' \in L^2(\lambda) : h(s) = \int_{[0, s]} h'(x) \lambda(dx), s \in [0, 1] \right\}$$

den zur Brownschen Bewegung W auf $[0, 1]$ gehörigen kernreproduzierenden Hilbertraum bezeichnet, vgl. Beispiel A.25 in Anhang A.6. Insbesondere ist Z_f ein Gaußprozess mit stetigen Pfaden. Bezeichnen P^{Z_f} bzw. P^W die Verteilungen von Z_f bzw. W auf dem Messraum $(C[0, 1], \mathcal{B}(C[0, 1]))$, so sind die Maße P^{Z_f} und P^W nach Anhang A.6 äquivalent. Aus dem Transformationssatz A.17 und Beispiel 1.27 folgt direkt

Lemma 2.17 *Unter den gerade beschriebenen Voraussetzungen, insbesondere sei $\frac{f}{\sigma} \in L^2(H)$ für alle $f \in \mathcal{F}$, ist das Experiment*

$$E = (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{Z_f} \mid f \in \mathcal{F}\}) \quad (2.33)$$

ein Gaußexperiment mit Beobachtungsraum $C[0, 1]$. Es besitzt für $f_1, f_2 \in \mathcal{F}$ den bilinearen Covarianzkern zur Basis 0:

$$\text{Cov} \left(\log \left(\frac{dP^{Z_{f_1}}}{dP^W} \right), \log \left(\frac{dP^{Z_{f_2}}}{dP^W} \right) \right) = \langle h_{f_1}, h_{f_2} \rangle_{H(R)} = J \int_{[a,b]} \frac{f_1 f_2}{\sigma^2} dH = J \langle \frac{f_1}{\sigma}, \frac{f_2}{\sigma} \rangle_{L^2(H)}.$$

Also ist E für $0 \in \mathcal{F}$ ein Gaußshift im Sinne von Definition 1.19. Für die Hellingertransformierte von E gilt

$$\begin{aligned} H(E_\alpha)(z) &= \exp \left\{ \frac{J}{2} \left(\sum_{i,j=1}^r z_i z_j \langle \frac{f_i}{\sigma}, \frac{f_j}{\sigma} \rangle_{L^2(H)} - \sum_{i=1}^r z_i \left\| \frac{f_i}{\sigma} \right\|_{L^2(H)}^2 \right) \right\} \\ &= \exp \left\{ -\frac{J}{4} \sum_{i,j=1}^r z_i z_j \left\| \frac{f_i - f_j}{\sigma} \right\|_{L^2(H)}^2 \right\} \end{aligned}$$

für $\alpha := \{f_1, \dots, f_r\} \in A(\mathcal{F})$, $z = (z_1, \dots, z_r)^\top \in S_\alpha$.

Ist $0 \notin \mathcal{F}$, so ist E zwar ein Gaußexperiment, aber kein Gaußshift. Der Covarianzkern zur Basis $f_0 \in \mathcal{F}$ ist dann nicht mehr bilinear; die Darstellung der Hellingertransformierten bleibt aber unverändert, da sie unabhängig von der Wahl der Basis ist, vgl. Lemma 1.22.

Aus Lemma 2.17 und Lemma A.11 folgt direkt

Korollar 2.18 Für $f_1, f_2 \in \mathcal{F}$ gilt

$$\begin{aligned} a(P^{Z_{f_1}}, P^{Z_{f_2}}) &= \exp \left\{ -\frac{J}{8} \left\| \frac{f_1 - f_2}{\sigma} \right\|_{L^2(H)}^2 \right\}, \\ d_H^2(P^{Z_{f_1}}, P^{Z_{f_2}}) &= 1 - a(P^{Z_{f_1}}, P^{Z_{f_2}}) = 1 - \exp \left\{ -\frac{J}{8} \left\| \frac{f_1 - f_2}{\sigma} \right\|_{L^2(H)}^2 \right\}. \end{aligned}$$

Identifiziert man die Funktionen f mit ihren integrierten Versionen h_f , so kann man den Parameterraum \mathcal{F} als eine Teilmenge von $H(R)$ betrachten. Anhand des Covarianzkerns des Experiments E zur Basis 0, wie er in Lemma 2.17 angegeben ist, erkennen wir, dass E die Einschränkung $G_{H(R), \mathcal{F}}$ des Standard-Gaußshifts $G_{H(R)}$ auf den Parameterraum \mathcal{F} ist, vgl. Definition 1.24, also dass

$$E = G_{H(R), \mathcal{F}} \tag{2.34}$$

gilt. Es stellt sich die Frage, ob wir für den Fall des Parameterraums $\mathcal{F} = \mathcal{F}_\sigma = \{f \mid \frac{f}{\sigma} \in L^2(H)\}$ einen „vollständigen“ Standard-Gaußshift auf $H(R)$ erhalten. Wegen der fest vorgegebenen Verteilungsfunktion H , die im Experiment E eine wichtige Rolle spielt, ist dies im Allgemeinen nicht der Fall. Schränken wir den Raum $H(R)$ jedoch ein, so erhalten wir einen Standard-Gaußshift auf einem etwas kleineren Hilbertraum.

Lemma 2.19 *Sei*

$$\tilde{H}(R) := \left\{ s \mapsto \sqrt{J} \int_{[0,s]} \frac{f \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} \lambda(dx), s \in [0, 1] \mid f \in \mathcal{F}_\sigma \right\} \subset H(R),$$

und sei weiter $\langle \cdot, \cdot \rangle_{\tilde{H}(R)}$ das auf $\tilde{H}(R)$ eingeschränkte Innenprodukt $\langle \cdot, \cdot \rangle_{H(R)}$. Dann ist $(\tilde{H}(R), \langle \cdot, \cdot \rangle_{\tilde{H}(R)})$ wiederum ein Hilbertraum.

Beweis. Es ist nur die Vollständigkeit zu zeigen. Für Elemente von $\tilde{H}(R)$ verwenden wir die Bezeichnung aus (2.31). Ist $(h_{f_n})_{n \in \mathbb{N}}$ eine Cauchy-Folge in $(\tilde{H}(R), \langle \cdot, \cdot \rangle_{\tilde{H}(R)})$, so bilden wegen

$$\|h_{f_n} - h_{f_m}\|_{\tilde{H}(R)}^2 = \int_{[a,b]} J \frac{(f_n - f_m)^2}{\sigma^2} dH = \left\| \sqrt{J} \frac{f_n - f_m}{\sigma} \right\|_{L^2(H)}^2 \quad (2.35)$$

die Funktionen $\frac{\sqrt{J}f_n}{\sigma}$ eine Cauchy-Folge in $L^2(H)$. Also existiert eine Funktion $\tilde{f} \in L^2(H)$ mit $\left\| \frac{\sqrt{J}f_n}{\sigma} - \tilde{f} \right\|_{L^2(H)} \rightarrow 0$ für $n \rightarrow \infty$. Damit ist die Funktion $f := \frac{\tilde{f}\sigma}{\sqrt{J}} \in \mathcal{F}_\sigma$, und genau wie in (2.35) folgt die Konvergenz

$$\|h_{f_n} - h_f\|_{\tilde{H}(R)} \rightarrow 0$$

für $n \rightarrow \infty$. □

Jeder Funktion $f \in \mathcal{F}_\sigma$ können wir eindeutig ein Hilbertraumelement

$$h_f(\cdot) = \sqrt{J} \int_{[0,\cdot]} \frac{f \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} \lambda(dx)$$

aus $\tilde{H}(R)$ zuordnen. Umgekehrt können wir auf Grund von Lemma 2.19 jedem Hilbertraumelement aus $\tilde{H}(R)$ eindeutig eine Funktion aus \mathcal{F}_σ zuordnen. Liegt also dem Experiment E aus (2.33) als Parameterraum \mathcal{F}_σ zu Grunde, so können wir E schreiben als

$$E = \left(C[0, 1], \mathcal{B}(C[0, 1]), \{P^{Z_f} \mid h_f \in \tilde{H}(R)\} \right), \quad (2.36)$$

und aus Lemma 2.17 und Definition 1.24 folgt direkt

Korollar 2.20 *Das Experiment E aus (2.36) ist der Standard-Gaußshift auf dem Hilbertraum $(\tilde{H}(R), \langle \cdot, \cdot \rangle_{\tilde{H}(R)})$.*

Bemerkung 2.21 Der Parameterraum \mathcal{F}_σ , versehen mit der kanonischen Norm des $L^2(H)$, ist im Allgemeinen kein vollständiger Raum mehr. Wählt man z.B. auf $[0, 1]$ als Verteilungsfunktion $H := id_{[0,1]}$ und weiter

$$\sigma(x) := 1_{\{0\}}(x) + x, \quad f_n(x) := x^{\frac{1}{2} + \frac{1}{n}}, \quad f(x) := \sqrt{x}, \quad x \in [0, 1], \quad n \in \mathbb{N},$$

so ist $f_n \in \mathcal{F}_\sigma$ für alle $n \in \mathbb{N}$, es gilt $\|f_n - f\|_{L^2(H)} \rightarrow 0$, aber $f \notin \mathcal{F}_\sigma$.

Nach diesen Vorüberlegungen kommen wir zu einem Konvergenzsatz für allgemeine Regressionsmodelle, einem Analogon zu Satz 2.16.

Satz 2.22 *Es gelten die Voraussetzungen 2.9 für die Fehlerverteilung Q . Weiter setzen wir voraus:*

1. *Es existiere ein Grenzdesign H , gegen das die Designfolge H_n im Sinne von (2.11) oder (2.12) konvergiert,*
2. *die Klasse $\frac{1}{\sigma}\mathcal{F} := \{\frac{f}{\sigma} \mid f \in \mathcal{F}\}$ sei eine H_n - H -Konvergenzklasse (insbesondere ist dann $\frac{f}{\sigma} \in L^2(H)$ für alle $f \in \mathcal{F}$), für alle Funktionen $f \in \mathcal{F}$ gelte*

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{|f(t_{ni})|}{\sqrt{n}\sigma(t_{ni})} = 0,$$

und es sei $0 \in \mathcal{F}$.

Für $f \in \mathcal{F}$ betrachten wir den stochastischen Prozess, vgl. (2.32),

$$Z_f(s) := \sqrt{J} \int_{[0,s]} \frac{f \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} \lambda(dx) + W(s) = h_f(s) + W(s), \quad s \in [0, 1].$$

Hierbei seien J wie in (2.20) die Fisher-Information des Shiftexperiments E_Q und W eine Version der Brownschen Bewegung mit stetigen Pfaden. P^{Z_f} sei die Verteilung der Zufallsvariablen Z_f auf dem Messraum $(C[0, 1], \mathcal{B}(C[0, 1]))$ und E das zugehörige statistische Experiment, vgl. (2.33),

$$E = (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{Z_f} \mid f \in \mathcal{F}\}).$$

Dann konvergiert die Folge der um den Ursprung lokalisierten Regressionsmodelle

$$E_n^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{1}{\sqrt{n}}, \frac{f(t_{ni})}{\sigma(t_{ni})}} \mid f \in \mathcal{F} \right\} \right)$$

schwach gegen E , also

$$E_n^{loc} \xrightarrow{w} E.$$

E ist nach Lemma 2.17 ein Gaußshift im Sinne von Definition 1.19.

Beweis. Wir beweisen die Aussage mit Hilfe von Satz 1.32. Ist $\alpha = \{f_1, \dots, f_l\} \in A(\mathcal{F})$ beliebig, so müssen wir die punktweise Konvergenz der Hellingertransformierten $H(E_{n,\alpha}^{loc})$ gegen $H(E_\alpha)$ nachweisen. Schränken wir den Parameterraum $\frac{1}{\sigma}\mathcal{F}$ auf die lineare Hülle (in $L^2(H)$) von $\frac{1}{\sigma}\alpha$ ein, so liefert uns Satz 2.16 zusammen mit Satz 1.32 die Konvergenz der Hellingertransformierten des eingeschränkten Experiments gegen die Hellingertransformierte des zugehörigen endlichdimensionalen Gaußshifts, den wir zur Abkürzung G nennen. Letztere ist nach Beispiel 1.15 (der Parameterraum ist hier die lineare Hülle von α , und α selbst ist die endliche Teilmenge) für $z \in S_\alpha$ gegeben durch

$$H(G_\alpha)(z) = \exp \left\{ -\frac{J}{4} \sum_{i,j=1}^l z_i z_j \int \frac{(f_i - f_j)^2}{\sigma^2} dH \right\},$$

also nach Lemma 2.17 gleich der Hellingertransformierten $H(E_\alpha)(z)$. \square

Den Gaußshift E können wir sogar als Teilexperiment eines Standard-Gaußshifts auf einem Hilbertraum sehen.

Korollar 2.23 *Nach Satz 2.22 sowie (2.34) ist das Grenzexperiment E der auf \mathcal{F} eingeschränkte Standard-Gaußshift auf $H(R)$, also*

$$E = G_{H(R),\mathcal{F}}. \quad (2.37)$$

Analog zu der zentralen Folge aus Satz 2.16 definieren wir mittels

$$X_{n,f}(x) := \sum_{i=1}^n \frac{f(t_{ni})}{\sqrt{n}\sigma(t_{ni})} \cdot \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i), \quad (2.38)$$

$f \in \mathcal{F}$, $n \in \mathbb{N}$, $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, eine Folge linearer Prozesse $(X_n)_{n \in \mathbb{N}}$. Dann folgt auch analog zum Beweis von Satz 2.16 die schwache Konvergenz

$$\mathcal{L}(X_{n,f} \mid Q^n) \xrightarrow{\mathcal{D}} N(0, \|h_f\|_{H(R)}^2) \quad (2.39)$$

mit $h_f(\cdot) = \sqrt{J} \int_{[0,1]} \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} d\lambda$.

Wegen der Translationsinvarianz des Lebesgue-Maßes ergibt sich aus der Definition der Hellingertransformierten A.5 und den Sätzen 1.32 und 2.22 sofort das

Korollar 2.24 *Es gelten die Voraussetzungen von Satz 2.22. Wir fordern jedoch nicht mehr $0 \in \mathcal{F}$, sondern betrachten ein festes $f_0 \in \mathcal{F}$. Dann konvergiert die Folge der um f_0 lokalisierten Regressionsmodelle*

$$E_n^{loc, f_0} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f_0(t_{ni})}{\sigma(t_{ni})} + \frac{1}{\sqrt{n}} \cdot \frac{f(t_{ni})}{\sigma(t_{ni})}} \mid f \in \mathcal{F} \right\} \right)$$

schwach gegen das Gaußexperiment E aus Satz 2.22, also

$$E_n^{loc, f_0} \xrightarrow{w} E = G_{H(R),\mathcal{F}}.$$

Die Experimente

$$G_n^{loc, f_0} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n N \left(\frac{f_0(t_{ni})}{\sigma(t_{ni})} + \frac{1}{\sqrt{n}} \cdot \frac{f(t_{ni})}{\sigma(t_{ni})}, \frac{1}{J} \right) \mid f \in \mathcal{F} \right\} \right)$$

konvergieren ebenfalls schwach gegen das Gaußexperiment E aus Satz 2.22. Man setze dort lediglich für Q die $N(0, \frac{1}{J})$ -Verteilung ein, die nach Beispiel 2.10 1. ebenfalls die Fisher-Information J besitzt. Mit Bemerkung 1.59 folgt

Korollar 2.25 *Die Experimentfolgen $(E_n^{loc, f_0})_{n \in \mathbb{N}}$ und $(G_n^{loc, f_0})_{n \in \mathbb{N}}$ sind schwach asymptotisch äquivalent. Diese Äquivalenzaussage ist insoweit unabhängig von der Fehlerverteilung Q , als sie für alle Fehlerverteilungen gilt, die den Voraussetzungen 2.9 genügen und die dieselbe Fisher-Information J haben.*

2.4.4 Δ -Konvergenz Linearer Regressionsmodelle mit Lokalisierungsrate $\frac{1}{\sqrt{n}}$

In Abschnitt 2.4.2 haben wir die schwache Konvergenz Linearer Regressionsmodelle gegen einen Standard-Gaußhft nachgewiesen. Es stellt sich die Frage, ob nicht nur schwache sondern sogar die starke Δ -Konvergenz vorliegt. Dies wird im Allgemeinen nur unter stärkeren Voraussetzungen der Fall sein. Man muss entweder

1. den Parameterraum einschränken - Grundlage für derartige Überlegungen ist der Satz von Lindae 1.53 - oder
2. die Klasse der Fehlerverteilungen einschränken; wir untersuchen den Fall normalverteilter Fehler.

Es sei jedoch erwähnt, dass die folgenden Ausführungen zur Δ -Konvergenz lokalisierter Regressionsmodelle im weiteren Verlauf der vorliegenden Arbeit keine entscheidende Rolle spielen und bei einem ersten Lesen übergangen werden können. Sie werden hier dennoch erwähnt, denn einerseits liefern sie weitergehende Einblicke in die Theorie der Experimente, und andererseits sind sie Ausgangspunkt für die in Abschnitt 2.5 folgende globale Theorie.

Wir betrachten in diesem Abschnitt 2.4.4 das lokalisierte Lineare Regressionsmodell

$$E_n^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{1}{\sqrt{n}} \cdot \frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})}} \mid \beta \in \mathbb{R}^k \right\} \right)$$

aus (2.24) auf Seite 52. Die folgenden Sätze sind starke Varianten von Satz 2.16. Wir setzen dabei alle Bezeichnungen dieses Satzes voraus, insbesondere seien die Matrizen

$$B = \left(\int_{[a,b]} \frac{f_i f_j}{\sigma^2} dH \right)_{1 \leq i, j \leq k} \quad \text{bzw.} \quad B_n = \left(\int_{[a,b]} \frac{f_i f_j}{\sigma^2} dH_n \right)_{1 \leq i, j \leq k}, \quad n \in \mathbb{N},$$

wie in (2.26) bzw. (2.28) definiert.

2.4.4.1 Beschränkter Parameterraum

Zunächst erhalten wir eine starke Variante von Satz 2.16, indem wir den Parameterraum verkleinern.

Satz 2.26 *Es gelten die Voraussetzungen 2.9 an die Fehlerverteilung und die folgenden Voraussetzungen an die Regressoren, die Varianzfunktion und die Designfolge:*

1. *Es existiere ein Grenzdesign H , gegen das die Designfolge H_n im Sinne von (2.11) oder (2.12) konvergiert,*
2. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien allesamt beschränkt und bilden eine H_n - H -Konvergenzklasse,*
3. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien linear unabhängig in $L^2(H)$.*

Weiter seien $\Theta \subset \mathbb{R}^k$ beschränkt und

$$E_{n,\Theta}^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{1}{\sqrt{n}} \cdot \frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})}} \mid \beta \in \Theta \right\} \right)$$

die auf den Parameterraum Θ eingeschränkten Experimente E_n^{loc} aus (2.24). Ebenso sei $G_{H_{JB},\Theta}$ der auf Θ eingeschränkte Standard-Gaußshft $G_{H_{JB}}$, der (analog zu Satz 2.16) äquivalent zum Experiment

$$G_\Theta := \left(\mathbb{R}^k, \mathcal{B}^k, \left\{ N\left(\sqrt{J}B^{\frac{1}{2}}\beta, I_k\right) \mid \beta \in \Theta \right\} \right)$$

ist. Dann gilt

$$\Delta(E_{n,\Theta}^{loc}, G_{H_{JB},\Theta}) \xrightarrow{n \rightarrow \infty} 0 \quad \text{bzw.} \quad \Delta(E_{n,\Theta}^{loc}, G_\Theta) \xrightarrow{n \rightarrow \infty} 0.$$

Beweis. Satz 2.16 liefert uns die schwache Konvergenz

$$E_{n,\Theta}^{loc} \xrightarrow{w} G_\Theta.$$

Um die Behauptung zu beweisen, müssen wir die beiden Voraussetzungen des Satzes von Lindae 1.53 nachprüfen. Da die Begriffe *beschränkt* und *totalbeschränkt* im \mathbb{R}^k übereinstimmen, folgt mit Lemma 1.55 direkt dessen erste Voraussetzung. Zum Nachweis der zweiten Voraussetzung benötigen wir zusätzliche Überlegungen.

Da die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ nach Voraussetzung beschränkt sind und da Θ beschränkt ist, existiert ein $M > 0$ mit

$$\frac{f^\top(t)(\beta_1 - \beta_2)}{\sigma(t)} \leq M \quad \text{für alle } t \in [a, b] \text{ und alle } \beta_1, \beta_2 \in \Theta.$$

In Abschnitt 2.4.1.1 haben wir gesehen, dass unter den Voraussetzungen 2.9 an die Fehlerverteilung

$$\lim_{\substack{s \rightarrow 0 \\ s \neq 0}} \frac{d_H^2(Q, Q_s)}{s^2} = \frac{J}{8}$$

gilt. Insgesamt folgt

$$\frac{d_H^2 \left(Q, Q_{\frac{f^\top(t)(\beta_1 - \beta_2)}{\sqrt{n\sigma(t)}}} \right)}{\left(\frac{f^\top(t)(\beta_1 - \beta_2)}{\sqrt{n\sigma(t)}} \right)^2} \xrightarrow{n \rightarrow \infty} \frac{J}{8} \quad \text{gleichmäßig in } t \in [a, b], \beta_1, \beta_2 \in \Theta, \beta_1 \neq \beta_2. \quad (2.40)$$

Ebenso gilt wegen der elementweisen Konvergenz der Matrizen B_n gegen B und der Beschränktheit von Θ

$$J \sum_{i=1}^n \left(\frac{f^\top(t_{ni})(\beta_1 - \beta_2)}{\sqrt{n\sigma(t_{ni})}} \right)^2 = J(\beta_1 - \beta_2)^\top B_n (\beta_1 - \beta_2) \xrightarrow{n \rightarrow \infty} \|\beta_1 - \beta_2\|_{JB}^2 \quad (2.41)$$

gleichmäßig in $\beta_1, \beta_2 \in \Theta$. Aus (2.40) und (2.41) folgt unter Verwendung von (A.6) und Beispiel 1.16

$$\begin{aligned} d_H^2 \left(\bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta_1}{\sqrt{n\sigma(t_{ni})}}}, \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta_2}{\sqrt{n\sigma(t_{ni})}}} \right) &\stackrel{(A.6)}{=} 1 - \prod_{i=1}^n \left(1 - d_H^2 \left(Q_{\frac{f^\top(t_{ni})\beta_1}{\sqrt{n\sigma(t_{ni})}}, Q_{\frac{f^\top(t_{ni})\beta_2}{\sqrt{n\sigma(t_{ni})}}} \right) \right) \\ &= 1 - \prod_{i=1}^n \left(1 - d_H^2 \left(Q, Q_{\frac{f^\top(t_{ni})(\beta_1 - \beta_2)}{\sqrt{n\sigma(t_{ni})}}} \right) \right) \end{aligned} \quad (2.42)$$

$$= 1 - \exp \left\{ \sum_{i=1}^n \ln \left(1 - d_H^2 \left(Q, Q_{\frac{f^\top(t_{ni})(\beta_1 - \beta_2)}{\sqrt{n\sigma(t_{ni})}}} \right) \right) \right\} \quad (2.43)$$

$$= 1 - \exp \left\{ \sum_{i=1}^n -d_H^2 \left(Q, Q_{\frac{f^\top(t_{ni})(\beta_1 - \beta_2)}{\sqrt{n\sigma(t_{ni})}}} \right) + \frac{1}{2} d_H^4 \left(Q, Q_{\frac{f^\top(t_{ni})(\beta_1 - \beta_2)}{\sqrt{n\sigma(t_{ni})}}} \right) \mp \dots \right\} \quad (2.44)$$

$$\stackrel{(2.40)}{=} 1 - \exp \left\{ - \sum_{i=1}^n \frac{d_H^2 \left(Q, Q_{\frac{f^\top(t_{ni})(\beta_1 - \beta_2)}{\sqrt{n\sigma(t_{ni})}}} \right)}{\left(\frac{f^\top(t_{ni})(\beta_1 - \beta_2)}{\sqrt{n\sigma(t_{ni})}} \right)^2} \left(\frac{f^\top(t_{ni})(\beta_1 - \beta_2)}{\sqrt{n\sigma(t_{ni})}} \right)^2 + O \left(\frac{1}{n^2} \right) \right\}$$

$$\xrightarrow{n \rightarrow \infty} 1 - \exp \left\{ - \frac{1}{8} \|\beta_1 - \beta_2\|_{JB}^2 \right\} \stackrel{1.16}{=} d_H^2 \left(N \left(\sqrt{JB}^{\frac{1}{2}} \beta_1, I_k \right), N \left(\sqrt{JB}^{\frac{1}{2}} \beta_2, I_k \right) \right)$$

gleichmäßig in $\beta_1, \beta_2 \in \Theta$. Dabei folgt (2.42) mit der Translationsinvarianz des Lebesgue-Maßes aus der Definition des Hellingerabstandes, (2.43) ist erst für hinreichend großes n definiert, und bei (2.44) wurde die Potenzreihenentwicklung des Logarithmus verwendet. Somit ist (1.22) erfüllt, und die Behauptung folgt aus dem Satz von Lindae. \square

2.4.4.2 Normalverteilungsannahme

Jetzt werden wir als Parameterraum wieder den kompletten \mathbb{R}^k zulassen, dafür aber als Fehlerverteilung $Q := N(0, 1)$ wählen. Damit erhalten wir die lokalisierten Experimente, vgl. Beispiel 2.1,

$$\mathcal{N}\left(\frac{1}{\sqrt{n}}A_n, \Sigma_n^2\right)$$

mit

$$A_n^\top = \begin{pmatrix} f_1(t_{n1}) & \cdots & f_k(t_{n1}) \\ \vdots & \ddots & \vdots \\ f_1(t_{nn}) & \cdots & f_k(t_{nn}) \end{pmatrix}, \quad \Sigma_n^2 = \begin{pmatrix} \sigma^2(t_{n1}) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(t_{nn}) \end{pmatrix}$$

und dem Parameterraum \mathbb{R}^k .

Satz 2.27 *Unter den Voraussetzungen*

1. *es existiere ein Grenzdesign H , gegen das die Designfolge H_n im Sinne von (2.11) oder (2.12) konvergiert,*
2. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ bilden eine H_n - H -Konvergenzklasse,*
3. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien linear unabhängig in $L^2(H)$*

gilt

$$\Delta\left(\mathcal{N}\left(\frac{1}{\sqrt{n}}A_n, \Sigma_n^2\right), \mathcal{N}\left(B^{\frac{1}{2}}, I_k\right)\right) \xrightarrow{n \rightarrow \infty} 0$$

bzw.

$$\Delta\left(\mathcal{N}\left(\frac{1}{\sqrt{n}}A_n, \Sigma_n^2\right), \mathcal{N}(B, B)\right) \xrightarrow{n \rightarrow \infty} 0.$$

Beweis. Wir beweisen zunächst die erste Aussage. Durch eine suffiziente Transformation, vgl. Beispiel 2.1, erhalten wir

$$\Delta\left(\mathcal{N}\left(\frac{1}{\sqrt{n}}A_n, \Sigma_n^2\right), \mathcal{N}\left(\frac{1}{\sqrt{n}}A_n \Sigma_n^{-1}, I_n\right)\right) = 0.$$

Nun können wir Satz 1.11 auf $\mathcal{N}\left(\frac{1}{\sqrt{n}}A_n\Sigma_n^{-1}, I_n\right)$ und $\mathcal{N}\left(B^{\frac{1}{2}}, I_k\right)$ anwenden. Dabei benötigen wir

$$\begin{aligned} A_n\Sigma_n^{-1}(A_n\Sigma_n^{-1})^\top &= A_n\Sigma_n^{-2}A_n^\top = \left(\sum_{i=1}^n \frac{f_j(t_{ni})f_l(t_{ni})}{\sigma^2(t_{ni})}\right)_{1 \leq j, l \leq k} \\ &= n \left(\int_{[a, b]} \frac{f_j f_l}{\sigma^2} dH_n\right)_{1 \leq j, l \leq k} = nB_n. \end{aligned} \quad (2.45)$$

Da die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ eine H_n - H -Konvergenzklasse bilden, konvergiert die Matrix B_n elementweise gegen B . Wegen Ungleichung (A.3) können wir in Satz 1.11 an Stelle des Variationsabstandes auch den Hellingerabstand verwenden, mit Beispiel A.13 sowie der Stetigkeit der Determinanten folgt

$$\begin{aligned} d_H^2 \left(N \left(0, B^{\frac{1}{2}} \left(B^{\frac{1}{2}} \right)^\top \right), N \left(0, \frac{1}{\sqrt{n}} A_n \Sigma_n^{-1} \left(\frac{1}{\sqrt{n}} A_n \Sigma_n^{-1} \right)^\top \right) \right) \\ = d_H^2 (N(0, B), N(0, B_n)) = 1 - 2^{\frac{k}{2}} \cdot \frac{(\det B)^{\frac{1}{4}} (\det B_n)^{\frac{1}{4}}}{(\det(B + B_n))^{\frac{1}{2}}} \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

und schließlich mit Satz 1.11

$$\Delta \left(\mathcal{N} \left(\frac{1}{\sqrt{n}} A_n \Sigma_n^{-1}, I_n \right), \mathcal{N} \left(B^{\frac{1}{2}}, I_k \right) \right) \xrightarrow{n \rightarrow \infty} 0.$$

Die zweite Behauptung ergibt sich mit Hilfe der ersten aus

$$\Delta \left(\mathcal{N} \left(B^{\frac{1}{2}}, I_k \right), \mathcal{N} (B, B) \right) = 0,$$

was wiederum mittels einer suffizienten Transformation (analog zu Beispiel 2.1) folgt. \square

2.4.5 Δ -Konvergenz allgemeiner Regressionsmodelle mit Lokalisierungsrate $\frac{1}{\sqrt{n}}$

Wie im Linearen Regressionsmodell wollen wir nun auch im allgemeinen Regressionsmodell die Δ -Konvergenz lokalisierter Experimente nachweisen. Die folgenden Sätze sind starke Varianten von Satz 2.22. Wir verwenden alle Bezeichnungen dieses Satzes.

2.4.5.1 Totalbeschränkter Parameterraum

Zunächst werden wir uns auf totalbeschränkte Teilmengen von $\frac{1}{\sigma}\mathcal{F} \subset L^2(H)$ zurückziehen, also im Wesentlichen auf kompakte Mengen. Dies ist eine sehr starke Einschränkung, da kompakte Mengen in unendlichdimensionalen Räumen sehr klein sind.

Satz 2.28 *Es gelten die Voraussetzungen (2.9) an die Fehlerverteilung Q . Weiter setzen wir voraus:*

1. *Es existiere ein Grenzdesign H , gegen das die Designfolge H_n im Sinne von (2.11) oder (2.12) konvergiert,*
2. *die Klasse $\frac{1}{\sigma}\mathcal{F} := \{\frac{f}{\sigma} \mid f \in \mathcal{F}\}$ sei eine H_n - H -Konvergenzklasse.*

Weiter sei $\Theta \subset \mathcal{F}$ so, dass $\frac{1}{\sigma}\Theta$ eine totalbeschränkte Teilmenge von $L^2(H)$ ist, und es existiere ein $M > 0$ mit

$$\frac{f(t)}{\sigma(t)} \leq M \quad \text{für alle } t \in [a, b] \text{ und alle } f \in \Theta,$$

d.h. die Funktionen in Θ seien gleichmäßig beschränkt. Es seien

$$E_{n,\Theta}^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{1}{\sqrt{n}} \cdot \frac{f(t_{ni})}{\sigma(t_{ni})}} \mid f \in \Theta \right\} \right)$$

die auf Θ eingeschränkten Experimente E_n^{loc} aus Satz 2.22 und

$$E_{\Theta} = (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{Z_f} \mid f \in \mathcal{F}\}).$$

das auf Θ eingeschränkte Experiment E aus Satz 2.22. Dann gilt

$$\Delta(E_{n,\Theta}^{loc}, E_{\Theta}) \xrightarrow{n \rightarrow \infty} 0.$$

Beweis. Der Beweis verläuft analog zum Beweis von Satz 2.26. □

2.4.5.2 Normalverteilungsannahme

In diesem Abschnitt werden wir die Normalverteilung als Fehlerverteilung voraussetzen, also $Q := N(0, 1)$ mit Fisher-Information $J = 1$. Im Gegensatz zum Linearen Regressionsmodell wird dies beim allgemeinen Regressionsmodell jedoch nicht genügen, um Δ -Konvergenz nachzuweisen. Vielmehr müssen wir auch noch den Parameterraum einschränken. Dabei betrachten wir speziell Räume von Funktionen von gleichmäßig beschränkter Variation. Desweiteren konzentrieren wir uns auf H -äquidistante Designfolgen H_n , vgl. Beispiel 2.3.

Satz 2.29 *Unter den Voraussetzungen*

1. *H_n sei eine H -äquidistante Designfolge, also $t_{ni} = H^{-1}(\frac{i}{n})$, $i = 1, \dots, n$,*

2. $C > 0$ sei beliebig, aber fest vorgegeben, und der Parameterraum sei

$$\mathcal{F}_C := \left\{ f : [a, b] \rightarrow \mathbb{R} \mid \frac{f}{\sigma} \in BV[a, b]; V\left(\frac{f}{\sigma}\right) \leq C \right\}, \quad (2.46)$$

wobei $V(\cdot)$ die Totalvariation bezeichne,

gilt für die Experimente

$$F_n^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n N\left(\frac{1}{\sqrt{n}}f(t_{ni}), \sigma^2(t_{ni})\right) \mid f \in \mathcal{F}_C \right\} \right)$$

und (wie in Satz 2.22)

$$E = (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{Z_f} \mid f \in \mathcal{F}_C\})$$

mit

$$Z_f(s) = \int_{[0, s]} \frac{f \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} \lambda(dx) + W(s), \quad s \in [0, 1],$$

die starke Konvergenz

$$\Delta(F_n^{loc}, E) \xrightarrow{n \rightarrow \infty} 0.$$

Wir beweisen dieses Ergebnis mit Hilfe des folgenden Satzes.

Satz 2.30 Sei $C > 0$ fest vorgegeben und

$$\mathcal{G}_C := \{g : [0, 1] \rightarrow \mathbb{R} \mid g \in BV[0, 1]; V(g) \leq C\}.$$

Dann gilt für die Experimente

$$G_n^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n N\left(\frac{1}{\sqrt{n}}g\left(\frac{i}{n}\right), 1\right) \mid g \in \mathcal{G}_C \right\} \right)$$

und

$$G = (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{X_g} \mid g \in \mathcal{G}_C\})$$

mit

$$X_g(s) = \int_{[0, s]} g(x) \lambda(dx) + W(s), \quad s \in [0, 1],$$

die starke Konvergenz

$$\Delta(G_n^{loc}, G) \xrightarrow{n \rightarrow \infty} 0.$$

Beweis. Im ersten Schritt werden wir die Regressoren $g \in \mathcal{G}_C$ diskretisieren. Zu g definieren wir die Funktionen $g_n : [0, 1] \rightarrow \mathbb{R}$ durch

$$g_n(t) := \sum_{i=1}^{n-1} g\left(\frac{i}{n}\right) 1_{[\frac{i-1}{n}, \frac{i}{n})}(t) + g(1) 1_{[\frac{n-1}{n}, 1]}(t), \quad t \in [0, 1],$$

und damit die stochastischen Prozesse

$$X_{g_n}(s) = \int_{[0, s]} g_n(x) \lambda(dx) + W(s), \quad s \in [0, 1],$$

sowie die Experimente

$$\tilde{G}_n = (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{X_{g_n}} \mid g \in \mathcal{G}_C\}).$$

Mit Korollar 2.18 (setze $H = id_{[0, 1]}$ und $\sigma \equiv 1$) folgt

$$d_H^2(P^{X_g}, P^{X_{g_n}}) = 1 - \exp\left\{-\frac{1}{8}\|g - g_n\|_{L^2(\lambda)}^2\right\},$$

und damit gelten die Äquivalenzen

$$\begin{aligned} & \sup \{d_H(P^{X_g}, P^{X_{g_n}}) \mid g \in \mathcal{G}_C\} \rightarrow 0 \\ \iff & \sup \left\{1 - \exp\left\{-\frac{1}{8}\|g - g_n\|_{L^2(\lambda)}^2\right\} \mid g \in \mathcal{G}_C\right\} \rightarrow 0 \\ \iff & \inf \left\{\exp\left\{-\frac{1}{8}\|g - g_n\|_{L^2(\lambda)}^2\right\} \mid g \in \mathcal{G}_C\right\} \rightarrow 1 \\ \iff & \sup \left\{\|g - g_n\|_{L^2(\lambda)}^2 \mid g \in \mathcal{G}_C\right\} \rightarrow 0 \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Nach Satz A.19 ist

$$\sup \left\{\|g - g_n\|_{L^2(\lambda)}^2 \mid g \in \mathcal{G}_C\right\} \leq \frac{C^3 + C}{n},$$

und da nach Lemma 1.6 6. die Ungleichung

$$\Delta(G, \tilde{G}_n) \leq \sqrt{2} \sup \{d_H(P^{X_g}, P^{X_{g_n}}) \mid g \in \mathcal{G}_C\}$$

gilt, folgt

$$\Delta(G, \tilde{G}_n) \rightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (2.47)$$

Im zweiten Schritt werden wir mittels einer suffizienten Abbildung nachweisen, dass die Experimente \tilde{G}_n und G_n^{loc} für jedes $n \in \mathbb{N}$ äquivalent sind. Die Funktionen g_n sind nach

Konstruktion rechtsseitig stetig sowie von beschränkter Variation. Mit Formel (A.14) und Beispiel A.25 aus Anhang A.6 folgt für $u \in C[0, 1]$

$$\begin{aligned} \frac{dP^{X_{g_n}}}{dP^W}(u) &= \exp \left\{ \int_0^1 g_n(x) u(dx) - \frac{1}{2} \|g_n\|_{L^2(\lambda)}^2 \right\} \\ &= \exp \left\{ \sum_{i=1}^n g \left(\frac{i}{n} \right) \left[u \left(\frac{i}{n} \right) - u \left(\frac{i-1}{n} \right) \right] \right\} \exp \left\{ -\frac{1}{2} \|g_n\|_{L^2(\lambda)}^2 \right\}. \end{aligned}$$

Definieren wir

$$\gamma_{ni}(g) := g \left(\frac{i}{n} \right), \quad R_{ni}(u) := u \left(\frac{i}{n} \right) - u \left(\frac{i-1}{n} \right), \quad A_n(g) := \exp \left\{ -\frac{1}{2} \|g_n\|_{L^2(\lambda)}^2 \right\},$$

so erhalten wir

$$\frac{dP^{X_{g_n}}}{dP^W}(u) = A_n(g) \exp \left\{ \sum_{i=1}^n \gamma_{ni}(g) R_{ni}(u) \right\},$$

die Maße $\{P^{X_{g_n}} \mid g \in \mathcal{G}_C\}$ bilden also eine Exponentialfamilie. Nach Witting und Müller-Funk (1995), Korollar 3.20 und Satz 3.10, ist somit die Statistik $R_n : C[0, 1] \rightarrow \mathbb{R}^n$,

$$\begin{aligned} R_n(u) &= (\sqrt{n} R_{n1}(u), \dots, \sqrt{n} R_{nn}(u))^\top \\ &= \left(\sqrt{n} \left[u \left(\frac{1}{n} \right) - u(0) \right], \dots, \sqrt{n} \left[u(1) - u \left(\frac{n-1}{n} \right) \right] \right)^\top, \quad u \in C[0, 1], \end{aligned}$$

suffizient für \tilde{G}_n . Wir wenden R_n auf dieses Experiment an und erhalten als i -te Komponente, $i = 1, \dots, n$:

$$\begin{aligned} \sqrt{n} \left[X_{g_n} \left(\frac{i}{n} \right) - X_{g_n} \left(\frac{i-1}{n} \right) \right] &= \\ &= \sqrt{n} \left[\int_0^{\frac{i}{n}} g_n(x) \lambda(dx) + W \left(\frac{i}{n} \right) - \int_0^{\frac{i-1}{n}} g_n(x) \lambda(dx) + W \left(\frac{i-1}{n} \right) \right] \\ &= \sqrt{n} \left[\int_{\frac{i-1}{n}}^{\frac{i}{n}} g_n(x) \lambda(dx) + W \left(\frac{i}{n} \right) - W \left(\frac{i-1}{n} \right) \right] \\ &= \frac{1}{\sqrt{n}} g \left(\frac{i}{n} \right) + \sqrt{n} \left[W \left(\frac{i}{n} \right) - W \left(\frac{i-1}{n} \right) \right] \\ &\sim N \left(\frac{1}{\sqrt{n}} g \left(\frac{i}{n} \right), 1 \right). \end{aligned}$$

Da die Brownsche Bewegung unabhängige Zuwächse besitzt, geht für $g \in \mathcal{G}_C$ das Maß $P^{X_{g_n}}$ durch Anwendung von R_n über in

$$\bigotimes_{i=1}^n N \left(\frac{1}{\sqrt{n}} g \left(\frac{i}{n} \right), 1 \right),$$

d.h. \tilde{G}_n geht über in G_n^{loc} . Mit Satz 1.10 folgt $\Delta(\tilde{G}_n, G_n^{loc}) = 0$ und zusammen mit (2.47) die Behauptung. \square

Beweis von Satz 2.29. In Abschnitt 2.1 wurde mittels suffizienter Transformationen die Äquivalenz $\Delta(F_n^{loc}, E_n^{loc}) = 0$, $n \in \mathbb{N}$, gezeigt mit

$$\begin{aligned} E_n^{loc} &= \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n N \left(\frac{f(t_{ni})}{\sqrt{n}\sigma(t_{ni})}, 1 \right) \mid f \in \mathcal{F}_C \right\} \right) \\ &= \left(\left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n N \left(\frac{1}{\sqrt{n}} \cdot \frac{f \circ H^{-1} \left(\frac{i}{n} \right)}{\sigma \circ H^{-1} \left(\frac{i}{n} \right)}, 1 \right) \mid f \in \mathcal{F}_C \right\} \right) \right). \end{aligned}$$

Ist \mathcal{G}_C der Parameterraum aus Satz 2.30, so gilt nach Voraussetzung und nach Satz A.20

$$\frac{f \circ H^{-1}}{\sigma \circ H^{-1}} \in \mathcal{G}_C \quad \text{für alle } f \in \mathcal{F}_C.$$

Wir erhalten die Experimente E_n^{loc} somit aus den Experimenten G_n^{loc} von Satz 2.30 durch Einschränken des Parameterraums und analog das Experiment E aus G . Jetzt folgt aus der Definition der Δ -Metrik 1.5 und Satz 2.30

$$\Delta(E_n^{loc}, E) \leq \Delta(G_n^{loc}, G) \xrightarrow{n \rightarrow \infty} 0$$

und damit die Behauptung. \square

2.5 Globale Theorie - Asymptotische Äquivalenz unter Normalverteilungsannahme

In dem letzten Abschnitt von Kapitel 2 wollen wir eines der lokalen Ergebnisse aus 2.4 „globalisieren“. Wie wir bereits zu Beginn von Abschnitt 2.4.1 gesehen haben, werden globale, d.h. nicht lokalisierte Experimente im Allgemeinen schwach gegen ein vollständig informatives Experiment konvergieren. Deshalb erweist sich in diesem Zusammenhang der Begriff der „starken asymptotischen Äquivalenz“ von Experimenten aus Definition 1.56 als adäquat. An Stelle eines Grenzexperiments werden wir nun eine Folge von „Grenz-“experimenten angeben, deren Abstand zu den ursprünglichen Experimenten in der „starken“ Δ -Metrik gegen Null konvergiert. Während die Konvergenz von lokalisierten Experimenten vor allem in der Testtheorie eine große Rolle spielt, werden wir die globale Theorie in unseren Überlegungen zur Designtheorie in Kapitel 4 nutzen. Der vorliegende Abschnitt beleuchtet allerdings nur einen sehr kleinen Teil der globalen Theorie.

2.5.1 Ein asymptotisches Äquivalent zum Linearen Regressionsmodell unter Normalverteilungsannahme

Der folgende Satz 2.31 beinhaltet die Globalisierung von Satz 2.27 und behandelt die starke Konvergenz lokalisierter Linearer Regressionsmodelle unter Normalverteilungsannahme.

Satz 2.31 *Unter den Voraussetzungen*

1. *es existiere ein Grenzdesign H , gegen das die Designfolge H_n im Sinne von (2.11) oder (2.12) konvergiert,*
2. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ bilden eine H_n - H -Konvergenzklasse,*
3. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien linear unabhängig in $L^2(H)$*

sind die Experimentfolgen $(\mathcal{N}(A_n, \Sigma_n^2))_{n \in \mathbb{N}}$ und $(\mathcal{N}(\sqrt{n}B^{\frac{1}{2}}, I_k))_{n \in \mathbb{N}}$ (beziehungsweise $(\mathcal{N}(B, \frac{1}{n}B))_{n \in \mathbb{N}}$) asymptotisch äquivalent, d.h.

$$\Delta \left(\mathcal{N}(A_n, \Sigma_n^2), \mathcal{N}(\sqrt{n}B^{\frac{1}{2}}, I_k) \right) \xrightarrow{n \rightarrow \infty} 0$$

bzw.

$$\Delta \left(\mathcal{N}(A_n, \Sigma_n^2), \mathcal{N}\left(B, \frac{1}{n}B\right) \right) \xrightarrow{n \rightarrow \infty} 0.$$

Beweis. Wir könnten die beiden Aussagen analog zu Satz 2.27 beweisen, wollen sie aber stattdessen aus diesem Satz herleiten. Dies geschieht mit Hilfe des anschließenden Lemmas 2.32. \square

Lemma 2.32 *Es seien $\mathcal{N}(C_n, I_{\mathcal{X}_{1n}})$, $\mathcal{N}(D_n, I_{\mathcal{X}_{2n}})$, $n \in \mathbb{N}$, Lineare Modelle mit gemeinsamem Parameterraum \mathbb{R}^k und reellen Beobachtungsräumen $\mathcal{X}_{1n}, \mathcal{X}_{2n}$, und $(\delta_n)_{n \in \mathbb{N}}$ sei eine beliebige reelle Zahlenfolge mit $\delta_n \neq 0$, $n \in \mathbb{N}$. Dann gilt die Äquivalenz*

$$\Delta(\mathcal{N}(\delta_n C_n, I_{\mathcal{X}_{1n}}), \mathcal{N}(D_n, I_{\mathcal{X}_{2n}})) \xrightarrow{n \rightarrow \infty} 0 \iff \Delta\left(\mathcal{N}(C_n, I_{\mathcal{X}_{1n}}), \mathcal{N}\left(\frac{1}{\delta_n} D_n, I_{\mathcal{X}_{2n}}\right)\right) \xrightarrow{n \rightarrow \infty} 0.$$

Beweis. Der Beweis beruht auf Satz 1.11, Ungleichung (A.3) und Beispiel A.13:

$$\begin{aligned} \Delta(\mathcal{N}(\delta_n C_n, I_{\mathcal{X}_{1n}}), \mathcal{N}(D_n, I_{\mathcal{X}_{2n}})) \xrightarrow{n \rightarrow \infty} 0 &\iff \\ &\stackrel{1.11}{\iff} \left\| N(0, \delta_n^2 C_n C_n^\top) - N(0, D_n D_n^\top) \right\|_V \xrightarrow{n \rightarrow \infty} 0 \\ &\stackrel{(A.3)}{\iff} d_H^2(N(0, \delta_n^2 C_n C_n^\top) - N(0, D_n D_n^\top)) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

$$\begin{aligned}
&\stackrel{A.13}{\iff} 1 - 2^{\frac{k}{2}} \cdot \frac{(\det \delta_n^2 C_n C_n^\top)^{\frac{1}{4}} (\det D_n D_n^\top)^{\frac{1}{4}}}{(\det(\delta_n^2 C_n C_n^\top + D_n D_n^\top))^{\frac{1}{2}}} \xrightarrow{n \rightarrow \infty} 0 \\
&\iff 1 - 2^{\frac{k}{2}} \cdot \frac{(\det C_n C_n^\top)^{\frac{1}{4}} (\det \frac{1}{\delta_n^2} D_n D_n^\top)^{\frac{1}{4}}}{(\det(C_n C_n^\top + \frac{1}{\delta_n^2} D_n D_n^\top))^{\frac{1}{2}}} \xrightarrow{n \rightarrow \infty} 0 \\
&\iff \Delta \left(\mathcal{N}(C_n, I_{\mathcal{X}_{1n}}), \mathcal{N}\left(\frac{1}{\delta_n} D_n, I_{\mathcal{X}_{2n}}\right) \right) \xrightarrow{n \rightarrow \infty} 0.
\end{aligned}$$

□

2.5.2 Ausblick

An dieser Stelle wollen wir die Betrachtung der Asymptotik von Regressionsmodellen abschließen und zu den Anwendungen der bisherigen Ergebnisse kommen. Es soll noch auf weitere globale Ergebnisse hingewiesen werden, deren Beweis wir allerdings schuldig bleiben.

Zum einen ist zu erwarten, dass sich analog zu Satz 2.27 auch Satz 2.26 globalisieren lässt. Ein ganz ähnliches Resultat wurde bereits in der relativ unbekanntenen Arbeit LeCam (1985) bewiesen. Es ist zu prüfen, ob die dortigen Voraussetzungen auch in unserem Fall erfüllt sind und LeCams Resultat (dort Théorème 4.3) anwendbar ist oder ob zumindest ein ähnlicher Beweis wie dort durchgeführt werden kann.

Zum anderen scheint auch die Globalisierung von Satz 2.29 möglich. Ein solches Ergebnis würde dann eine Verallgemeinerung der Arbeit Brown und Low (1996) darstellen, wo nur Grenzdesigns betrachtet werden, die äquivalent zum Lebesgue-Maß auf dem Versuchsbereich sind. Es zeigt sich jedoch, dass die Techniken, die Brown und Low verwenden, nicht direkt übertragbar sind. Vielmehr scheint es sinnvoll, zuerst eine lokale Version ähnlich wie in Satz 2.29 zu beweisen und diese dann mittels eines „preliminary estimator“ zu globalisieren - eine Technik, wie sie zum Beispiel in der Arbeit Nussbaum (1996) verwendet wird.

Kapitel 3

Anwendungen in der Testtheorie

In diesem Kapitel sollen aus den Konvergenzaussagen von Kapitel 2 asymptotische Aussagen für Testprobleme der Form

$$H_\ell = \{\theta \in \Theta \mid \ell(\theta) \leq 0\}, \quad K_\ell = \{\theta \in \Theta \mid \ell(\theta) > 0\}$$

gewonnen werden, wobei $\ell : \Theta \rightarrow \mathbb{R}$ eine gegebene stetige lineare Abbildung vom Parameterraum Θ nach \mathbb{R} ist. Es genügt bereits die *schwache* Konvergenz lokalisierter Experimente, um asymptotisch optimale Testfolgen sowohl für Lineare als auch für allgemeine Regressionsmodelle angeben zu können. Wir verwenden im Folgenden die Bezeichnungen und Definitionen von Abschnitt 1.8.1.

Im ersten Abschnitt 3.1 des vorliegenden Kapitels betrachten wir ausschließlich Lineare Regressionsmodelle. Zunächst geben wir eine auf Satz 1.49 basierende asymptotisch optimale Testfolge für Testprobleme obiger Form an. Diese beruht entscheidend auf der Kenntnis der Fehlerverteilung. Anschließend untersuchen wir eine auf dem Gauß-Markoff-Schätzer basierende Testfolge, die lediglich die Existenz zweiter Momente an die Fehlerverteilung nicht aber deren Kenntnis voraussetzt und unter Normalverteilungsannahme dem Likelihood-Quotienten-Test entspricht, und vergleichen sie mit der asymptotisch optimalen Testfolge bezüglich ihrer Asymptotischen Relativen Effizienz.

Im zweiten Abschnitt 3.2 von Kapitel 3 untersuchen wir Testprobleme obiger Form für allgemeine Regressionsmodelle und bestimmen eine auf Satz 1.52 basierende asymptotisch optimale Testfolge.

Aus mehreren Gründen werden in diesem Kapitel nur solche einfachen Testprobleme untersucht. Zum einen lassen diese „eindimensionalen“ Testprobleme Optimalitätsaussagen zu, ohne weitere Begriffe wie Invarianz bzw. asymptotische Invarianz zu benötigen. Zum anderen reichen sie aus, um die weitreichenden Folgerungen zu demonstrieren, die bereits die schwache Konvergenz im Sinne LeCam's ermöglicht.

3.1 Tests für das Lineare Regressionsmodell

Grundlage dieses Abschnitts ist die schwache Konvergenzaussage für Lineare Regressionsmodelle aus Satz 2.16, dessen Voraussetzungen erfüllt seien. Wir verwenden die dort auftretenden Bezeichnungen.

3.1.1 Ein asymptotisch optimaler Test φ_n^*

Der Parameterraum der Satz 2.16 zu Grunde liegenden Experimente E_n^{loc} , vgl. (2.24), ist der \mathbb{R}^k , der mit dem Innenprodukt $\langle \cdot, \cdot \rangle_{JB}$ zum Hilbertraum H_{JB} wird. Wir betrachten für fest vorgegebenes $\ell \in \mathbb{R}^k \setminus \{0\}$ das Testproblem

$$H_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta \leq 0\}, \quad K_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta > 0\}. \quad (3.1)$$

Grob gesagt interessiert man sich bei diesem Testproblem für eine Umgebung der Null, auf die man aus einer vorgegebenen Richtung ℓ zuläuft. Folglich wird man die zugehörigen Experimente um den Nullpunkt lokalisieren, was in Satz 2.16 bereits geschehen ist. (Natürlich kann an Stelle der Null jedes beliebige Element x_0 des \mathbb{R}^k stehen. Man benötigt dann lediglich ein um x_0 lokalisiertes Analogon zu Satz 2.16.)

Im Folgenden wollen wir Satz 1.49 anwenden. Das Testproblem (H_ℓ, K_ℓ) liegt allerdings noch nicht in der Form vor, wie sie in Satz 1.49 benötigt wird. Definieren wir aber

$$\gamma := \frac{1}{J} \cdot B^{-1} \ell \in \mathbb{R}^k \setminus \{0\},$$

so ist

$$\|\gamma\|_{JB} = \sqrt{\frac{\ell^\top B^{-1} \ell}{J}},$$

und für $\beta \in \mathbb{R}^k$ gilt

$$\langle \gamma, \beta \rangle_{JB} = J \gamma^\top B \beta = \ell^\top \beta.$$

Damit können wir das Testproblem (H_ℓ, K_ℓ) in der Form

$$H_\ell = \{\beta \in \mathbb{R}^k \mid \langle \gamma, \beta \rangle_{JB} \leq 0\}, \quad K_\ell = \{\beta \in \mathbb{R}^k \mid \langle \gamma, \beta \rangle_{JB} > 0\}$$

schreiben und Satz 1.49 direkt anwenden.

Satz 3.1 *Es mögen die Voraussetzungen von Satz 2.16 gelten. Für festes $\ell \in \mathbb{R}^k \setminus \{0\}$ betrachten wir das Testproblem*

$$H_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta \leq 0\}, \quad K_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta > 0\}.$$

Dann ist die Testfolge $(\varphi_n^)_{n \in \mathbb{N}}$ mit $\varphi_n^* : \mathbb{R}^n \rightarrow [0, 1]$,*

$$\varphi_n^*(x) = \begin{cases} 1, & \frac{1}{\sqrt{nJ\ell^\top B^{-1}\ell}} \sum_{i=1}^n \frac{f^\top(t_{ni})B^{-1}\ell}{\sigma(t_{ni})} \cdot \left(-\frac{q'(x_i)}{q(x_i)}\right) \cdot 1_{\{q>0\}}(x_i) > \Phi^{-1}(1-\alpha) \\ 0, & \frac{1}{\sqrt{nJ\ell^\top B^{-1}\ell}} \sum_{i=1}^n \frac{f^\top(t_{ni})B^{-1}\ell}{\sigma(t_{ni})} \cdot \left(-\frac{q'(x_i)}{q(x_i)}\right) \cdot 1_{\{q>0\}}(x_i) < \Phi^{-1}(1-\alpha) \end{cases},$$

$x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, asymptotisch optimal in der Klasse aller asymptotisch α -ähnlichen Testfolgen im Sinne von Definition 1.41 bzw. Bemerkung 1.42 mit einem fest vorgegebenen Niveau $\alpha \in (0, 1)$. Für die Folge der Gütefunktionen $(g_{\varphi_n^*})_{n \in \mathbb{N}}$ gilt

$$g^*(\beta) := \lim_{n \rightarrow \infty} g_{\varphi_n^*}(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \sqrt{J} \frac{\ell^\top \beta}{\sqrt{\ell^\top B^{-1} \ell}} \right) \quad \text{für alle } \beta \in \mathbb{R}^k.$$

Bemerkung 3.2 Die optimale Testfolge $(\varphi_n^*)_{n \in \mathbb{N}}$ aus Satz 3.1 hängt von der zu Grunde liegenden Fehlerverteilung Q ab. Zentraler Bestandteil der Teststatistik ist der Quotient $\frac{q'}{q}$, in dem die Lebesgue-Dichte q von Q auftritt. Einen Test, dessen Teststatistik von der Fehlerverteilung unabhängig ist, nennen wir einen **verteilungsfreien Test**. Die Tests φ_n^* , $n \in \mathbb{N}$, sind also *nicht verteilungsfrei*. Bemerkenswert ist aber, dass in die Grenzgüte dieser Tests lediglich die Fisher-Information J der Fehlerverteilung eingeht und die Dichte selbst dort keine Rolle mehr spielt.

3.1.2 Vergleich von φ_n^* mit dem Likelihood-Quotienten-Test bei normalverteiltem Fehler

Um die asymptotisch optimale Testfolge aus Satz 3.1 mit anderen Tests, die aus der Theorie Linearer Modelle bekannt sind, zu vergleichen, setzen wir als Fehlerverteilung zunächst $Q := N(0, 1)$ voraus. Dann gilt für das lokalisierte Lineare Modell E_n^{loc} aus Fromel (2.24) auf Seite 52

$$E_n^{loc} = \mathcal{N} \left(\frac{1}{\sqrt{n}} A_n \Sigma_n^{-1}, I_n \right)$$

mit

$$A_n^\top = \begin{pmatrix} f_1(t_{n1}) & \cdots & f_k(t_{n1}) \\ \vdots & \ddots & \vdots \\ f_1(t_{nn}) & \cdots & f_k(t_{nn}) \end{pmatrix}, \quad \Sigma_n^2 = \begin{pmatrix} \sigma^2(t_{n1}) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(t_{nn}) \end{pmatrix}$$

wie in den Formeln (2.8), (2.9) aus Beispiel 2.1. Setzen wir

$$X_n := (A_n \Sigma_n^{-1})^\top = \Sigma_n^{-1} A_n^\top, \quad (3.2)$$

so können wir dieses Modell auch in der Form

$$Z_n^{loc} = \frac{1}{\sqrt{n}} X_n \beta + \epsilon_n, \quad \beta \in \mathbb{R}^k,$$

schreiben, wobei $\epsilon_n \sim N(0, I_n)$ -verteilt ist.

Für diese normalverteilten Linearen Regressionsmodelle und für das Testproblem

$$H_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta \leq 0\}, \quad K_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta > 0\}$$

mit festem $\ell \in \mathbb{R}^k \setminus \{0\}$ heißt für jedes feste $n \in \mathbb{N}$ der Test $\varphi_n^{LQ} : \mathbb{R}^n \rightarrow [0, 1]$ mit

$$\varphi_n^{LQ}(x) = \begin{cases} 1, & \text{falls } \frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top x}{\sqrt{\ell^\top (X_n^\top X_n)^{-1} \ell}} > \Phi^{-1}(1 - \alpha) \\ 0, & \text{falls } \frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top x}{\sqrt{\ell^\top (X_n^\top X_n)^{-1} \ell}} < \Phi^{-1}(1 - \alpha) \end{cases}, \quad (3.3)$$

$x \in \mathbb{R}^n$, **Likelihood-Quotienten-Test** (oder kurz LQ-Test) für das Testproblem (H_ℓ, K_ℓ) . Für jedes feste $n \in \mathbb{N}$ und jedes Niveau $\alpha \in (0, 1)$ ist der Test φ_n^{LQ} im Sinne von Definition 1.40 optimal in der Klasse $\mathcal{F}_\alpha^s(H_\ell, K_\ell)$ aller α -ähnlichen Tests. Er besitzt die Güte

$$g_n^{LQ}(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\ell^\top \beta}{\sqrt{n \ell^\top (X_n^\top X_n)^{-1} \ell}} \right) \quad \text{für alle } \beta \in \mathbb{R}^k,$$

vgl. z.B. Stapleton (1995), S. 83 ff..

Wegen $n(X_n^\top X_n)^{-1} = B_n^{-1}$, vgl. (2.45) auf Seite 64, können wir die Teststatistik des LQ-Tests auch in der Form

$$\frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top x}{\sqrt{\ell^\top (X_n^\top X_n)^{-1} \ell}} = \frac{\ell^\top B_n^{-1} X_n^\top x}{\sqrt{n \ell^\top B_n^{-1} \ell}}, \quad x \in \mathbb{R}^n,$$

schreiben, und wegen $B_n^{-1} \rightarrow B^{-1}$ (für $n \rightarrow \infty$) elementweise ist

$$\lim_{n \rightarrow \infty} g_n^{LQ}(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\ell^\top \beta}{\sqrt{\ell^\top B^{-1} \ell}} \right) \quad \text{für alle } \beta \in \mathbb{R}^k.$$

Man beachte, dass die Ausdrücke $(X_n^\top X_n)^{-1}$ bzw. B_n^{-1} erst ab hinreichend großem n definiert sind.

Da nach Beispiel 2.10 die Fisher-Information der $N(0, 1)$ -Verteilung $J = 1$ ist, hat die Folge der LQ-Tests $(\varphi_n^{LQ})_{n \in \mathbb{N}}$ die gleiche Grenzgüte wie die Folge $(\varphi_n^*)_{n \in \mathbb{N}}$ aus Satz 3.1 und ist somit nach Definition 1.41 ebenfalls asymptotisch optimal in der Klasse aller asymptotisch α -ähnlichen Testfolgen.

Es stellt sich jetzt natürlich die Frage, ob die Tests φ_n^* und φ_n^{LQ} bereits für festes $n \in \mathbb{N}$ identisch sind. Dies ist zu verneinen. Denn setzt man in φ_n^* für q die Dichte der Standardnormalverteilung ein, so erhält man für $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$

$$\begin{aligned} & \frac{1}{\sqrt{n J \ell^\top B^{-1} \ell}} \sum_{i=1}^n \frac{f^\top(t_{ni}) B^{-1} \ell}{\sigma(t_{ni})} \cdot \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i) \\ &= \frac{1}{\sqrt{n \ell^\top B^{-1} \ell}} \sum_{i=1}^n \frac{f^\top(t_{ni}) B^{-1} \ell x_i}{\sigma(t_{ni})} = \frac{\ell^\top B^{-1} X_n^\top x}{\sqrt{n \ell^\top B^{-1} \ell}}. \end{aligned}$$

Die „Covarianz“ B des Grenzmodells, das nach Satz 2.16 der Standard-Gaußshift auf dem Hilbertraum H_B ist (man beachte $J = 1$), spiegelt sich in den Tests φ_n^* auf Stufe n wider. Dies ist bei den Tests φ_n^{LQ} nicht der Fall. Hier steht an Stelle von B^{-1} der Ausdruck $n(X_n^\top X_n)^{-1} = B_n^{-1}$, der gegen B^{-1} konvergiert.

3.1.3 Asymptotik des Likelihood-Quotienten-Tests bei unbekanntem Fehler

Wir wollen nun die Testfolge $(\varphi_n^{LQ})_{n \in \mathbb{N}}$ aus (3.3) für den Fall untersuchen, dass *kein normalverteilter Fehler* vorliegt, d.h. für Modelle der Form

$$Z_n^{loc} = \frac{1}{\sqrt{n}} X_n \beta + \epsilon_n, \quad \beta \in \mathbb{R}^k, \quad (3.4)$$

mit X_n aus (3.2) und $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nn})^\top$, wobei die ϵ_{ni} stochastisch unabhängig sind und dieselbe Verteilung Q besitzen. Zwar sind diese Tests dann keine Likelihood-Quotienten-Tests mehr, da die zugehörigen Teststatistiken nicht mehr auf dem Dichtequotienten beruhen, wir wollen sie aber trotzdem als LQ-Tests bezeichnen, um ihren Ursprung anzudeuten. Diese Testfolge hat gegenüber der Folge asymptotisch optimaler Tests $(\varphi_n^*)_{n \in \mathbb{N}}$ den entscheidenden Vorteil, dass sie verteilungsfrei ist, dass sie also ohne Kenntnis der wahren Fehlerverteilung Q angewandt werden kann. Dafür muss ein Güteverlust in Kauf genommen werden, den wir im nächsten Abschnitt 3.1.4 untersuchen.

Für die folgenden Überlegungen benötigen wir im Vergleich zu Satz 2.16 teils speziellere, teils zusätzliche Voraussetzungen.

Voraussetzungen 3.3

1. *Es existiere ein Grenzdesign H , gegen das die Designfolge H_n im Sinne von (2.12), d.h. gleichmäßig konvergiert.*
2. *Die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien linear unabhängige Funktionen aus $L^2(H)$.*
3. *Die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien von beschränkter Variation.*
4. *Die Fehlerverteilung Q sei zentriert und besitze die Varianz τ^2 , also*

$$\int_{\mathbf{R}} x Q(dx) = 0 \quad \text{und} \quad \int_{\mathbf{R}} x^2 Q(dx) = \tau^2.$$

Die Asymptotik der Teststatistik der LQ-Testfolge bei unbekannter Fehlerverteilung ist bereits ausführlich untersucht worden, vgl. z.B. Stapleton (1995), S. 197 ff.. Da die bekannten Grenzwertsätze jedoch unter abstrakten Voraussetzungen an die Modellmatrizen X_n formuliert sind, wollen wir für unser speziell vorliegendes Regressionsmodell

die Konvergenz der Teststatistik selbst nachweisen. Ähnliche Konvergenzsätze wurden bereits in Bischoff und Miller (2000b) und Bischoff und Miller (2000a) hergeleitet. Mit den dortigen Hilfsmitteln können wir leicht die folgende Aussage für die Teststatistik der LQ-Testfolge beweisen.

Satz 3.4 *Unter den Voraussetzungen 3.3 gilt*

$$S_n := \frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top \epsilon_n}{\sqrt{\tau^2 \ell^\top (X_n^\top X_n)^{-1} \ell}} \xrightarrow{\mathcal{D}} N(0, 1),$$

d.h. die Statistiken S_n konvergieren schwach gegen eine $N(0, 1)$ -verteilte Zufallsvariable.

Beweis. Der Beweis beruht auf der Einbettung der S_n in den Raum $C[0, 1]$ mittels des Partialsummenprozesses. Definieren wir die linearen Einbettungsoperatoren $T_n : \mathbb{R}^n \rightarrow C[0, 1]$ wie in (2.14) und $V_n : C[0, 1] \rightarrow \mathbb{R}^n$ durch

$$V_n(u) := \left(u \left(\frac{1}{n} \right) - u(0), \dots, u(1) - u \left(\frac{n-1}{n} \right) \right)^\top, \quad u \in C[0, 1],$$

so ist $V_n \circ T_n = id_{\mathbb{R}^n}$ für alle $n \in \mathbb{N}$, und wir können S_n in der Form

$$S_n = \frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top \epsilon_n}{\sqrt{\tau^2 \ell^\top (X_n^\top X_n)^{-1} \ell}} = \frac{n \ell^\top (X_n^\top X_n)^{-1} X_n^\top V_n \left(T_n \left(\frac{1}{\sqrt{n\tau^2}} \epsilon_n \right) \right)}{\sqrt{n \ell^\top (X_n^\top X_n)^{-1} \ell}} \quad (3.5)$$

schreiben. In Bischoff und Miller (2000b) wird die Konvergenzaussage

$$\lim_{n \rightarrow \infty} X_n^\top V_n(u_n) = \int_0^1 \frac{f \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} u(dx) \in \mathbb{R}^k, \quad \text{für alle } u_n \rightarrow u \text{ in } C[0, 1] \quad (3.6)$$

bewiesen. Dabei ist obiges Integral komponentenweise zu verstehen. Weiter gilt nach dem Satz von Donsker, vgl. Billingsley (1968), Theorem 10.1,

$$T_n \left(\frac{\epsilon_n}{\sqrt{n\tau^2}} \right) \xrightarrow{\mathcal{D}} W,$$

wobei W eine Version der Brownschen Bewegung mit stetigen Pfaden bezeichnet. Zusammen mit $n (X_n^\top X_n)^{-1} \rightarrow B^{-1}$, vgl. Beweis von Satz 2.16, und Theorem 5.5 aus Billingsley (1968) folgt aus obiger Darstellung (3.5) von S_n die schwache Konvergenz

$$S_n \xrightarrow{\mathcal{D}} \frac{\ell^\top B^{-1} \int_0^1 \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} dW}{\sqrt{\ell^\top B^{-1} \ell}}, \quad (3.7)$$

wobei das Integral $\int_0^1 \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} dW$ pfadweise als Riemann-Stieltjes Integral zu verstehen ist. Jetzt ist nur noch nachzuweisen, dass der Grenzwert aus (3.7) eine $N(0, 1)$ -verteilte Zufallsvariable ist. Da der pfadweise Integralbegriff fast sicher mit dem stochastischen Integralbegriff aus Anhang A.5 übereinstimmt, ist $\int_0^1 \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} dW$ nach Konstruktion, vgl. Anhang A.5 ein L^2 -Grenzwert normalverteilter Zufallsvariablen und somit selbst wieder eine normalverteilte Zufallsvariable, vgl. Bauer (1991), S. 268, Aufgabe 8. Nach Lemma A.23 ist

$$\mathbb{E} \left(\int_0^1 \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} dW \right) = 0,$$

und nach Lemma A.23, Beispiel A.22 und Satz A.17 berechnet sich die Kovarianz zu

$$\begin{aligned} \text{Cov} \left(\int_0^1 \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} dW \right) &= \left(\int_{[0,1]} \frac{(f_i \circ H^{-1}) \cdot (f_j \circ H^{-1})}{(\sigma \circ H^{-1})^2} d\lambda \right)_{1 \leq i, j \leq k} \\ &= \left(\int_{[a,b]} \frac{f_i f_j}{\sigma^2} dH \right)_{1 \leq i, j \leq k} = B. \end{aligned}$$

Damit ergibt sich $\frac{\ell^\top B^{-1} \int_0^1 \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} dW}{\sqrt{\ell^\top B^{-1} \ell}} \sim N \left(0, \frac{\ell^\top B^{-1} B B^{-1} \ell}{\ell^\top B^{-1} \ell} \right) = N(0, 1)$. \square

Jetzt können wir die asymptotische Güte der LQ-Testfolge für das Testproblem (H_ℓ, K_ℓ) mit festem $\ell \in \mathbb{R}^k \setminus \{0\}$ berechnen. Unter Berücksichtigung der Varianz τ^2 definieren wir die LQ-Tests $\varphi_n^{LQ} : \mathbb{R}^n \rightarrow [0, 1]$ durch

$$x \mapsto \varphi_n^{LQ}(x) := \begin{cases} 1, & \text{falls } \frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top x}{\sqrt{\tau^2 \ell^\top (X_n^\top X_n)^{-1} \ell}} > \Phi^{-1}(1 - \alpha) \\ 0, & \text{falls } \frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top x}{\sqrt{\tau^2 \ell^\top (X_n^\top X_n)^{-1} \ell}} < \Phi^{-1}(1 - \alpha). \end{cases} \quad (3.8)$$

Satz 3.5 *Unter den Voraussetzungen 3.3 gilt für die Güten $g_{\varphi_n^{LQ}}$ der LQ-Tests aus (3.8), angewandt auf die lokalisierten Zufallsvariablen Z_n^{loc} aus (3.4)*

$$g^{LQ}(\beta) := \lim_{n \rightarrow \infty} g_{\varphi_n^{LQ}}(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\ell^\top \beta}{\sqrt{\tau^2 \ell^\top B^{-1} \ell}} \right) \quad \text{für alle } \beta \in \mathbb{R}^k.$$

Beweis. Wegen $n (X_n^\top X_n)^{-1} \rightarrow B^{-1}$ und $S_n \xrightarrow{\mathcal{D}} N(0, 1)$, vgl. Satz 3.4, folgt für Z_n^{loc} aus (3.4) und festes $\beta \in \mathbb{R}^k$

$$\frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top Z_n^{loc}}{\sqrt{\tau^2 \ell^\top (X_n^\top X_n)^{-1} \ell}} = \frac{\ell^\top \beta}{\sqrt{n \tau^2 \ell^\top (X_n^\top X_n)^{-1} \ell}} + S_n \xrightarrow{\mathcal{D}} N \left(\frac{\ell^\top \beta}{\sqrt{\tau^2 \ell^\top B^{-1} \ell}}, 1 \right).$$

Damit ergibt sich die Grenzgüte

$$\begin{aligned} \lim_{n \rightarrow \infty} g_{\varphi_n^{LQ}}(\beta) &= \lim_{n \rightarrow \infty} P \left(\frac{\ell^\top (X_n^\top X_n)^{-1} X_n^\top Z_n^{loc}}{\sqrt{\tau^2 \ell^\top (X_n^\top X_n)^{-1} \ell}} > \Phi^{-1}(1 - \alpha) \right) \\ &= \Phi \left(\Phi^{-1}(\alpha) + \frac{\ell^\top \beta}{\sqrt{\tau^2 \ell^\top B^{-1} \ell}} \right) \end{aligned}$$

für $\beta \in \mathbb{R}^k$. □

Korollar 3.6 *Ist die Varianz τ^2 der Fehlerverteilung unbekannt, so bleiben die Ergebnisse der Sätze 3.4 und 3.5 gültig, wenn man τ^2 durch eine konsistente Schätzfolge τ_n^2 ersetzt.*

3.1.4 Vergleich von φ_n^* mit dem Likelihood-Quotienten-Test bei unbekannter Fehlerverteilung

Unter den Voraussetzungen 3.3 konnten wir das lokale asymptotische Verhalten der Testfolge $(\varphi_n^{LQ})_{n \in \mathbb{N}}$ aus (3.3) angeben. Wir wollen nun diese verteilungsfreie Testfolge mit der optimalen Testfolge $(\varphi_n^*)_{n \in \mathbb{N}}$ aus Satz 3.1 anhand der zugehörigen Grenzgüten g^* und g^{LQ} vergleichen. Es stellt sich die Frage, auf welche Weise dies sinnvoll geschehen kann. Ein klassisches Konzept ist das der Asymptotisch Relativen Effizienz, auch Pitman-Effizienz genannt.

Definition 3.7 (Asymptotisch Relative Effizienz (Pitman-Effizienz))

Seien $E_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,\theta} \mid \theta \in \Theta\}) \in \mathcal{E}(\Theta)$, $n \in \mathbb{N}$, Experimente mit gemeinsamem Parameterraum Θ und (H, K) ein Testproblem, also eine Partition von Θ . Weiter seien $\varphi_{in} : \mathcal{X}_n \rightarrow [0, 1]$, $n \in \mathbb{N}$, $i = 1, 2$, zwei asymptotisch unverfälschte Folgen von Tests zum Niveau $\alpha \in (0, 1)$ mit zugehörigen Güten $g_{\varphi_{in}}$, für die für ein $\theta \in K$ Werte $\psi_i(\theta) \in (0, \infty)$ existieren sollen mit

$$\lim_{n \rightarrow \infty} g_{\varphi_{in}}(\theta) = \Phi \left(\Phi^{-1}(\alpha) + \psi_i(\theta) \right), \quad i = 1, 2. \quad (3.9)$$

Dann heißt der Quotient

$$ARE(\varphi_{2n} : \varphi_{1n} \mid \theta) := \frac{\psi_2^2(\theta)}{\psi_1^2(\theta)}$$

die **Asymptotisch Relative Effizienz** oder **Pitman-Effizienz** der Testfolge $(\varphi_{2n})_{n \in \mathbb{N}}$ bezüglich der Testfolge $(\varphi_{1n})_{n \in \mathbb{N}}$ unter der Alternative θ .

Existieren die Grenzwerte in (3.9) für alle $\theta \in K$ mit Konstanten $\psi_i(\theta) \in (0, \infty)$, und ist der Quotient $\frac{\psi_2^2(\theta)}{\psi_1^2(\theta)}$ unabhängig von θ , so heißt

$$ARE(\varphi_{2n} : \varphi_{1n}) := \frac{\psi_2^2(\theta_0)}{\psi_1^2(\theta_0)} \quad (\theta_0 \in K \text{ beliebig})$$

die **Asymptotisch Relative Effizienz** oder **Pitman-Effizienz** der Testfolge $(\varphi_{2n})_{n \in \mathbb{N}}$ bezüglich der Testfolge $(\varphi_{1n})_{n \in \mathbb{N}}$.

Bemerkung 3.8 Unter den Voraussetzungen von Definition 3.7 gilt

$$\text{ARE}(\varphi_{2n} : \varphi_{1n}) \begin{cases} < 1 \\ = 1 \\ > 1 \end{cases} \iff \lim_{n \rightarrow \infty} g_{\varphi_{2n}}(\theta) \begin{cases} < \\ = \\ > \end{cases} \lim_{n \rightarrow \infty} g_{\varphi_{1n}}(\theta), \quad \theta \in K.$$

Dementsprechend nennen wir die Testfolge

$$(\varphi_{2n})_{n \in \mathbb{N}} \text{ asymptotisch } \begin{cases} \text{schlechter als} \\ \text{äquivalent zu} \\ \text{besser als} \end{cases} (\varphi_{1n})_{n \in \mathbb{N}} \text{ (auf der Alternative } K).$$

Jetzt haben wir ein Maß dafür, um wieviel die asymptotisch optimale Testfolge besser ist als die Folge der LQ-Tests. Die folgende Aussage ist eine direkte Folgerung aus Satz 3.1 und Satz 3.5.

Satz 3.9 *Es seien die Voraussetzungen von Satz 3.1 und Satz 3.5 erfüllt, insbesondere genüge die Fehlerverteilung den Voraussetzungen 2.9 und 3.3. Für die Folge $(\varphi_n^*)_{n \in \mathbb{N}}$ der asymptotisch optimalen Tests aus Satz 3.1 und für die Folge der LQ-Tests $(\varphi_n^{LQ})_{n \in \mathbb{N}}$ aus Formel (3.8) gilt*

$$\text{ARE}(\varphi_n^* : \varphi_n^{LQ}) = J\tau^2.$$

Bemerkung 3.10

1. Die Satz 3.9 zu Grunde liegenden Experimente sind mit Rate $\frac{1}{\sqrt{n}}$ um den Ursprung $\beta = 0 \in \overline{H}_\ell \cap \overline{K}_\ell$ lokalisiert. Wir können die Asymptotisch Relative Effizienz somit wie folgt interpretieren (man vergleiche dazu Witting und Müller-Funk (1995), S. 282 ff., speziell Satz 6.101): Wird der asymptotisch optimale Test φ_n^* bei einem Stichprobenumfang $n \in \mathbb{N}$ angewandt, so würde man bei Anwendung des LQ-Tests in etwa $\text{ARE}(\varphi_n^* : \varphi_n^{LQ}) = J\tau^2$ -mal so viele Beobachtungen benötigen, also $J\tau^2 \cdot n$, um dieselbe Güte zu erreichen. Eine solche asymptotische Aussage gilt natürlich nur für hinreichend große Beobachtungsumfänge n .
2. Aus Satz 3.9 erhalten wir auch eine rein analytische Aussage: Für jede Wahrscheinlichkeitsverteilung Q mit Fisher-Information J und Varianz τ^2 , die den Voraussetzungen 2.9 und 3.3 genügt, gilt

$$J\tau^2 \geq 1.$$

Wir wollen das Ergebnis an einem Beispiel verdeutlichen.

Beispiel 3.11 Sei Q wie in Beispiel 2.10 4. die zweiseitige Exponentialverteilung mit Parameter $r := 1$, auch Laplace-Verteilung genannt. Dann ist Q zentriert mit Varianz $\tau^2 = 2$ und Fisher-Information $J = 1$. Weiter wollen wir ein einfaches Change-Point Problem mit bekannter Stelle eines möglichen Changes betrachten. Seien dazu $f_1 := 1_{[0,1]}$ und $f_2 := 1_{[\frac{1}{2},1]}$ vorgegebene Regressoren auf dem Versuchsbereich $\mathcal{V} = [0, 1]$. Als Grenzdesign \bar{H} wählen wir das Lebesgue-Maß λ auf \mathcal{V} und als Varianzfunktion $\sigma \equiv 1$. Damit erhalten wir

$$B = \begin{pmatrix} \int_{[0,1]} f_1^2 d\lambda & \int_{[0,1]} f_1 f_2 d\lambda \\ \int_{[0,1]} f_1 f_2 d\lambda & \int_{[0,1]} f_2^2 d\lambda \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad \text{bzw.} \quad B^{-1} = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}.$$

Wir wollen testen, ob im Punkt $\frac{1}{2}$ ein Change-Point mit positivem Change vorliegt, d.h. ob die wahre Regressionsfunktion an der Stelle $\frac{1}{2}$ einen Sprung nach oben macht. Ist $\beta = (\beta_1, \beta_2)^\top \in \mathbb{R}^2$ der unbekannte Parameter, so testen wir also die Hypothese $\beta_2 \leq 0$ gegen die Alternative $\beta_2 > 0$. Mit dem Vektor $\ell = (0, 1)^\top$ haben wir somit ein Testproblem der Form (H_ℓ, K_ℓ) , vgl. (3.1). Als Grenzgüten des optimalen Tests sowie des LQ-Tests ergeben sich damit bei einem Niveau $\alpha \in (0, 1)$

$$g^*(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\beta_2}{2} \right) \quad \text{bzw.} \quad g^{LQ}(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\beta_2}{2\sqrt{2}} \right)$$

für $\beta = (\beta_1, \beta_2)^\top \in \mathbb{R}^2$. Abbildung 3.1 zeigt die beiden Gütekurven in Abhängigkeit von β_2 bei einem Niveau von $\alpha = 0.1$. Eine sinnvolle Interpretation dieser Grenzgüten

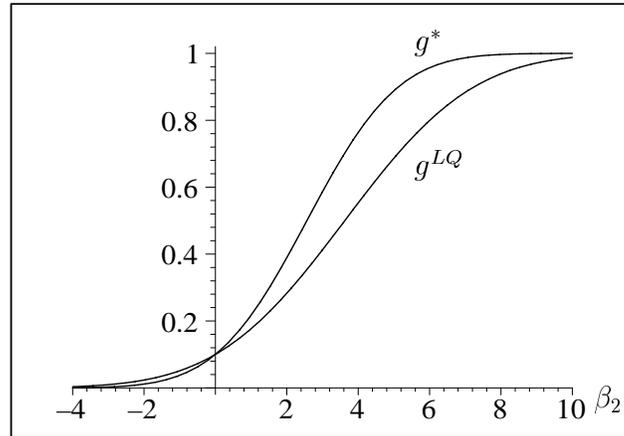


Abbildung 3.1: Grenzgüten

erhält man aber erst durch Satz 3.9. Dieser liefert die Asymptotisch Relative Effizienz

$$\text{ARE}(\varphi_n^* : \varphi_n^{LQ}) = J\tau^2 = 2$$

des asymptotisch optimalen Tests gegenüber dem LQ-Test. Nach Bemerkung 3.10 benötigt der LQ-Test also in etwa doppelt so viele Beobachtungen wie der asymptotisch optimale Test, um dieselbe Güte zu erreichen.

3.2 Tests für das allgemeine Regressionsmodell

Während wir in Abschnitt 3.1 Testprobleme für das Lineare Regressionsmodell untersucht haben, werden wir jetzt in Analogie zum Vorgehen in 3.1.1 optimale Testfolgen zu „einfachen“ Testproblemen für das allgemeine Regressionsmodell herleiten. Grundlage dieses Abschnitts sind Satz 2.22 sowie Korollar 2.23. Die dortigen Voraussetzungen seien erfüllt, insbesondere identifizieren wir die Regressoren $f \in \mathcal{F} \subseteq \{f \mid \frac{f}{\sigma} \in L^2(H)\}$ wieder mit den zugehörigen Elementen

$$h_f(\cdot) = \sqrt{J} \int_{[0, \cdot]} \frac{f \circ H^{-1}}{\sigma \circ H^{-1}} d\lambda$$

des kernreproduzierenden Hilbertraumes $H(R)$ der Brownschen Bewegung, vgl. (2.31). Im Gegensatz zum Linearen Regressionsmodell haben wir hier als Grenzexperiment keinen „vollständigen“ Standard-Gaußshift auf $H(R)$ vorliegen, sondern nur eine Einschränkung $G_{H(R), \mathcal{F}}$, vgl. (2.37).

In diesem Abschnitt sei die *Varianzfunktion* σ^2 *beschränkt*. Dies ist eine sehr schwache Voraussetzung, wenn man bedenkt, dass der Versuchsbereich $\mathcal{V} = [a, b]$ ein kompaktes Intervall ist. Sie hat aber zur Folge, dass aus $\frac{f}{\sigma} \in L^2(H)$ sofort $f \in L^2(H)$ folgt. Dieser Hilbertraum $L^2(H)$ ist zwar nicht der „passende“ Hilbertraum für unsere Theorie, aber er ist der natürliche Hilbertraum, in dem man Testprobleme für f formulieren wird.

Wir wollen wieder ein „eindimensionales“, lineares Testproblem untersuchen, geben uns dazu eine Funktion $\ell \in L^2(H) \setminus \{0\}$ vor und betrachten das Testproblem

$$H_\ell = \{f \in \mathcal{F} \mid \langle f, \ell \rangle_{L^2(H)} \leq 0\}, \quad K_\ell = \{f \in \mathcal{F} \mid \langle f, \ell \rangle_{L^2(H)} > 0\},$$

das wir mit Hilfe des Innenproduktes $\langle \cdot, \cdot \rangle_{H(R)}$ umformulieren. Sei dazu

$$\gamma := \frac{\sigma^2 \ell}{J}.$$

Da σ nach Voraussetzung beschränkt ist, folgt $\frac{\gamma}{\sigma} \in L^2(H)$, und wir können γ mit dem zugehörigen Element

$$h_\gamma(\cdot) = \sqrt{J} \int_{[0, \cdot]} \frac{\gamma \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} \lambda(dx) = \int_{[0, \cdot]} \frac{(\sigma \circ H^{-1}(x)) \cdot (\ell \circ H^{-1}(x))}{\sqrt{J}} \lambda(dx)$$

aus dem Hilbertraum $H(R)$ identifizieren. Dann folgt mit Beispiel A.25 und dem Transformationssatz A.17

$$\begin{aligned} \|h_\gamma\|_{H(R)}^2 &= J \int_{[0, 1]} \frac{(\gamma \circ H^{-1}(x))^2}{(\sigma \circ H^{-1}(x))^2} \lambda(dx) \\ &= \frac{1}{J} \int_{[0, 1]} (\sigma \circ H^{-1}(x))^2 (\ell \circ H^{-1}(x))^2 \lambda(dx) \stackrel{\text{A.17}}{=} \frac{1}{J} \|\sigma \ell\|_{L^2(H)}^2, \end{aligned}$$

und für $f \in \mathcal{F}$ mit zugehörigem $h_f \in H(R)$ ist

$$\begin{aligned} \langle h_\gamma, h_f \rangle_{H(R)} &= J \int_{[0,1]} \frac{(\gamma \circ H^{-1}(x)) \cdot (f \circ H^{-1}(x))}{(\sigma \circ H^{-1}(x))^2} \lambda(dx) \\ &= \int_{[0,1]} (\ell \circ H^{-1}(x)) \cdot (f \circ H^{-1}(x)) \lambda(dx) \stackrel{A.17}{=} \langle f, \ell \rangle_{L^2(H)}. \end{aligned}$$

Damit lässt sich das Testproblem (H_ℓ, K_ℓ) in der Form

$$H_\ell = \{f \in \mathcal{F} \mid \langle h_\gamma, h_f \rangle_{H(R)} \leq 0\}, \quad K_\ell = \{f \in \mathcal{F} \mid \langle h_\gamma, h_f \rangle_{H(R)} > 0\}$$

schreiben.

Im Folgenden wollen wir Satz 1.52 anwenden und die dortigen Voraussetzungen, speziell die schwache Konvergenz (1.21) nachprüfen. Dazu benötigen wir zusätzliche Anforderungen an das „Testproblem“ ℓ .

Voraussetzungen 3.12 Für $\ell \in L^2(H)$ gelte

1.

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{\sigma(t_{ni}) |\ell(t_{ni})|}{\sqrt{n}} = 0, \quad (3.10)$$

2.

$$\int_{[a,b]} \sigma^2 \ell^2 dH_n \xrightarrow{n \rightarrow \infty} \int_{[a,b]} \sigma^2 \ell^2 dH, \quad (3.11)$$

3.

$$\int_{[a,b]} \ell f dH_n \xrightarrow{n \rightarrow \infty} \int_{[a,b]} \ell f dH \quad \text{für alle } f \in \mathcal{F}. \quad (3.12)$$

Diese Voraussetzungen sind z.B. dann erfüllt, wenn H_n gleichmäßig gegen H konvergiert und die Funktionen σ , ℓ und $f \in \mathcal{F}$ alle von beschränkter Variation sind, vgl. Beispiel 2.6. Im Wesentlichen muss also die „Testfunktion“ ℓ dieselben Eigenschaften wie die Regressoren besitzen.

Satz 3.13 Es mögen die Voraussetzungen 3.12 an ℓ sowie die Voraussetzungen von Satz 2.22 erfüllt sein, und die Varianzfunktion σ^2 sei beschränkt. Für festes $\ell \in L^2(H) \setminus \{0\}$ betrachten wir das Testproblem

$$H_\ell = \{f \in \mathcal{F} \mid \langle f, \ell \rangle_{L^2(H)} \leq 0\}, \quad K_\ell = \{f \in \mathcal{F} \mid \langle f, \ell \rangle_{L^2(H)} > 0\}.$$

Dann ist die Testfolge $(\varphi_n^*)_{n \in \mathbb{N}}$ mit $\varphi_n^* : \mathbb{R}^n \rightarrow [0, 1]$,

$$\varphi_n^*(x) = \begin{cases} 1, & \frac{1}{\sqrt{nJ\|\sigma\ell\|_{L^2(H)}^2}} \sum_{i=1}^n \sigma(t_{ni})\ell(t_{ni}) \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i) > \Phi^{-1}(1-\alpha) \\ 0, & \frac{1}{\sqrt{nJ\|\sigma\ell\|_{L^2(H)}^2}} \sum_{i=1}^n \sigma(t_{ni})\ell(t_{ni}) \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i) < \Phi^{-1}(1-\alpha), \end{cases}$$

$x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, asymptotisch optimal in der Klasse aller asymptotisch α -ähnlichen Testfolgen im Sinne von Definition 1.41 bzw. Bemerkung 1.42 mit einem fest vorgegebenen Niveau $\alpha \in (0, 1)$. Man beachte, dass wir die Regressorenklasse $\mathcal{F} \subset L^2(H)$ so ergänzen können, dass $\overline{H_\ell} \cap \overline{K_\ell} \neq \emptyset$ und damit der Begriff „asymptotisch α -ähnlich“ nicht trivial ist. Für die Folge der Gütefunktionen $(g_{\varphi_n^*})_{n \in \mathbb{N}}$ gilt

$$g^*(f) := \lim_{n \rightarrow \infty} g_{\varphi_n^*}(f) = \Phi \left(\Phi^{-1}(\alpha) + \sqrt{J} \frac{\langle f, \ell \rangle_{L^2(H)}}{\|\sigma\ell\|_{L^2(H)}} \right) \quad \text{für alle } f \in \mathcal{F}.$$

Beweis. Der Beweis beruht auf Satz 1.52, dessen Voraussetzungen hier überprüft werden müssen. Das Analogon

$$X_{n,f} = \sum_{i=1}^n \frac{f(t_{ni})}{\sqrt{n}\sigma(t_{ni})} \cdot \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i)$$

zur zentralen Folge gemäß Formel (2.38) aus Korollar 2.23 erfüllt die zugehörigen Bedingungen (1.14), (1.15) und (1.16) von Satz 1.52. Man beachte die Identifikation von Elementen des Hilbertraums $H(R)$ mit ihren Ableitungen. Weiter gilt für die Teststatistik der Tests φ_n^*

$$\frac{1}{\sqrt{nJ\|\sigma\ell\|_{L^2(H)}^2}} \sum_{i=1}^n \sigma(t_{ni})\ell(t_{ni}) \left(-\frac{q'(x_i)}{q(x_i)} \right) \cdot 1_{\{q>0\}}(x_i) = \frac{X_{n,\gamma}(x)}{\|h_\gamma\|_{H(R)}}$$

für $x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$. Es ist nur noch Voraussetzung (1.21) nachzuweisen, und die Aussage folgt aus Satz 1.52. Setzen wir

$$\mathcal{F}_\gamma := \{ \alpha\gamma + \beta f \mid \alpha, \beta \in \mathbb{R}, f \in \mathcal{F} \},$$

so erfüllt dieser „erweiterte“ Parameterraum \mathcal{F}_γ ebenfalls die Voraussetzungen von Satz 2.22, d.h. $\frac{1}{\sigma}\mathcal{F}_\gamma$ ist eine H_n - H -Konvergenzklasse, und es gilt

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{|\tilde{f}(t_{ni})|}{\sqrt{n}\sigma(t_{ni})} = 0 \quad \text{für alle } \tilde{f} \in \mathcal{F}_\gamma,$$

was direkt aus den Voraussetzungen 3.12 folgt. Damit können wir mit Satz 2.22 auf die schwache Konvergenz der „erweiterten“ Experimente

$$E_{n,\mathcal{F}_\gamma}^{loc} \xrightarrow{w} G_{H(R),\mathcal{F}_\gamma}$$

schließen, und aus der schwachen Konvergenz (2.39) in Korollar 2.23 folgt Bedingung (1.21). \square

Beispiel 3.14 Gegeben seien das Grenzdesign $H = \lambda$ auf dem Versuchsbereich $\mathcal{V} = [0, 1]$ und eine H -äquidistante Designfolge H_n . Weiter seien $\mathcal{F} = BV[0, 1] \cap L^2(H)$, die Varianzfunktion $\sigma \equiv 1$, und als Fehlerverteilung Q legen wir wie schon in Beispiel 3.11 eine Laplace-Verteilung mit Parameter $r = 1$ und Fisher-Information $J = 1$ zu Grunde. Dann hat Q die Dichte $q(x) = \frac{1}{2}e^{-|x|}$ mit Lebesgue-Ableitung $q'(x) = -\text{sgn}(x)q(x)$. Wir interessieren uns dafür, ob das Lebesgue-Integral über den unbekanntes Regressor positiv oder negativ ist. Mittels der Funktion $\ell \equiv 1$ formulieren wir das zugehörige Testproblem

$$H_\ell = \{f \in \mathcal{F} \mid \langle f, \ell \rangle_{L^2(H)} \leq 0\}, \quad K_\ell = \{f \in \mathcal{F} \mid \langle f, \ell \rangle_{L^2(H)} > 0\},$$

also

$$H_\ell = \left\{ f \in \mathcal{F} \mid \int_{[0,1]} f \, d\lambda \leq 0 \right\}, \quad K_\ell = \left\{ f \in \mathcal{F} \mid \int_{[0,1]} f \, d\lambda > 0 \right\},$$

und nach Satz 3.13 erhalten wir zum Niveau α eine asymptotisch optimale Testfolge $(\varphi_n^*)_{n \in \mathbb{N}}$ durch $\varphi_n^* : \mathbb{R}^n \rightarrow [0, 1]$,

$$\varphi_n^*(x) = \begin{cases} 1, & \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sgn}(x_i) > \Phi^{-1}(1 - \alpha) \\ 0, & \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{sgn}(x_i) < \Phi^{-1}(1 - \alpha), \end{cases}$$

$x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$, mit der Grenzgüte

$$g^*(f) := \Phi \left(\Phi^{-1}(\alpha) + \int_{[0,1]} f \, d\lambda \right) \quad \text{für alle } f \in \mathcal{F}.$$

Kapitel 4

Anwendungen in der Designtheorie

In diesem Kapitel beschäftigen wir uns mit dem Gebiet der Versuchsplanung. In Abschnitt 4.1 schildern wir die zu Grunde liegende Problemstellung und gehen auf klassische Ansätze aus der Theorie Linearer Modelle ein, wo man ausgehend vom linearen erwartungstreuen Schätzen bei quadratischem Schaden und einem fest vorgegebenen Stichprobenumfang versucht, einen möglichst „guten“ Versuchsplan zu bestimmen. Im Gegensatz dazu, aber ganz im Kontext der vorliegenden Arbeit, liegt es nahe, die Designtheorie im Rahmen der Entscheidungstheorie zu sehen, d.h. Versuchspläne zu bestimmen, die „universell gut“ sind für alle möglichen Entscheidungsprobleme. Dieser neue Ansatz führt uns zu denselben Methoden, wie sie in der klassischen Designtheorie angewandt werden, er erlaubt aber - im Gegensatz zu den bisherigen Ansätzen - eine wesentlich weitreichendere Interpretation der Ergebnisse.

In den Abschnitten 4.2 und 4.3 beschäftigen wir uns mit asymptotischer Designtheorie. Es werden Versuchspläne bestimmt, die nicht für einen festen Beobachtungsumfang optimal sind, sondern asymptotisch optimal, d.h. „gut“ für alle hinreichend großen Beobachtungsumfänge. Dieser asymptotische Ansatz, der bisher in der Literatur noch sehr wenig untersucht worden ist, wird in der vorliegenden Arbeit erstmals im entscheidungstheoretischen Rahmen betrachtet, wobei die Konvergenzaussagen für Regressionsmodelle aus Kapitel 2 die entscheidende Grundlage sind. Dabei basiert Abschnitt 4.2 auf dem globalen Konvergenzsatz 2.31 für Lineare Regressionsmodelle mit normalverteilten Fehlern, Abschnitt 4.3 auf den lokalen Konvergenzsätzen aus Abschnitt 2.4.

Zum Abschluss des Kapitels untersuchen wir in Abschnitt 4.4 ein klassisches Optimalitätskriterium der Designtheorie Linearer Modelle, nämlich die E -Optimalität aus entscheidungstheoretischer Sicht.

4.1 Designtheorie bei fest vorgegebenem Stichprobenumfang

4.1.1 Problemstellung

Bei der Untersuchung der Regressionsmodelle

$$E_n = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f(t_{ni})}{\sigma(t_{ni})}} \mid f \in \mathcal{F} \right\} \right) \quad (4.1)$$

aus Formel (2.6) war das Dreiecksschema (t_{ni}) , $i = 1, \dots, n$, $n \in \mathbb{N}$, der Versuchspunkte, das wir auch als **Designfolge** bezeichnet haben, stets fest vorgegeben. Für $n \in \mathbb{N}$ haben wir die Versuchspunkte t_{n1}, \dots, t_{nn} mit der zugehörigen Verteilungsfunktion H_n aus (2.10) identifiziert. Eine solche aus n (nicht notwendig verschiedenen) Punkten hervorgehende Verteilungsfunktion H_n nennen wir künftig ein **konkretes (oder exaktes) Design für n Versuchspunkte**. Um die Abhängigkeit der Regressionsmodelle von dem zugehörigen konkreten Design auszudrücken, schreiben wir ab jetzt $E_n(H_n)$.

Beispiel 4.1 Im Fall eines Linearen Regressionsmodells mit den Regressoren f_1, \dots, f_k , dem Beobachtungsraum $\mathcal{X} = \mathbb{R}^n$ und dem Parameterraum \mathbb{R}^k schreiben wir für die Modellmatrix künftig

$$A_n^\top(H_n) = \begin{pmatrix} f_1(t_{n1}) & \cdots & f_k(t_{n1}) \\ \vdots & \ddots & \vdots \\ f_1(t_{nn}) & \cdots & f_k(t_{nn}) \end{pmatrix}$$

und für die Kovarianzmatrix

$$\Sigma_n^2(H_n) = \begin{pmatrix} \sigma^2(t_{n1}) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(t_{nn}) \end{pmatrix}.$$

Unter Normalverteilungsannahme ergeben sich dann die Linearen Modelle

$$\mathcal{N}(A_n(H_n), \Sigma_n^2(H_n)) \quad \text{bzw. transformiert} \quad \mathcal{N}(A_n(H_n)\Sigma_n^{-1}(H_n), I_n),$$

vgl. Beispiel 2.1. Ist die Fehlerverteilung unbekannt mit Erwartungswert 0 und Varianz 1, so erhält man die Linearen Modelle, vgl. Beispiel 2.2,

$$\mathcal{L}(A_n(H_n), \Sigma_n^2(H_n)) \quad \text{bzw. transformiert} \quad \mathcal{L}(A_n(H_n)\Sigma_n^{-1}(H_n), I_n).$$

Es stellt sich jetzt die Frage, wie man die Versuchspunkte t_{n1}, \dots, t_{nn} , d.h. das konkrete Design H_n , zu wählen hat, wenn man ein möglichst informatives statistisches Experiment erhalten möchte. Auf Grund der bisherigen Überlegungen ist es naheliegend, als Vergleichsmaßstab zweier Experimente mit verschiedenen Designs die Δ -Metrik LeCam's zu wählen. Der Zugang in der klassischen Designtheorie ist jedoch ein anderer.

4.1.2 Approximative Designtheorie für das Lineare Regressionsmodell - klassische Ansätze

Wir wollen die Prinzipien der klassischen Designtheorie kurz erläutern, um sie mit unseren entscheidungstheoretischen Ansätzen vergleichen zu können. Eine ausführliche Darstellung der klassischen Theorie findet man zum Beispiel in den Büchern Silvey (1980) und Pukelsheim (1993). Wir beschränken uns hier auf den Fall der Linearen Regression.

Bei festem Stichprobenumfang $n_0 \in \mathbb{N}$ betrachten wir für zwei verschiedene konkrete Designs H_{1n_0} und H_{2n_0} die Linearen Regressionsmodelle $\mathcal{L}(A_{n_0}(H_{1n_0}), \Sigma_{n_0}^2(H_{1n_0}))$ bzw. $\mathcal{L}(A_{n_0}(H_{2n_0}), \Sigma_{n_0}^2(H_{2n_0}))$ aus Beispiel 4.1. Beiden Modellen liegen dieselben Regressoren f_1, \dots, f_k sowie dieselbe Varianzfunktion σ^2 zu Grunde. Jetzt vergleichen wir diese Experimente bezüglich des linearen erwartungstreuen Schätzens bei quadratischem Schaden, wie wir das bereits in Abschnitt 1.3 erläutert haben. Dabei erhalten wir in der klassischen Designtheorie das folgende Ergebnis als eine direkte Folgerung der Sätze 1.12 und 1.14 sowie Gleichung (2.45) auf Seite 64.

Satz 4.2 *Das zum konkreten Design H_{1n_0} gehörige Modell $\mathcal{L}(A_{n_0}(H_{1n_0}), \Sigma_{n_0}^2(H_{1n_0}))$ ist genau dann bezüglich des linearen erwartungstreuen Schätzens bei quadratischem Schaden mindestens so informativ wie das zum Design H_{2n_0} gehörige Lineare Modell $\mathcal{L}(A_{n_0}(H_{2n_0}), \Sigma_{n_0}^2(H_{2n_0}))$, d.h. in der Schreibweise aus (1.6)*

$$\mathcal{L}(A_{n_0}(H_{1n_0}), \Sigma_{n_0}^2(H_{1n_0})) \supseteq_q \mathcal{L}(A_{n_0}(H_{2n_0}), \Sigma_{n_0}^2(H_{2n_0})),$$

wenn im Sinne der Löwner-Ordnung

$$\begin{pmatrix} \int_{[a,b]} \frac{f_1^2}{\sigma^2} dH_{1n_0} & \cdots & \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_{1n_0} \\ \vdots & \ddots & \vdots \\ \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_{1n_0} & \cdots & \int_{[a,b]} \frac{f_k^2}{\sigma^2} dH_{1n_0} \end{pmatrix} \geq_L \begin{pmatrix} \int_{[a,b]} \frac{f_1^2}{\sigma^2} dH_{2n_0} & \cdots & \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_{2n_0} \\ \vdots & \ddots & \vdots \\ \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_{2n_0} & \cdots & \int_{[a,b]} \frac{f_k^2}{\sigma^2} dH_{2n_0} \end{pmatrix}$$

*gilt. In diesem Fall nennen wir das Design H_{1n_0} **mindestens so informativ** wie H_{2n_0} . Die obigen Matrizen werden **Informationsmatrizen** der jeweiligen Modelle genannt; sie waren bereits Gegenstand der Untersuchungen in Abschnitt 2.4.2, vgl. (2.26) und (2.28) auf Seite 53.*

Auf der Suche nach „guten“ Designs treten vor allem zwei Probleme auf. Zum einen existieren im Allgemeinen keine optimalen Designs, da die Löwner-Ordnung nur eine Halbordnung ist. Man kann sich dann entweder auf verschiedene Klassen von konkreten Designs einschränken, in denen es optimale Vertreter gibt, oder man optimiert ein Funktional der Informationsmatrizen, was z.B. zum Begriff der E -Optimalität führt, die in Abschnitt 4.4 untersucht wird. Zum anderen stellt sich die diskrete Struktur ($n_0 \in \mathbb{N}$ fest) als schwerwiegendes Problem heraus, da dadurch einfache analytische Hilfsmittel

bei den Optimierungsproblemen nicht anwendbar sind. Diese Schwierigkeit führt uns zur **approximativen Designtheorie**. Wir betrachten hierbei die Informationsmatrizen selbst, lassen jetzt aber alle Verteilungsfunktionen H auf $[a, b]$ als Integratoren zu. Man spricht dann von **approximativen** oder **stetigen Designs**, wobei der Begriff „stetig“ nichts mit den Eigenschaften der Verteilungsfunktion H zu tun hat, diese kann durchaus Sprünge besitzen. Die Informationsmatrizen sind also von der Form

$$B(H) := \begin{pmatrix} \int_{[a,b]} \frac{f_1^2}{\sigma^2} dH & \cdots & \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH \\ \vdots & \ddots & \vdots \\ \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH & \cdots & \int_{[a,b]} \frac{f_k^2}{\sigma^2} dH \end{pmatrix} \quad (4.2)$$

und man wird versuchen, sie selbst bzw. Funktionale davon zu „optimieren“, wobei auf übliche Methoden der Analysis zurückgegriffen werden kann. Ein optimales stetiges Design wird jedoch i.A. nicht durch ein konkretes Design mit n_0 Versuchspunkten realisierbar sein, d.h. es wird kein solches konkretes Design mit derselben Informationsmatrix geben. Zwar besagt der Satz von Carathéodory, vgl. Pukelsheim (1993), S. 188, dass jede von einem beliebigen stetigen Design erzeugte Informationsmatrix auch von einem Design mit höchstens $\frac{1}{2}k(k+1)+1$ Massepunkten erzeugt werden kann. Die Wahrscheinlichkeitsmasse in diesen Punkten ist aber i.A. kein Vielfaches von $\frac{1}{n_0}$. Somit bleibt die Frage, inwieweit man ein optimales Design für n_0 Versuchspunkte durch ein konkretes Design mit n_0 Versuchspunkten annähern kann. Diese Problemstellung korrespondiert, wie wir sehen werden, mit den asymptotischen Überlegungen von Abschnitt 4.2.

4.1.3 Entscheidungstheoretische Ansätze in der Designtheorie

In Abschnitt 4.1.1 haben wir bereits angedeutet, dass es auf Grund der bisherigen entscheidungstheoretischen Überlegungen nahe liegt, Designs im Sinne von LeCam's Δ -Metrik zu vergleichen. Die allgemeinen Regressionsmodelle E_n aus (4.1) entsprechen dem Beobachten der Zufallsvariablen

$$Y_{ni} = f(t_{ni}) + \sigma(t_{ni})\epsilon_{ni}, \quad i = 1, \dots, n,$$

wie wir das in Abschnitt 2.1 beschrieben haben.

Definition 4.3 *Es seien der Versuchsbereich \mathcal{V} , die Fehlerverteilung Q , die Regressorenklasse \mathcal{F} sowie die Varianzfunktion σ^2 fest vorgegeben, und weiter seien $n_0 \in \mathbb{N}$ ein fester Stichprobenumfang, H_{1n_0}, H_{2n_0} zwei konkrete Designs mit n_0 Versuchspunkten und $E_{n_0}(H_{1n_0}), E_{n_0}(H_{2n_0})$ die zugehörigen Regressionsmodelle gemäß Formel (4.1). Dann nennen wir das konkrete Design H_{1n_0} **mindestens so informativ** wie H_{2n_0} , falls im Sinne der LeCam-Halbordnung, vgl. Definition 1.4, gilt*

$$E_{n_0}(H_{1n_0}) \supseteq E_{n_0}(H_{2n_0}).$$

Sind die beiden Regressionsmodelle äquivalent, also $E_{n_0}(H_{1n_0}) \sim E_{n_0}(H_{2n_0})$, so nennen wir auch die zugehörigen Designs **äquivalent**. Mit

$$\delta_D(H_{1n_0}, H_{2n_0}) := \delta(E_{n_0}(H_{1n_0}), E_{n_0}(H_{2n_0}))$$

bezeichnen wir den **(einseitigen) Informationsabstand** des Designs H_{1n_0} vom Design H_{2n_0} und mit

$$\Delta_D(H_{1n_0}, H_{2n_0}) := \Delta(E_{n_0}(H_{1n_0}), E_{n_0}(H_{2n_0}))$$

den **Informationsabstand** der beiden Designs H_{1n_0} und H_{2n_0} , vgl. Definition 1.5.

Interpretation 4.4 (von Definition 4.3) Gemäß Definition 1.4 ist ein Design H_{1n_0} genau dann mindestens so informativ wie ein zweites Design H_{2n_0} , falls es für jedes Entscheidungsproblem mit von unten halbstetiger und nach unten beschränkter Verlustfunktion (Verlustfunktionen sind definitionsgemäß nach unten beschränkt, man beachte aber, dass auf eine obere Schranke für die Verlustfunktion verzichtet werden kann, vgl. Definition 1.2) für das zugehörige Regressionsmodell $E_{n_0}(H_{1n_0})$ gleichmäßig nicht schlechtere Entscheidungen gibt wie für $E_{n_0}(H_{2n_0})$. Wir werden dies an zwei Beispielen erläutern: Ist ein beliebiges Testproblem, d.h. eine Partition (H, K) von \mathcal{F} gegeben, so existiert zu jedem Test ψ für $E_{n_0}(H_{2n_0})$ ein Test φ für $E_{n_0}(H_{1n_0})$ mit gleichmäßig nicht schlechterer Güte, d.h.

$$g_\varphi(f) \leq g_\psi(f) \quad \text{für alle } f \in H \quad \text{und} \quad g_\varphi(f) \geq g_\psi(f) \quad \text{für alle } f \in K.$$

Soll dagegen die unbekannte Regressionsfunktion $f \in \mathcal{F}$ geschätzt werden, und ist eine beliebige von unten halbstetige Verlustfunktion $(W_f)_{f \in \mathcal{F}}$ gegeben, so existiert zu jedem Schätzer T_2 für $E_{n_0}(H_{2n_0})$ ein Schätzer T_1 für $E_{n_0}(H_{1n_0})$ mit gleichmäßig nicht größerem Risiko, d.h.

$$r_{T_1}(f) \leq r_{T_2}(f) \quad \text{für alle } f \in \mathcal{F}.$$

Etwas Vorsicht ist bei dieser Interpretation jedoch geboten, da LeCam's Δ -Metrik mittels „verallgemeinerter Entscheidungsfunktionen“ eingeführt wurde. Streng genommen müsste man z.B. sagen, dass zu jedem Schätzer T_2 eine verallgemeinerte Entscheidungsfunktion β_1 mit gleichmäßig nicht größerem Risiko existiert. Man beachte dazu aber Satz 1.1 oder die Tatsache, dass die Menge der randomisierten Entscheidungsfunktionen in der Menge der verallgemeinerten Entscheidungsfunktionen bezüglich der Topologie der punktweisen Konvergenz, vgl. Abschnitt 1.8.2, dicht liegt, vgl. Strasser (1985), S. 232.

Anhand dieser Interpretation erkennt man den entscheidenden Unterschied zur klassischen Designtheorie für Lineare Modelle aus Abschnitt 4.1.2, wo man „gute“ Designs für das erwartungstreue Schätzen bei quadratischem Schaden sucht. Bei unserer Definition ist ein Design „gut“, wenn es für „alle“ Problemstellungen „gut“ ist. Nach Lemma 1.6 ist Δ_D eine Pseudometrik auf dem Raum aller konkreten Designs mit n_0 Versuchspunkten bei gegebenen $Q, \mathcal{V}, \mathcal{F}$ und σ^2 . Bildet man Äquivalenzklassen, so erhält man einen metrischen Raum, dessen topologische Eigenschaften hier jedoch nicht untersucht werden.

Zum Abschluss dieses Abschnitts werden wir noch zwei wichtige Beispiele anführen, die den Zusammenhang zur klassischen Designtheorie herstellen.

Beispiel 4.5 (Konstantenregression) Betrachten wir das Problem der Konstantenregression bei konstanter Varianzfunktion, d.h. $\mathcal{F} = \{c \cdot 1_{\mathcal{V}} \mid c \in \mathbb{R}\}$ und o.B.d.A. $\sigma^2 \equiv 1$, für beliebigen Versuchsbereich \mathcal{V} und beliebige Fehlerverteilung Q , so hängen die Regressionsmodelle

$$E_n(H) = (\mathbb{R}^n, \mathcal{B}^n, \{Q_c^n \mid c \in \mathbb{R}\})$$

nicht mehr von den Versuchspunkten ab. Alle Designs H sind somit äquivalent.

Beispiel 4.6 (Lineare Modelle mit normalverteilten Fehlern)

Betrachten wir wie in Beispiel 4.1 Lineare Modelle mit normalverteilten Fehlern $\mathcal{N}(A_{n_0}(H_{i n_0})\Sigma_{n_0}^{-1}(H_{i n_0}), I_{n_0})$, $i = 1, 2$, so ist nach Satz 1.12 das konkrete Design $H_{1 n_0}$ genau dann mindestens so informativ im Sinne von Definition 4.3 wie $H_{2 n_0}$, wenn gilt

$$A_{n_0}(H_{1 n_0})\Sigma_{n_0}^{-2}(H_{1 n_0})A_{n_0}(H_{1 n_0})^\top \geq_L A_{n_0}(H_{2 n_0})\Sigma_{n_0}^{-2}(H_{2 n_0})A_{n_0}(H_{2 n_0})^\top,$$

was auf Grund von Gleichung (2.45) auf Seite 64 äquivalent zu

$$\begin{pmatrix} \int_{[a,b]} \frac{f_1^2}{\sigma^2} dH_{1 n_0} & \cdots & \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_{1 n_0} \\ \vdots & \ddots & \vdots \\ \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_{1 n_0} & \cdots & \int_{[a,b]} \frac{f_k^2}{\sigma^2} dH_{1 n_0} \end{pmatrix} \geq_L \begin{pmatrix} \int_{[a,b]} \frac{f_1^2}{\sigma^2} dH_{2 n_0} & \cdots & \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_{2 n_0} \\ \vdots & \ddots & \vdots \\ \int_{[a,b]} \frac{f_1 \cdot f_k}{\sigma^2} dH_{2 n_0} & \cdots & \int_{[a,b]} \frac{f_k^2}{\sigma^2} dH_{2 n_0} \end{pmatrix}$$

ist. Die Begriffsbildung eines „besseren Designs“ im Sinne von Definition 4.3 stimmt folglich mit derjenigen der klassischen Designtheorie überein, vgl. Satz 4.2. Interpretation 4.4 erlaubt uns aber viel weitergehende Schlussfolgerungen, als das bisher möglich war.

4.2 Asymptotische Designtheorie für Lineare Regressionsmodelle mit normalverteilten Fehlern

In der asymptotischen Designtheorie gehen wir nicht von einem festen Beobachtungsumfang aus, sondern geben Folgen von Designs an, die für große Beobachtungsumfänge „nahezu optimal“ sind, also asymptotisch optimal. Die Grundidee ist dabei die folgende:

- (i) In Kapitel 2 haben wir den Regressionsmodellen „geeignete“ Grenzmodelle zugeordnet, in denen Grenzdesigns auftraten. Man bestimmt für diese Grenzmodelle ein „gutes“ stetiges Design H . Dabei vergleicht man zwei Grenzexperimente mit verschiedenen Designs analog zu Definition 4.3 mittels der LeCam-Halbordnung.

- (ii) Man bestimmt eine Designfolge H_n , die das gewählte Grenzdesign H gemäß den Voraussetzungen des den Experimentfolgen zu Grunde liegenden Konvergenzsatzes approximiert.

Dieses Vorgehen wird uns zu denselben Ansätzen führen, die wir für die approximative Designtheorie in Abschnitt 4.1.2 beschrieben haben, ermöglicht aber auf Grund der neuen Herangehensweise wesentlich weitreichendere Interpretationen.

Die soeben skizzierte Methodik wird anhand von speziellen Situationen präzisiert, und wir werden ausführlich erläutern, was unter asymptotischer Optimalität einer Designfolge zu verstehen ist. Hierzu werden wir die in Kapitel 2 erarbeiteten Konvergenzsätze heranziehen.

Da globale asymptotische Untersuchungen zunächst eine größere Rolle spielen als lokale Konvergenzsätze, werden wir zuerst die Auswirkungen von Abschnitt 2.5 und speziell von Satz 2.31, wo (starke) asymptotische Äquivalenz von Experimentfolgen vorliegt, für die asymptotische Designtheorie analysieren. Dazu geben wir uns einen festen Versuchsbereich $\mathcal{V} = [a, b]$ sowie Regressoren f_1, \dots, f_k und eine Varianzfunktion $\sigma^2 > 0$ vor. Für ein konkretes Design H_n betrachten wir $A_n(H_n)$, $\Sigma_n^2(H_n)$ bzw. $\mathcal{N}(A_n(H_n), \Sigma_n^2(H_n))$ wie in Beispiel 4.1. Für ein stetiges Design H definieren wir die Informationsmatrix $B(H)$ gemäß Formel (4.2). Man beachte, dass die Matrix $B(H)$ genau dann regulär ist, wenn die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ linear unabhängig in $L^2(H)$ sind.

Definition 4.7 *Gegeben seien zwei Folgen $\mathcal{N}(A_n(H_{in}), \Sigma_n^2(H_{in}))$, $i = 1, 2$, Linearer Regressionsmodelle mit normalverteilten Fehlern wie in Beispiel 4.1. Für die Regressoren, die Varianzfunktion und die beiden Designfolgen gelte wie in Satz 2.31*

1. *es existieren Grenzdesigns H_1, H_2 , gegen das die Designfolgen H_{1n}, H_{2n} im Sinne von (2.11) oder (2.12) konvergieren,*
2. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ bilden eine $H_{1n}-H_1$ - und $H_{2n}-H_2$ -Konvergenzklasse,*
3. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien linear unabhängig in $L^2(H_1)$ und in $L^2(H_2)$.*

*Dann heißt die Folge konkreter Designs $(H_{1n})_{n \in \mathbb{N}}$ **asymptotisch mindestens so informativ** wie die Folge $(H_{2n})_{n \in \mathbb{N}}$, falls für die zugehörigen Grenzdesigns*

$$B(H_1) \geq_L B(H_2).$$

*gilt. Existiert keine asymptotisch informativere Designfolge als $(H_{1n})_{n \in \mathbb{N}}$, so nennen wir die Folge $(H_{1n})_{n \in \mathbb{N}}$ **asymptotisch zulässig**.*

Der folgende Satz illustriert die Bedeutung dieser Definition.

Satz 4.8 *Gegeben sei eine Designfolge $(H_{1n})_{n \in \mathbb{N}}$, die im Sinne von Definition 4.7 echt informativer sei als eine Folge $(H_{2n})_{n \in \mathbb{N}}$, d.h. es gelte $B(H_1) >_L B(H_2)$. Dann gilt für die zugehörigen Linearen Regressionsmodelle mit normalverteilten Fehlern*

$$\mathcal{N}(A_n(H_{1n}), \Sigma_n^2(H_{1n})) \supseteq \mathcal{N}(A_n(H_{2n}), \Sigma_n^2(H_{2n})) \quad (4.3)$$

ab hinreichend großem $n \in \mathbb{N}$, d.h. die zu der Designfolge $(H_{1n})_{n \in \mathbb{N}}$ gehörigen Regressionsmodelle $\mathcal{N}(A_n(H_{1n}), \Sigma_n^2(H_{1n}))$ sind ab einem $n_0 \in \mathbb{N}$ informativer als die zur Folge $(H_{2n})_{n \in \mathbb{N}}$ gehörigen Modelle $\mathcal{N}(A_n(H_{2n}), \Sigma_n^2(H_{2n}))$. Der Δ -Abstand der beiden Modelle konvergiert:

$$\Delta(\mathcal{N}(A_n(H_{1n}), \Sigma_n^2(H_{1n})), \mathcal{N}(A_n(H_{2n}), \Sigma_n^2(H_{2n}))) \xrightarrow{n \rightarrow \infty} \|N(0, B(H_1)) - N(0, B(H_2))\|_V \quad (4.4)$$

mit der Abschätzung

$$\begin{aligned} 1 - 2^{\frac{k}{2}} \cdot \frac{(\det B(H_1))^{\frac{1}{4}} (\det B(H_2))^{\frac{1}{4}}}{(\det(B(H_1) + B(H_2)))^{\frac{1}{2}}} &\leq \\ &\leq \|N(0, B(H_1)) - N(0, B(H_2))\|_V \leq \sqrt{1 - 2^k \cdot \frac{(\det B(H_1))^{\frac{1}{2}} (\det B(H_2))^{\frac{1}{2}}}{\det(B(H_1) + B(H_2))}} \end{aligned} \quad (4.5)$$

für den Grenzwert. Für die zugehörigen Linearen Modelle bei unbekannter Fehlerverteilung gilt

$$\mathcal{L}(A_n(H_{1n}), \Sigma_n^2(H_{1n})) \supseteq_q \mathcal{L}(A_n(H_{2n}), \Sigma_n^2(H_{2n})) \quad (4.6)$$

ab hinreichend großem $n \in \mathbb{N}$.

Beweis. In Satz 2.31 wurde die asymptotische Äquivalenz

$$\Delta(\mathcal{N}(A_n(H_{in}), \Sigma_n^2(H_{in})), \mathcal{N}(\sqrt{n}B^{\frac{1}{2}}(H_i), I_k)) \xrightarrow{n \rightarrow \infty} 0$$

für $i = 1, 2$ bewiesen. Für die Grenzmodelle gilt nach Satz 1.13

$$\begin{aligned} \mathcal{N}(\sqrt{n}B^{\frac{1}{2}}(H_1), I_k) &\supseteq \mathcal{N}(\sqrt{n}B^{\frac{1}{2}}(H_2), I_k) \\ &\iff \mathcal{N}(B^{\frac{1}{2}}(H_1), I_k) \supseteq \mathcal{N}(B^{\frac{1}{2}}(H_2), I_k) \\ &\iff B(H_1) \geq_L B(H_2), \end{aligned}$$

und damit folgt (4.3). Weiter folgt aus Satz 1.13 und Lemma A.7

$$\begin{aligned} \Delta(\mathcal{N}(\sqrt{n}B^{\frac{1}{2}}(H_1), I_k), \mathcal{N}(\sqrt{n}B^{\frac{1}{2}}(H_2), I_k)) &= \Delta(\mathcal{N}(B^{\frac{1}{2}}(H_1), I_k), \mathcal{N}(B^{\frac{1}{2}}(H_2), I_k)) \\ &= \|N(0, B(H_1)) - N(0, B(H_2))\|_V \end{aligned}$$

und damit die Konvergenz (4.4). Die Abschätzung (4.5) folgt aus Beispiel A.13 und Ungleichung (A.3); (4.6) folgt aus (4.3) in Verbindung mit Satz 1.14. \square

Bemerkung 4.9

1. Wir wollen nochmals auf die Interpretation von Definition 4.7 eingehen. Ist eine Designfolge asymptotisch informativer als eine andere, so heißt das, dass die Folge zugehöriger Linearer Regressionsmodelle unter Normalverteilungsannahme ab einem Stichprobenumfang n_0 immer im „starken“ Sinne LeCam’s informativer ist als die entsprechende andere Modellfolge. Der Informationsabstand der Modelle konvergiert gegen den Informationsabstand der Grenzmodelle aus Formel (4.4). Die Designfolge ist somit für jedes Entscheidungsproblem mit von unten halbsteiger Verlustfunktion ab hinreichend großem Stichprobenumfang geeigneter als die andere, vgl. Interpretation 4.4. Betrachtet man Lineare Regressionsmodelle bei unbekannter Fehlerverteilung mit Erwartungswert 0 und Varianz 1, so ist eine informativere Designfolge zwar nicht mehr im „starken“ Sinne LeCam’s besser, aber sie ist ab hinreichend großem Stichprobenumfang besser für das erwartungstreue lineare Schätzen bei quadratischem Schaden.
2. Auch wenn keines von zwei Grenzdesigns im Sinne von Definition 4.7 informativer ist als das andere, kann man den Abstand der zugehörigen Grenzmodelle angeben. Er ist unabhängig vom Stichprobenumfang n und lässt sich als Variationsabstand zweier Normalverteilungen gemäß Theorem 8.5.7 aus Torgersen (1991) berechnen. Damit kann man, vergleichbar zu Definition 4.3, einen asymptotischen Abstand zweier Designfolgen definieren.
3. Schließlich soll kurz erläutert werden, wie Definition 4.7 von einem Anwender zu verstehen ist, der sich im Vorfeld eines Experimentes ein „gutes“ Design auswählen möchte. Gemäß Satz 4.8 kann er wie folgt vorgehen:
 - (i) Er wählt für die Grenzmodelle ein geeignetes stetiges Design H aus, z.B. ein Design, das in einer vorgegebenen Klasse optimal ist.
 - (ii) Er bestimmt eine Designfolge H_n , die das gewählte Grenzdesign H gemäß den Voraussetzungen von Definition 4.7 approximiert. So kann man zum Beispiel eine H -äquidistante Designfolge wählen, vgl. Beispiel 2.3.
 - (iii) Er beachtet dabei, dass die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ linear unabhängig in $L^2(H)$ sind und eine H_n - H -Konvergenzklasse bilden.

Bei der Bestimmung eines geeigneten konkreten Designs für fest vorgegebenen (großen) Stichprobenumfang geht man also vom Grenzmodell aus. Das folgende Beispiel soll dies verdeutlichen.

Beispiel 4.10 (Geradenregression) Seien $\mathcal{V} = [0, 1]$, $\sigma^2 \equiv 1$, $f_1 \equiv 1$, $f_2 = id_{[0,1]}$. Sei $\mathcal{H} := \{H : \mathcal{V} \rightarrow [0, 1] \mid H \text{ Verteilungsfunktion}, H \not\equiv 1, H \neq 1_{\{1\}}\}$. Dann sind für alle $H \in \mathcal{H}$ die Funktionen f_1 und f_2 linear unabhängig in $L^2(H)$. Beide Funktionen sind stetig und von beschränkter Variation, so dass sie eine H_n - H -Konvergenzklasse bilden für jede Designfolge $(H_n)_{n \in \mathbb{N}}$, die im Sinne von (2.11) oder (2.12) gegen ein Grenzdesign

H konvergiert. Die Voraussetzungen von Satz 2.31 sind somit diesbezüglich erfüllt. Weiter geben wir uns für $\gamma \in (0, 1)$ die Klasse

$$\mathcal{H}_\gamma := \left\{ H \in \mathcal{H} \mid \int_{[0,1]} x H(dx) = \gamma \right\}$$

möglicher Grenzdesigns vor. Dann ist das Design $H_\gamma := (1 - \gamma)1_{[0,1]} + \gamma 1_{\{1\}}$ optimal in der Klasse \mathcal{H}_γ , d.h. es gilt $B(H_\gamma) \geq_L B(H)$ für alle $H \in \mathcal{H}_\gamma$. Jedes der Designs H_γ ist zulässig in \mathcal{H} , d.h. es gibt kein informativeres Design in \mathcal{H} für das Grenzmodell. Die Menge

$$\{H_\gamma \mid \gamma \in [0, 1]\}$$

ist die Menge aller zulässigen Designs. Betrachten wir die Klasse $\mathcal{H}_{\frac{1}{2}}$, so ist das optimale Design $H_{\frac{1}{2}}$ besser als das zur Dichte $g(x) := 12(x - \frac{1}{2})^2$, $x \in [0, 1]$, gehörige Design H_g mit $H_g(x) = 4x^3 - 6x^2 + 3x$, $x \in [0, 1]$. Das Design H_g ist wiederum informativer als das zum Lebesgue-Maß gehörige Design $H_\lambda = id_{[0,1]}$. Für den Informationsabstand im asymptotischen Modell ergibt sich z.B. für die Designs $H_{\frac{1}{2}}$ und H_λ die Abschätzung

$$0.069 \approx 1 - \frac{\sqrt[4]{3}}{\sqrt{2}} \leq \Delta \left(\mathcal{N}(B^{\frac{1}{2}}(H_{\frac{1}{2}}), I_2), \mathcal{N}(B^{\frac{1}{2}}(H_\lambda), I_2) \right) \leq \sqrt{\frac{2 - \sqrt{3}}{2}} \approx 0.366.$$

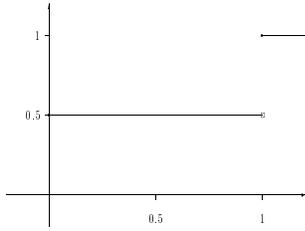
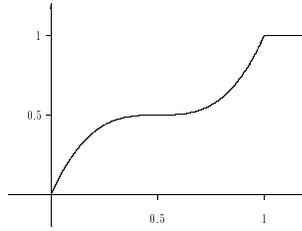
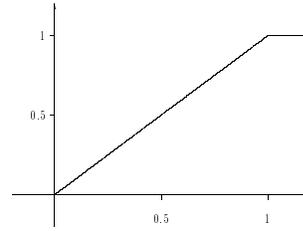
Bild 1: $H_{\frac{1}{2}}$ Bild 2: H_g Bild 3: H_λ

Abbildung 4.1: Grenzdesigns

Beweis. Für $\gamma \in (0, 1)$ und $H \in \mathcal{H}_\gamma$ ist

$$B(H) = \begin{pmatrix} 1 & \gamma \\ \gamma & \int_{[0,1]} x^2 H(dx) \end{pmatrix} \quad \text{und speziell} \quad B(H_\gamma) = \begin{pmatrix} 1 & \gamma \\ \gamma & \gamma \end{pmatrix}.$$

Wegen $\int_{[0,1]} x^2 H(dx) \leq \int_{[0,1]} x H(dx) = \gamma$ ist $B(H_\gamma) - B(H)$ positiv semidefinit, also $B(H_\gamma) \geq_L B(H)$. Damit ist das Design H_γ optimal in der Klasse \mathcal{H}_γ . Für $\gamma_1 \neq \gamma_2$ ist

$$B(H_{\gamma_1}) - B(H_{\gamma_2}) = \begin{pmatrix} 0 & \gamma_1 - \gamma_2 \\ \gamma_1 - \gamma_2 & \gamma_1 - \gamma_2 \end{pmatrix}$$

stets indefinit. Damit sind alle Designs H_γ zulässig in der Klasse \mathcal{H} .

Wegen $\int_{[0,1]} x^2 g(x) dx = \frac{2}{5} > \frac{1}{3} = \int_{[0,1]} x^2 dx$ ist H_g informativer als H_λ . Die Abschätzung für $H_{\frac{1}{2}}$ und H_λ erhält man durch einfache Rechnung aus Ungleichung (4.5). \square

4.3 Lokale asymptotische Designtheorie für Regressionsmodelle

Während wir in Abschnitt 4.2 globale asymptotische Äquivalenz von Regressionsmodellen dazu genutzt haben, um globale asymptotische Aussagen zur Designtheorie herzuleiten, werden wir nun die lokalen Untersuchungen aus Abschnitt 2.4 verwenden. Es sei hier ausdrücklich darauf hingewiesen, dass bei den folgenden Überlegungen jeweils eine *Lokalisierung im Parameterraum* zu Grunde liegt. Diese lokalisierten Modelle sind speziell für statistische Fragestellungen, die sich „auf den Lokalisierungspunkt konzentrieren“, geeignet. Dementsprechend sind die „guten“ Designs dieses Abschnitts als „gut in Hinblick auf solche lokale Fragestellungen“ zu verstehen.

4.3.1 Lokalisierte Lineare Regressionsmodelle

Grundlage der folgenden Betrachtungen sind die Sätze 2.16 und 2.26. Beide Sätze untersuchen die Asymptotik lokalisierter Linearer Regressionsmodelle mit einer bekannten Fehlerverteilung, die den Voraussetzungen 2.9 genügt. In einem Fall liegt schwache, im anderen Fall starke Konvergenz im Sinne LeCam's vor. Dabei ist der folgende Aspekt besonders wichtig: Für die Asymptotik der Modelle müssen wir die Fehlerverteilung als bekannt voraussetzen, für die asymptotischen Untersuchungen zur Designtheorie ist sie jedoch irrelevant, solange sie nur den Voraussetzungen 2.9 genügt. Die Aussagen zur Designtheorie sind also *unabhängig von der Fehlerverteilung*.

Zunächst greifen wir nochmals die Ergebnisse der beiden Sätze auf. Man beachte dabei die jeweiligen Voraussetzungen und die Tatsache, dass wir uns auf den Ursprung als Lokalisierungspunkt beschränken. Wir werden im Unterschied zu Kapitel 2 alle auftretenden Experimente in Abhängigkeit von der Designfolge $(H_n)_{n \in \mathbb{N}}$ und vom Grenzdesign H schreiben. In Satz 2.16 hatten wir die Modelle

$$E_n^{loc}(H_n) = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q \left. \frac{1}{\sqrt{n}} \cdot \frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})} \right| \beta \in \mathbb{R}^k \right\} \right)$$

aus (2.24) betrachtet und die schwache Konvergenz $E_n^{loc}(H_n) \xrightarrow{w} \mathcal{N}(\sqrt{J}B^{\frac{1}{2}}(H), I_k)$ nachgewiesen. In Satz 2.26 hatten wir die auf einen beschränkten Parameterraum $\Theta \subseteq \mathbb{R}^k$ eingeschränkten Modelle $E_{n,\Theta}^{loc}(H_n)$ untersucht und die starke Konvergenz gegen den auf Θ eingeschränkten Gaußshift

$$G_\Theta(H) := \left(\mathbb{R}^k, \mathcal{B}^k, \left\{ N(\sqrt{J}B^{\frac{1}{2}}(H)\beta, I_k) \mid \beta \in \Theta \right\} \right)$$

nachgewiesen, also $\Delta(E_{n,\Theta}^{loc}(H_n), G_\Theta(H)) \xrightarrow{n \rightarrow \infty} 0$. Die Fehlerverteilung Q tritt in den Grenzexperimenten lediglich in Form ihrer Fisher-Information J auf.

4.3.1.1 Starke lokale asymptotische Designtheorie

Wir untersuchen zunächst die designtheoretische Bedeutung des starken lokalen Konvergenzsatzes 2.26. In Analogie zu Definition 4.7 ergibt sich die folgende auf die vorliegende Situation angepasste Definition.

Definition 4.11 *Gegeben sei die Situation von Satz 2.26, d.h. es gelten die Voraussetzungen 2.9 an die Fehlerverteilung und die folgenden Voraussetzungen an die Regressoren, die Varianzfunktion und die Designfolgen:*

1. *Es existieren Grenzdesigns H_1, H_2 , gegen das die Designfolgen H_{1n}, H_{2n} im Sinne von (2.11) oder (2.12) konvergieren,*
2. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien beschränkt, und sie bilden eine $H_{1n}-H_1$ - und eine $H_{2n}-H_2$ -Konvergenzklasse,*
3. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien linear unabhängig in $L^2(H_1)$ und in $L^2(H_2)$.*

*Weiter seien $\Theta \subset \mathbb{R}^k$ beschränkt und die Modelle $E_{n,\Theta}^{loc}(H_{in})$, $i = 1, 2$, gegeben. Dann nennen wir eine Folge konkreter Designs $(H_{1n})_{n \in \mathbb{N}}$ **lokal asymptotisch mindestens so informativ** wie eine Folge $(H_{2n})_{n \in \mathbb{N}}$, falls für die zugehörigen Grenzmodelle gilt*

$$G_{\Theta}(H_1) \supseteq G_{\Theta}(H_2).$$

*Existiert keine informativere Designfolge als $(H_{1n})_{n \in \mathbb{N}}$, so nennen wir die Folge $(H_{1n})_{n \in \mathbb{N}}$ **lokal asymptotisch zulässig**.*

Analog zu Satz 4.8 beweist man den folgenden Satz, der Definition 4.11 verdeutlicht.

Satz 4.12 *Gegeben sei eine Designfolge $(H_{1n})_{n \in \mathbb{N}}$, die im Sinne von Definition 4.11 informativer als eine Folge $(H_{2n})_{n \in \mathbb{N}}$ sei, d.h. es gelte $G_{\Theta}(H_1) \supseteq G_{\Theta}(H_2)$. Dann gilt für die zugehörigen Regressionsmodelle*

$$E_{n,\Theta}^{loc}(H_{1n}) \supseteq E_{n,\Theta}^{loc}(H_{2n})$$

ab hinreichend großem $n \in \mathbb{N}$, d.h. die zu der Designfolge $(H_{1n})_{n \in \mathbb{N}}$ gehörigen Regressionsmodelle $E_{n,\Theta}^{loc}(H_{1n})$ sind ab einem $n_0 \in \mathbb{N}$ informativer als die zur Folge $(H_{2n})_{n \in \mathbb{N}}$ gehörigen Modelle $E_{n,\Theta}^{loc}(H_{2n})$. Desweiteren ist

$$G_{\Theta}(H_1) \supseteq G_{\Theta}(H_2) \iff \beta^{\top} B(H_1) \beta \geq \beta^{\top} B(H_2) \beta \text{ für alle } \beta \in \Theta,$$

vgl. Torgersen (1991), S. 429 ff.. Für den Δ -Abstand der beiden Modelle gilt

$$\Delta(E_{n,\Theta}^{loc}(H_{1n}), E_{n,\Theta}^{loc}(H_{2n})) \xrightarrow{n \rightarrow \infty} \Delta(G_{\Theta}(H_1), G_{\Theta}(H_2)) \leq \|N(0, B(H_1)) - N(0, B(H_2))\|_V.$$

Wir verdeutlichen die Bedeutung von Definition 4.11 und Satz 4.12 in Abbildung 4.2, für die die zu den Regressionsmodellen $E_{n,\Theta}^{loc}(H_{in})$ gehörigen Grenzmodelle $G_{\Theta}(H_i)$, $i = 1, 2$, den Δ -Abstand $\Delta_0 > 0$ haben sollen.

$$\begin{array}{ccc}
 B(H_1) & \geq_L & B(H_2) \\
 & \Downarrow & \\
 G_{\Theta}(H_1) & \underset{\substack{\supseteq \\ \Delta_0 > 0 \\ \text{unabhängig von } J}}{=} & G_{\Theta}(H_2) \\
 \uparrow \Delta & & \uparrow \Delta \\
 E_{n,\Theta}^{loc}(H_{1n}) & \underset{\substack{\supseteq \\ \text{für } n \text{ hinreichend groß} \\ \Delta \rightarrow \Delta_0 \text{ für } n \rightarrow \infty}}{=} & E_{n,\Theta}^{loc}(H_{2n})
 \end{array}$$

Abbildung 4.2: Bedeutung von Definition 4.11 und Satz 4.12

Auch Bemerkung 4.9 behält ihre Gültigkeit, berücksichtigt man die Tatsache, dass allen Überlegungen *lokalisierte Modelle* zu Grunde liegen und die Ergebnisse dahingehend zu interpretieren sind.

4.3.1.2 Schwache lokale asymptotische Designtheorie

Deutlich schwieriger ist es, die Bedeutung der schwachen Konvergenz im Sinne LeCam's aus Satz 2.16 für die asymptotische Designtheorie zu erfassen. Abbildung 4.3 soll diese Problematik analog zu Abbildung 4.2 verdeutlichen.

$$\begin{array}{ccc}
 B(H_1) & >_L & B(H_2) \\
 & \Downarrow & \\
 \mathcal{N}(\sqrt{J}B^{\frac{1}{2}}(H_1), I_k) & \underset{\substack{\supseteq \\ \Delta_0 > 0 \\ \text{unabhängig von } J}}{=} & \mathcal{N}(\sqrt{J}B^{\frac{1}{2}}(H_2), I_k) \\
 \uparrow w & & \uparrow w \\
 E_n^{loc}(H_{1n}) & ? & E_n^{loc}(H_{2n})
 \end{array}$$

Abbildung 4.3: Problematik der schwachen Konvergenz aus Satz 2.16

Auf Grund der nur schwachen Konvergenz können wir nicht mehr erwarten, dass die

lokalisierten Experimente $E_n^{loc}(H_{1n})$ ab hinreichend großem Beobachtungsumfang n im Sinne LeCam's informativer sind als $E_n^{loc}(H_{2n})$. Die folgende Definition bedarf also einer anderen Interpretation.

Definition 4.13 *Gegeben seien die Voraussetzungen von Satz 2.16, d.h. es gelten die Voraussetzungen 2.9 an die Fehlerverteilung und die folgenden Voraussetzungen an die Regressoren, die Varianzfunktion und die Designfolgen:*

1. *Es existieren Grenzdesigns H_1, H_2 , gegen das die Designfolgen H_{1n}, H_{2n} im Sinne von (2.11) oder (2.12) konvergieren,*
2. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ bilden sowohl eine $H_{1n}-H_1$ - als auch eine $H_{2n}-H_2$ -Konvergenzklasse, und es ist*

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{f_j(t_{ni}^{(l)})}{\sqrt{n}\sigma(t_{ni}^{(l)})} = 0 \quad \text{für } j = 1, \dots, k \text{ und } l = 1, 2,$$

wobei $t_{ni}^{(l)}$, $i = 1, \dots, n$, die zum Design H_{1n} gehörigen Versuchspunkte bezeichnen,

3. *die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ seien linear unabhängig in $L^2(H_1)$ und in $L^2(H_2)$.*

*Weiter seien die lokalisierten Experimente $E_n^{loc}(H_{in})$, $i = 1, 2$, gegeben. Dann nennen wir eine Folge konkreter Designs $(H_{1n})_{n \in \mathbb{N}}$ **schwach lokal asymptotisch mindestens so informativ** wie eine Folge $(H_{2n})_{n \in \mathbb{N}}$, falls für die zugehörigen Grenzdesigns gilt*

$$B(H_1) \geq_L B(H_2).$$

Die Bedeutung von Definition 4.13 lässt sich nicht mehr so anschaulich erklären, wie wir das im Fall der starken Konvergenz konnten. Die Folgerungen, die wir aus der schwachen Konvergenz im Sinne LeCam's ziehen können, beruhen im Wesentlichen auf der Hájek-LeCam Minimax-Schranke 1.46, die verschiedene Interpretationen von Definition 4.13 ermöglicht. Zunächst werden wir Definition 4.13 aber an einem Beispiel illustrieren.

Beispiel 4.14 In Satz 3.1 wurde aus der schwachen Konvergenz unter den Voraussetzungen von Satz 2.16 mit Hilfe der Hájek-LeCam Minimax-Schranke 1.46 eine asymptotisch optimale Testfolge für Testprobleme der Form

$$H_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta \leq 0\}, \quad K_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta > 0\}$$

hergeleitet. Dabei ergab sich die vom Grenzdesign H abhängige Grenzgüte

$$g_H^*(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \sqrt{J} \frac{\ell^\top \beta}{\sqrt{\ell^\top B^{-1}(H) \ell}} \right),$$

wobei $B(H)$ die Informationsmatrix bezeichnet.

Sind nun zwei Designfolgen $(H_{1n})_{n \in \mathbb{N}}$, $(H_{2n})_{n \in \mathbb{N}}$ mit Grenzdesigns H_1, H_2 gegeben und

ist die Folge $(H_{1n})_{n \in \mathbb{N}}$ im Sinne von Definition 4.13 schwach lokal asymptotisch mindestens so informativ wie die Folge $(H_{2n})_{n \in \mathbb{N}}$, d.h. es gilt $B(H_1) \geq_L B(H_2)$, so besitzt die zu den Designs H_{1n} gehörige optimale Testfolge $\varphi_n^*(H_{1n})$ bezüglich der zu H_{2n} gehörigen optimalen Testfolge $\varphi_n^*(H_{2n})$ die Asymptotisch Relative Effizienz, vgl. Definition 3.7,

$$\text{ARE}(\varphi_n^*(H_{1n}) : \varphi_n^*(H_{2n})) = \frac{\ell^\top B^{-1}(H_2)\ell}{\ell^\top B^{-1}(H_1)\ell} \geq 1.$$

Um ein Zahlenbeispiel zu geben, greifen wir das Beispiel 4.10 der Geradenregression nochmals auf, diesmal im Kontext von Satz 2.16. In der Klasse $\mathcal{H}_{\frac{1}{2}}$ betrachten wir die Grenzdesigns $H_{\frac{1}{2}}$ und H_λ , die von den Folgen $(H_{1n})_{n \in \mathbb{N}}$ bzw. $(H_{2n})_{n \in \mathbb{N}}$ approximiert werden. Die Fragestellung, ob die Regressionsfunktion im Punkt $x = 1$ unterhalb der x -Achse liegt, entspricht der Wahl $\ell = (1, 1)^\top$. Dann berechnet sich die Pitman-Effizienz zu

$$\text{ARE}(\varphi_n^*(H_{1n}) : \varphi_n^*(H_{2n})) = 2.$$

Will man wissen, ob die Steigung der Geraden negativ ist, so wählt man $\ell = (0, 1)^\top$ und erhält als Pitman-Effizienz der zugehörigen optimalen Testfolgen

$$\text{ARE}(\varphi_n^*(H_{1n}) : \varphi_n^*(H_{2n})) = 3.$$

Jetzt interpretieren wir Definition 4.13 theoretisch anhand von Korollar 1.48, einem Spezialfall der Hájek-LeCam Minimax-Schranke, und erhalten eine asymptotische Minimax-Interpretation für den Begriff eines schwach lokal asymptotisch informativeren Designs. Wie in Interpretation 4.4 argumentieren wir an Stelle von verallgemeinerten Entscheidungsfunktionen mit den „üblichen“ randomisierten Entscheidungsfunktionen.

Interpretation 4.15 (Minimax-Interpretation von Definition 4.13)

Gegeben seien die Situation aus Definition 4.13 und ein Entscheidungspaar (D, W) mit von unten halbstetiger Verlustfunktion W . Die Designfolge $(H_{1n})_{n \in \mathbb{N}}$ sei schwach lokal asymptotisch mindestens so informativ wie $(H_{2n})_{n \in \mathbb{N}}$, d.h. für die zugehörigen Grenzexperimente gilt

$$F_1 := \mathcal{N}(\sqrt{J}B^{\frac{1}{2}}(H_1), I_k) \supseteq \mathcal{N}(\sqrt{J}B^{\frac{1}{2}}(H_2), I_k) =: F_2.$$

Folglich existiert nach Definition 1.4 zu jeder Entscheidungsfunktion $\rho_2 \in \mathcal{R}(F_2, D)$ eine Entscheidungsfunktion $\rho_1 \in \mathcal{R}(F_1, D)$ mit gleichmäßig nicht größerem Risiko, vgl. (1.1), also

$$r_{\rho_1}(\beta) \leq r_{\rho_2}(\beta), \quad \beta \in \mathbb{R}^k.$$

Schränken wir den Parameterraum \mathbb{R}^k auf eine beliebige Teilmenge $\Theta_0 \subseteq \mathbb{R}^k$ ein, so gilt für die Minimax-Risiken in den Grenzmodellen die Ungleichung

$$\inf_{\rho_1 \in \mathcal{R}(F_1, D)} \sup_{\beta \in \Theta_0} r_{\rho_1}(\beta) \leq \inf_{\rho_2 \in \mathcal{R}(F_2, D)} \sup_{\beta \in \Theta_0} r_{\rho_2}(\beta).$$

Setzen wir weiter die Existenz von Entscheidungsfunktionen $\rho_{1n}^* \in \mathcal{R}(E_n^{loc}(H_{1n}), D)$, $n \in \mathbb{N}$, mit

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in \Theta_0} r_{\rho_{1n}^*}(\beta) = \inf_{\rho_1 \in \mathcal{R}(F_1, D)} \sup_{\beta \in \Theta_0} r_{\rho_1}(\beta)$$

voraus, so folgt mit Korollar 1.48

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in \Theta_0} r_{\rho_{1n}^*}(\beta) = \inf_{\rho_1 \in \mathcal{R}(F_1, D)} \sup_{\beta \in \Theta_0} r_{\rho_1}(\beta) \leq \inf_{\rho_2 \in \mathcal{R}(F_2, D)} \sup_{\beta \in \Theta_0} r_{\rho_2}(\beta) \leq \liminf_{n \rightarrow \infty} \sup_{\beta \in \Theta_0} r_{\rho_{2n}}(\beta)$$

für alle Folgen von Entscheidungsfunktionen $\rho_{2n} \in \mathcal{R}(E_n^{loc}(H_{2n}), D)$, $n \in \mathbb{N}$. Somit haben wir eine asymptotische Minimax-Interpretation dafür erhalten, dass die Designfolge $(H_{1n})_{n \in \mathbb{N}}$ schwach lokal asymptotisch mindestens so informativ ist wie $(H_{2n})_{n \in \mathbb{N}}$. Dabei mussten wir jedoch voraussetzen, dass für die Experimente $E_n^{loc}(H_{1n})$ das asymptotische Minimax-Risiko (für Θ_0) durch eine Folge von Entscheidungsfunktionen $(\rho_{1n}^*)_{n \in \mathbb{N}}$ realisiert werden kann. Die Existenz einer solchen Folge kann aber im Allgemeinen nur unter zusätzlichen Voraussetzungen an die Verlustfunktion W bewiesen werden.

4.3.2 Lokalisierte allgemeine Regressionsmodelle

Analog zu den Ausführungen für Lineare Regressionsmodelle in Abschnitt 4.3.1 werden wir nun lokalisierte allgemeine Regressionsmodelle in Hinblick auf asymptotische Aussagen zur Designtheorie untersuchen. Einziger Unterschied ist die Technik, mit der wir die Grenzmodelle (jetzt Gaußexperimente mit Beobachtungsraum $C[0, 1]$) vergleichen. An Stelle der Theorie Torgersens werden wir die Ergebnisse von Luschgy aus Abschnitt 1.5 verwenden. Im Fall allgemeiner Regressionsmodelle sind jedoch viele Designs im starken Sinne LeCam's nicht vergleichbar.

Grundlage der folgenden Ausführungen sind die Sätze 2.22, 2.28 und 2.29, deren Aussagen wir hier kurz wiederholen. Die auftretenden Experimente schreiben wir wieder in Abhängigkeit der Designfolge $(H_n)_{n \in \mathbb{N}}$ bzw. des Grenzdesigns H . In Satz 2.22 wurden um die Nullfunktion lokalisierte Experimente der Form

$$E_n^{loc}(H_n) = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{1}{\sqrt{n}} \frac{f(t_{ni})}{\sigma(t_{ni})}} \mid f \in \mathcal{F} \right\} \right)$$

untersucht, wobei wir diese Experimente in Satz 2.28 auf einen Parameterraum $\Theta \subseteq \mathcal{F}$, so dass $\frac{1}{\sigma}\Theta$ totalbeschränkt in $L^2(H)$ ist, und in Satz 2.29 unter Normalverteilungsannahme auf \mathcal{F}_C , vgl. (2.46) auf Seite 66, eingeschränkt haben. Unter den jeweiligen Voraussetzungen haben wir Grenzexperimente der Form

$$E(H) = (C[0, 1], \mathcal{B}(C[0, 1]), \{P^{Z_f} \mid f \in \mathcal{F}\})$$

mit

$$Z_f(s) = \sqrt{J} \int_{[0, s]} \frac{f \circ H^{-1}(x)}{\sigma \circ H^{-1}(x)} \lambda(dx) + W(s), \quad s \in [0, 1],$$

erhalten bzw. deren eingeschränkte Versionen $E_\Theta(H)$ und $E_{\mathcal{F}_C}(H)$ – in Satz 2.22 als schwachen Grenzwert, in den anderen beiden Sätzen als starken Grenzwert. Diese Grenzerperimente haben nach Lemma 2.17 den Kovarianzkern

$$K_H(f_1, f_2) := J \int_{[a,b]} \frac{f_1 f_2}{\sigma^2} dH \quad \text{für } f_1, f_2 \in \mathcal{F} \text{ (bzw. } \in \Theta \text{ oder } \in \mathcal{F}_C)$$

zur Basis 0. Analog zu den Definitionen 4.11 und 4.13 ergibt sich die folgende

Definition 4.16 *Gegeben seien die Voraussetzungen und Modelle von Satz 2.28 bzw. 2.29. Dann nennen wir eine Folge konkreter Designs $(H_{1n})_{n \in \mathbb{N}}$ **lokal asymptotisch mindestens so informativ** wie eine Folge $(H_{2n})_{n \in \mathbb{N}}$, falls für die zugehörigen Grenzerperimente H_1, H_2 gilt*

$$E_\Theta(H_1) \supseteq E_\Theta(H_2) \quad \text{bzw.} \quad E_{\mathcal{F}_C}(H_1) \supseteq E_{\mathcal{F}_C}(H_2).$$

*Sind die Voraussetzungen und das Modell von Satz 2.22 gegeben, wo schwache Konvergenz der Experimente vorliegt, so nennen wir eine Folge konkreter Designs $(H_{1n})_{n \in \mathbb{N}}$ **schwach lokal asymptotisch mindestens so informativ** wie eine Folge $(H_{2n})_{n \in \mathbb{N}}$, falls für die zugehörigen Grenzerperimente gilt*

$$E(H_1) \supseteq E(H_2).$$

Zur Erläuterung dieser Definition vergleiche man Bemerkung 4.9, Interpretation 4.15 sowie die Abbildungen 4.2 und 4.3. Für das allgemeine Regressionsmodell können wir aber nur eine hinreichende, keine charakterisierende Bedingung für eine informativere Designfolge angeben. Der folgende Satz ist eine direkte Anwendung von Satz 1.29.

Satz 4.17 *Es sei eine der Situationen aus Definition 4.16 gegeben. Weiter seien H_1, H_2 stetige Designs, die von Designfolgen $(H_{1n})_{n \in \mathbb{N}}$, $(H_{2n})_{n \in \mathbb{N}}$ approximiert werden. Die Funktionen $\frac{f}{\sigma}$ mit f aus dem jeweiligen Parameterraum seien paarweise unterscheidbar in $L^2(H_1)$ und in $L^2(H_2)$. K_{H_1} und K_{H_2} seien die zu den Grenzerperimenten gehörigen Kovarianzkern zur Basis 0.*

Ist $K_{H_1} - K_{H_2}$ positiv semidefinit (diese Bedingung ist unabhängig von J und damit unabhängig von der zu Grunde liegenden Fehlerverteilung), so ist die Designfolge $(H_{1n})_{n \in \mathbb{N}}$ (schwach) lokal asymptotisch mindestens so informativ wie die Folge $(H_{2n})_{n \in \mathbb{N}}$.

Ist umgekehrt die Designfolge $(H_{1n})_{n \in \mathbb{N}}$ (schwach) lokal asymptotisch mindestens so informativ wie die Folge $(H_{2n})_{n \in \mathbb{N}}$, so muss $\exp(K_{H_1}) - \exp(K_{H_2})$ positiv semidefinit sein und es muss insbesondere $K_{H_1}(f, f) \geq K_{H_2}(f, f)$ für alle Funktionen f aus dem jeweiligen Parameterraum gelten.

Wir wollen zum Abschluss diese Aussagen sowie die damit verbundenen Schwierigkeiten an einigen Beispielen verdeutlichen. Zunächst werden wir sehr kurz das Analogon zu Beispiel 4.14 für lokalisierte allgemeine Regressionsmodelle anführen.

Beispiel 4.18 Gegeben sei die Situation aus Satz 3.13. Die Designfolge $(H_{1n})_{n \in \mathbb{N}}$ mit Grenzdesign H_1 sei schwach lokal asymptotisch mindestens so informativ wie die Folge $(H_{2n})_{n \in \mathbb{N}}$ mit Grenzdesign H_2 . Dann gilt für die optimalen Testfolgen $\varphi_n^*(H_{1n})$, $\varphi_n^*(H_{2n})$ aus Satz 3.13

$$\text{ARE}(\varphi_n^*(H_{1n}) : \varphi_n^*(H_{2n}) \mid f) = \frac{\langle f, \ell \rangle_{L^2(H_1)}^2 \|\sigma \ell\|_{L^2(H_2)}^2}{\langle f, \ell \rangle_{L^2(H_2)}^2 \|\sigma \ell\|_{L^2(H_1)}^2} \geq 1$$

für alle f aus der Alternative K_ℓ . (Die Pitman-Effizienz ist demnach nicht mehr unabhängig vom Parameter f .)

Beispiel 4.19 Wir betrachten die Regressoren $\mathcal{F} := \{1_{[0, \vartheta]} \mid \vartheta \in [0, 1]\}$ auf dem Versuchsbereich $\mathcal{V} = [0, 1]$ bei gegebener Varianzfunktion $\sigma^2 \equiv 1$. Weiter seien H_1, H_2 stetige Designs mit echt positiver Lebesgue-Dichte. Dann sind die Regressoren unterscheidbar in $L^2(H_1)$ und in $L^2(H_2)$, und für die Kovarianzkerne der zugehörigen Grenzexperimente gilt

$$K_{H_i}(1_{[0, \vartheta_1]}, 1_{[0, \vartheta_2]}) = J \int_{[0, 1]} 1_{[0, \vartheta_1]} 1_{[0, \vartheta_2]} dH_i = JH_i(\min\{\vartheta_1, \vartheta_2\}), \quad i = 1, 2.$$

Für $H_1 \neq H_2$ kann dann $\exp(K_{H_1}) - \exp(K_{H_2})$ niemals positiv semidefinit sein. Somit ist nach Satz 1.29 keines der zu H_1 und H_2 gehörigen Grenzexperimente informativer als das andere. Nach Satz 4.17 lassen sich damit auch Designfolgen, die die Grenzdesigns approximieren, nicht im Sinne von Definition 4.16 ordnen.

Beweis. Wir müssen nur noch nachweisen, dass $\exp(K_{H_1}) - \exp(K_{H_2})$ für $H_1 \neq H_2$ niemals positiv semidefinit sein kann. Sei dazu o.B.d.A. ϑ_1 ein Punkt mit $H_1(\vartheta_1) > H_2(\vartheta_1)$. Wählen wir jetzt $\vartheta_2 = 1$ sowie die Skalare $\alpha_1 = -1$ und $\alpha_2 = 1$, so folgt

$$\begin{aligned} \sum_{i, j=1}^2 \alpha_i \alpha_j (\exp(K_{H_1}(\vartheta_i, \vartheta_j)) - \exp(K_{H_2}(\vartheta_i, \vartheta_j))) &= \\ &= \exp(JH_1(\vartheta_1)) - \exp(JH_2(\vartheta_1)) - 2 \{ \exp(JH_1(\vartheta_1)) - \exp(JH_2(\vartheta_1)) \} \\ &\quad + \exp(JH_1(1)) - \exp(JH_2(1)) \\ &= \exp(JH_2(\vartheta_1)) - \exp(JH_1(\vartheta_1)) < 0. \end{aligned}$$

Man beachte dabei $\exp(JH_1(1)) - \exp(JH_2(1)) = 0$. □

Während in Beispiel 4.19 keine zwei Designs vergleichbar waren, ist das in anderen Modellen zum Teil der Fall. Der Nachweis der Definitheits-Eigenschaft ist aber im Allgemeinen sehr mühsam und soll nicht Gegenstand dieser Arbeit sein, so dass wir uns hier mit einem trivialen Beispiel begnügen.

Beispiel 4.20 Gegeben sei die Situation von Beispiel 4.19, allerdings mit dem Versuchsbereich $\mathcal{V} = [0, 2]$. Damit bringen Versuchspunkte, die im Intervall $(1, 2]$ liegen, keinerlei

Information ein, da die Regressoren $1_{[0,\vartheta]}$, $\vartheta \in [0, 1]$, dort immer verschwinden. Somit ist anschaulich sofort klar, dass das Grenzdesign $H_1(x) = x1_{[0,1]}(x) + 1_{(1,2]}(x)$, $x \in [0, 2]$, das einer Gleichverteilung auf $[0, 1]$ entspricht, informativer ist als $H_2(x) = \frac{x}{2}$, $x \in [0, 2]$, das einer Gleichverteilung auf $[0, 2]$ entspricht. Für die Differenz der Covarianzkerne gilt nach Beispiel 4.19

$$K_{H_1}(1_{[0,\vartheta_1]}, 1_{[0,\vartheta_2]}) - K_{H_2}(1_{[0,\vartheta_1]}, 1_{[0,\vartheta_2]}) = J \min\{\vartheta_1, \vartheta_2\} - \frac{J}{2} \min\{\vartheta_1, \vartheta_2\} = \frac{J}{2} \min\{\vartheta_1, \vartheta_2\}.$$

Diese Funktion ist positiv definit, da sie im Wesentlichen dem Covarianzkern der Brownschen Bewegung auf $[0, 2]$ entspricht. Somit ist nach Satz 4.17 jede Designfolge, die H_1 approximiert lokal asymptotisch informativer im Sinne von Definition 4.16 als eine Designfolge, die H_2 approximiert.

4.4 Entscheidungstheoretische Interpretationen der E -Optimalität

Bereits in Abschnitt 4.1.2 wurde erwähnt, dass es im Allgemeinen kein optimales Design im Sinne der Löwner-Ordnung geben kann, da diese lediglich eine Halbordnung ist. Um dennoch „optimale“ Designs bestimmen zu können, geht man klassischer Weise dazu über, reelle Funktionale der Informationsmatrizen zu betrachten. In Beispiel 4.6 haben wir gesehen, dass die Löwner-Ordnung der Informationsmatrizen der LeCam-Halbordnung \subseteq für Experimente der Form $\mathcal{N}(A_n(H_n)\Sigma_n^{-1}(H_n), I_n)$ entspricht, wodurch wir die starke entscheidungstheoretische Interpretation 4.4 erhalten haben. Ebenso führte uns die LeCam-Halbordnung der in den Abschnitten 4.2 bzw. 4.3 auftretenden Grenzexperimente $\mathcal{N}(\sqrt{J}B^{\frac{1}{2}}(H), I_k)$ zur Löwner-Ordnung der Matrizen $B(H)$. Wir wollen jetzt die Frage untersuchen, welche entscheidungstheoretische Bedeutung die Optimierung eines Funktionals der Informationsmatrizen für solche Experimente hat.

Um die gerade angesprochenen Modelle gemeinsam untersuchen zu können, betrachten wir im Folgenden Experimente der Form

$$\mathcal{N}(A(H), I_n)$$

mit Beobachtungsraum \mathbb{R}^n , Parameterraum \mathbb{R}^k und einer vom Design H abhängigen Modellmatrix $A(H)^\top$. Für zwei Designs H_1 und H_2 gilt dann nach Satz 1.12 die Äquivalenz

$$\mathcal{N}(A(H_1), I_n) \supseteq \mathcal{N}(A(H_2), I_n) \iff A(H_1)A(H_1)^\top \geq_L A(H_2)A(H_2)^\top.$$

Ein klassisches Optimalitätskriterium der Designtheorie ist die E -Optimalität. In dem vorliegenden Kontext wird ein Design H_E E -optimal genannt, wenn für alle Designs H

$$\min_{\|\ell\|=1} \ell^\top A(H_E)A(H_E)^\top \ell \geq \min_{\|\ell\|=1} \ell^\top A(H)A(H)^\top \ell$$

gilt, wobei $\|\cdot\|$ die Euklidnorm des \mathbb{R}^k bezeichnet.

Zunächst wollen wir eine „Minimax“-Interpretation herleiten, die im Wesentlichen der üblichen designtheoretischen Interpretation der E -Optimalität entspricht. Dazu geben wir ein $\ell \in \mathbb{R}^k$ mit $\|\ell\| = 1$ vor und fassen es als Linearform von \mathbb{R}^k nach \mathbb{R} auf. Schränken wir den Parameterraum \mathbb{R}^k eines Experimentes $\mathcal{N}(A(H), I_n)$ auf $\{\ell^\top \beta \mid \beta \in \mathbb{R}^k\}$ ein, d.h. wir interessieren uns nicht mehr für den eigentlichen Parameter $\beta \in \mathbb{R}^k$ sondern nur noch für $\ell^\top \beta$, so erhalten wir das eindimensionale Experiment

$$\mathcal{N}(\ell^\top A(H), I_n) = (\mathbb{R}^n, \mathcal{B}^n, \{N(cA(H)^\top \ell, I_n) \mid c \in \mathbb{R}\})$$

und sprechen vom „zu ℓ gehörigen Telexperiment“. Sind zwei Richtungen ℓ_1, ℓ_2 mit $\|\ell_1\| = \|\ell_2\| = 1$ gegeben, so ist das zu ℓ_1 gehörige Telexperiment $\mathcal{N}(\ell_1^\top A(H), I_n)$ nach Satz 1.12 genau dann im Sinne LeCam's informativer als das zu ℓ_2 gehörige Telexperiment $\mathcal{N}(\ell_2^\top A(H), I_n)$, wenn gilt

$$\ell_1^\top A(H)A(H)^\top \ell_1 \geq \ell_2^\top A(H)A(H)^\top \ell_2.$$

Das zu einer Linearform ℓ_0 mit

$$\ell_0^\top A(H)A(H)^\top \ell_0 = \min_{\|\ell\|=1} \ell^\top A(H)A(H)^\top \ell$$

gehörige Telexperiment $\mathcal{N}(\ell_0^\top A(H), I_n)$ ist somit das uninformativste aus der Menge

$$\{\mathcal{N}(\ell^\top A(H), I_n) \mid \ell \in \mathbb{R}^k, \|\ell\| = 1\}.$$

Grob gesagt lässt jedes andere Experiment dieser Menge bessere Entscheidungen zu.

Interpretation 4.21 (Minimax-Interpretation für die E -Optimalität)

Gegeben seien Experimente der Form $\mathcal{N}(A(H), I_n)$. Ist H_E ein E -optimales Design sowie ℓ_E eine Linearform mit

$$\ell_E^\top A(H_E)A(H_E)^\top \ell_E = \min_{\|\ell\|=1} \ell^\top A(H_E)A(H_E)^\top \ell,$$

und ist H_1 ein weiteres Design sowie ℓ_1 eine Linearform mit

$$\ell_1^\top A(H_1)A(H_1)^\top \ell_1 = \min_{\|\ell\|=1} \ell^\top A(H_1)A(H_1)^\top \ell,$$

so gilt nach Satz 1.12 für die zu ℓ_E bzw. ℓ_1 gehörigen Telexperimente

$$\mathcal{N}(\ell_E^\top A(H_E), I_n) \supseteq \mathcal{N}(\ell_1^\top A(H_1), I_n).$$

Vergleicht man also nur die „uninformativsten eindimensionalen Telexperimente“, so beinhaltet das zu einem E -optimalen Design gehörige die meiste Information im Sinne LeCam's, oder kurz: „Von den eindimensionalen Telexperimenten mit minimaler Information enthält das zum E -optimalen Design gehörige die maximale Information“.

Wir wollen noch eine weitere „Mimimax“-Interpretation der E -Optimalität herleiten. Wir gehen dabei von der Problemstellung aus, den unbekannt Parameter $\beta \in \mathbb{R}^k$ bei gegebener von unten halbstetiger Verlustfunktion $(W_\beta)_{\beta \in \mathbb{R}^k}$ zu schätzen, als Entscheidungsraum ist folglich $D = \mathbb{R}^k$ versehen mit der σ -Algebra \mathcal{B}^k zu wählen. Eine randomisierte Entscheidung $\rho : \mathbb{R}^n \times \mathcal{B}^k \rightarrow [0, 1]$ hat dann das Risiko

$$r_{\rho, H}(\beta) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} W_\beta(a) \rho(x, da) N(A(H)^\top \beta, I_n)(dx), \quad \beta \in \mathbb{R}^k.$$

Lemma 4.22 *Es seien β_1, β_2 zwei feste Parameter mit $\|\beta_1\| = \|\beta_2\| = 1$, für zwei Designs H_1, H_2 gelte die Beziehung*

$$\beta_1^\top A(H_1)A(H_1)^\top \beta_1 \geq \beta_2^\top A(H_2)A(H_2)^\top \beta_2,$$

d.h. das Experiment $\mathcal{N}(\beta_1^\top A(H_1), I_n)$ ist informativer als $\mathcal{N}(\beta_2^\top A(H_2), I_n)$, und die Verlustfunktion erfülle die Invarianzbedingungen

$$W_\gamma(\gamma_1) = W_\gamma(\gamma_2) \quad \text{für alle } \gamma \in \mathbb{R}^k \text{ und alle } \gamma_1, \gamma_2 \in \mathbb{R}^k \text{ mit } \|\gamma_1 - \gamma\| = \|\gamma_2 - \gamma\|, \quad (4.7)$$

d.h. die Verlustfunktion ist auf Kugeloberflächen von Kugeln um γ konstant, und

$$W_{\gamma_1}(c\gamma_1) = W_{\gamma_2}(c\gamma_2) \quad \text{für alle } c \in \mathbb{R} \text{ und alle } \gamma_1, \gamma_2 \in \mathbb{R}^k \text{ mit } \|\gamma_1\| = \|\gamma_2\| = 1. \quad (4.8)$$

Dann existiert zu jeder randomisierten Entscheidungsfunktion $\rho_2 : \mathbb{R}^n \times \mathcal{B}^k \rightarrow [0, 1]$ eine weitere randomisierte Entscheidungsfunktion $\rho_1 : \mathbb{R}^n \times \mathcal{B}^k \rightarrow [0, 1]$ mit

$$r_{\rho_1, H_1}(c\beta_1) \leq r_{\rho_2, H_2}(c\beta_2) \quad \text{für alle } c \in \mathbb{R}.$$

Beweis. Ein beliebiger Markoff-Kern $\rho_2 : \mathbb{R}^n \times \mathcal{B}^k \rightarrow [0, 1]$ definiert mittels

$$\tilde{\rho}_2(x, [c_1, c_2]) := \rho_2(x, A_{c_1, c_2}(\beta_2)),$$

$c_1, c_2 \in \mathbb{R}$, $A_{c_1, c_2}(\beta_2) := \{\beta \in \mathbb{R}^k \mid c_1 \leq \|\beta - \beta_2\| \leq c_2\}$, $x \in \mathbb{R}^n$, einen weiteren Markoff-Kern $\tilde{\rho}_2 : \mathbb{R}^n \times \mathcal{B} \rightarrow [0, 1]$. Betrachten wir für das Teilerperiment $\mathcal{N}(\beta_2^\top A(H_2), I_n)$ das Problem, den unbekannt eindimensionalen Parameter $c \in \mathbb{R}$ unter der Verlustfunktion $W_{c,2}(\tilde{a}) := W_{c\beta_2}(\tilde{a}\beta_2)$, $\tilde{a}, c \in \mathbb{R}$, zu schätzen, so folgt wegen der Invarianzbedingung (4.7) für die zugehörigen Risikofunktionen

$$\begin{aligned} r_{\rho_2, H_2}(c\beta_2) &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} W_{c\beta_2}(a) \rho_2(x, da) N(cA(H_2)^\top \beta_2, I_n)(dx) \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}} W_{c,2}(\tilde{a}) \tilde{\rho}_2(x, d\tilde{a}) N(cA(H_2)^\top \beta_2, I_n)(dx) = r_{\tilde{\rho}_2, H_2}(c) \end{aligned}$$

für alle $c \in \mathbb{R}$. Da $\mathcal{N}(\beta_1^\top A(H_1), I_n)$ nach Voraussetzung informativer ist als $\mathcal{N}(\beta_2^\top A(H_2), I_n)$, existiert (man beachte die Anmerkungen am Ende von Interpretation 4.4) ein Markoff-Kern $\tilde{\rho}_1 : \mathbb{R}^n \times \mathcal{B} \rightarrow [0, 1]$ mit gleichmäßig kleinerem Risiko,

also $r_{\tilde{\rho}_1, H_1}(c) \leq r_{\tilde{\rho}_2, H_2}(c)$ für alle $c \in \mathbb{R}$, vorausgesetzt wir legen dieselbe Verlustfunktion $W_{c,2}$ zu Grunde. Diese stimmt aber wegen Invarianzbedingung (4.8) mit der Verlustfunktion $W_{c,1}(\tilde{a}) := W_{c\beta_1}(\tilde{a}, \beta_1)$, $\tilde{a}, c \in \mathbb{R}$, überein, so dass für jeden Markoff-Kern $\rho_1 : \mathbb{R}^n \times \mathcal{B}^k \rightarrow [0, 1]$ mit $\rho_1(x, A_{c_1, c_2}(\beta_1)) = \tilde{\rho}_1(x, [c_1, c_2])$, $c_1, c_2 \in \mathbb{R}$, $x \in \mathbb{R}^n$, wie oben die Gleichung $r_{\rho_1, H_1}(c\beta_1) = r_{\tilde{\rho}_1, H_1}(c)$ für alle $c \in \mathbb{R}$ gilt, und es folgt

$$r_{\rho_1, H_1}(c\beta_1) = r_{\tilde{\rho}_1, H_1}(c) \leq r_{\tilde{\rho}_2, H_2}(c) = r_{\rho_2, H_2}(c\beta_2)$$

für alle $c \in \mathbb{R}$. Dabei ist $A_{c_1, c_2}(\beta_1)$ analog zu $A_{c_1, c_2}(\beta_2)$ definiert. \square

Bemerkung 4.23 Auf Grund der Invarianzbedingungen 4.7 und 4.8 muss die Verlustfunktion von der Form

$$W_\beta(\tilde{\beta}) = w(\|\beta - \tilde{\beta}\|), \quad \beta, \tilde{\beta} \in \mathbb{R}^k,$$

mit einer von unten halbstetigen und nach unten beschränkten Funktion $w : \mathbb{R} \rightarrow \mathbb{R}$ sein.

Wir geben uns nun ein Design H sowie ein Experiment $\mathcal{N}(A(H), I_n)$ vor, und C sei die Menge aller β_0 mit

$$\beta_0^\top A(H)A(H)^\top \beta_0 = \min_{\|\beta\|=1} \beta^\top A(H)A(H)^\top \beta.$$

Desweiteren existiere ein Markoff-Kern $\tilde{\rho} : \mathbb{R}^n \times \mathcal{B}^k \rightarrow [0, 1]$ und ein $\tilde{\beta} \in \mathbb{R}^k$ mit

$$r_{\tilde{\rho}, H}(\tilde{\beta}) = \min_{\rho} \max_{\beta \in \mathbb{R}^k} r_{\rho, H}(\beta),$$

wobei ρ die Menge aller Markoff-Kerne von \mathbb{R}^n nach \mathcal{B}^k durchläuft. Dann gilt wegen Lemma 4.22 (angewandt auf $H_1 = H_2 = H$)

$$\tilde{\beta} \in [C],$$

d.h. der Minimax-Schaden wird für ein Element der linearen Hülle $[C]$ von C angenommen. Sind zwei Designs H_i , $i = 1, 2$, Experimente $\mathcal{N}(A(H_i), I_n)$ sowie β_i mit

$$\beta_i^\top A(H_i)A(H_i)^\top \beta_i = \min_{\|\beta\|=1} \beta^\top A(H_i)A(H_i)^\top \beta$$

gegeben, gilt

$$\beta_1^\top A(H_1)A(H_1)^\top \beta_1 \geq \beta_2^\top A(H_2)A(H_2)^\top \beta_2,$$

und existieren $\tilde{\rho}_i : \mathbb{R}^n \times \mathcal{B}^k \rightarrow [0, 1]$ und $\tilde{\beta}_i \in \mathbb{R}^k$ mit

$$r_{\tilde{\rho}_i, H_i}(\tilde{\beta}_i) = \min_{\rho} \max_{\beta \in \mathbb{R}^k} r_{\rho, H_i}(\beta),$$

so folgt aus den gerade angeführten Überlegungen bei erneuter Anwendung von Lemma 4.22

$$\min_{\rho} \max_{\beta \in \mathbb{R}^k} r_{\rho, H_1}(\beta) \leq \min_{\rho} \max_{\beta \in \mathbb{R}^k} r_{\rho, H_2}(\beta).$$

Interpretation 4.24 (Mimimax-Interpretation für die E -Optimalität)

Für Experimente der Form $\mathcal{N}(A(H), I_n)$ soll der unbekannte Parameter $\beta \in \mathbb{R}^k$ geschätzt werden, wobei eine von unten halbstetige Verlustfunktion $(W_\beta)_{\beta \in \mathbb{R}^k}$ gegeben ist, die die Invarianzbedingungen (4.7) und (4.8) erfüllt. Ist H_E ein E -optimales Design, H ein weiteres Design und existieren $\rho_E, \tilde{\rho} : \mathbb{R}^n \times \mathcal{B}^k \rightarrow [0, 1]$ und $\beta_E, \tilde{\beta} \in \mathbb{R}^k$ mit

$$r_{\rho_E, H_E}(\beta_E) = \min_{\rho} \max_{\beta \in \mathbb{R}^k} r_{\rho, H_E}(\beta) \quad \text{bzw.} \quad r_{\tilde{\rho}, H}(\tilde{\beta}) = \min_{\rho} \max_{\beta \in \mathbb{R}^k} r_{\rho, H}(\beta),$$

so gilt

$$\min_{\rho} \max_{\beta \in \mathbb{R}^k} r_{\rho, H_E}(\beta) \leq \min_{\rho} \max_{\beta \in \mathbb{R}^k} r_{\rho, H}(\beta),$$

d.h. der Minimax-Schaden für das gegebene Schätzproblem ist für das E -optimale Design am kleinsten.

Kapitel 5

Lineare Regressionsmodelle mit unbekannter Fehlerverteilung

Im bisherigen Verlauf dieser Arbeit sind uns bereits an verschiedenen Stellen Regressionsmodelle begegnet, bei denen die wahre Fehlerverteilung unbekannt war. So zum Beispiel die Linearen Regressionsmodelle aus Beispiel 2.2, wo an die Fehlerverteilung lediglich der Erwartungswert 0 und die Varianz 1 vorausgesetzt wurden. Oder die Regressionsmodelle aus Formel (2.13) auf Seite 43 mit denselben Voraussetzungen an die Fehlerverteilung.

In Kapitel 4 konnten wir feststellen, dass die designtheoretischen Überlegungen unabhängig von der Fehlerverteilung waren, solange diese nur den Voraussetzungen des jeweils zu Grunde liegenden Satzes genügte.

Im Gegensatz dazu standen jedoch die Betrachtungen aus Kapitel 3. Hier mussten wir feststellen, dass die Fehlerverteilung bei den dort betrachteten Testproblemen eine entscheidende Rolle spielt. Die asymptotisch optimale Testfolge aus Satz 3.1 hing direkt von der Dichte q der Fehlerverteilung ab, genauer gesagt vom sogenannten „Score“ $\frac{q'}{q}$, und von der Fisher-Information J der Fehlerverteilung, vgl. (2.20) auf Seite 48. Die asymptotische Güte dieser optimalen Testfolge war zwar nicht mehr von der Dichte, wohl aber von der Fisher-Information der Fehlerverteilung abhängig. In Satz 3.9 haben wir gesehen, dass der Likelihood-Quotienten-Test, dessen Teststatistik nicht mehr von der Fehlerverteilung abhängt, nicht dieselbe Effizienz erreicht wie die asymptotisch optimale Testfolge. Den Effizienzverlust konnten wir in Form der Pitman-Effizienz genau beziffern.

In diesem Kapitel wollen wir nun die Frage untersuchen, ob es Testverfahren gibt, die wie die Tests aus Satz 3.1 asymptotisch optimal sind, aber nicht mehr von der Fehlerverteilung abhängen. Wir suchen also eine Testfolge, die für möglichst viele verschiedene Fehlerverteilungen die optimale Grenzgüte aus Satz 3.1 erreicht, und zwar für dieselben Testprobleme sowie dieselbe Lokalisierung wie in Kapitel 3. Gibt es solche „adapti-

ven“ Testverfahren, so muss die Fehlerverteilung nicht mehr als bekannt vorausgesetzt werden, sondern sie muss lediglich in der Klasse von Verteilungen liegen, für die das gegebene Testverfahren adaptiv ist.

Ausgangspunkt für solche Tests sind adaptive Schätzverfahren, die wir in Abschnitt 5.2 einführen und mit deren Hilfe wir in Abschnitt 5.3 adaptive asymptotisch optimale Tests konstruieren. Wie wir sehen werden, ist die Konstruktion solcher Schätzer sehr aufwändig und die Implementierung der zugehörigen Tests mühsam. Daher liegt es nahe, diese Verfahren durch Simulationen mit dem asymptotisch nicht optimalen Likelihood-Quotienten-Test aus (3.8) zu vergleichen. Wir wollen in Abschnitt 5.4 an Fallstudien untersuchen, inwieweit es gerechtfertigt ist, den sehr viel einfacheren LQ-Test dem adaptiven asymptotisch optimalen Test vorzuziehen.

5.1 Modellbeschreibung

Zunächst werden wir das zu untersuchende Modell spezifizieren. Es mögen alle Bezeichnungen aus Abschnitt 2.1 gelten, wo Regressionsmodelle eingeführt wurden. Wir beschränken unsere Überlegungen in diesem Kapitel auf *Lineare Regressionsmodelle*.

Wie bisher seien die Versuchspunkte t_{ni} und die Varianzfunktion σ^2 fest vorgegeben, ebenso die Regressoren f_1, \dots, f_k . Wir setzen zur Abkürzung wieder

$$f := (f_1, \dots, f_k)^\top \quad \text{bzw.} \quad \frac{f}{\sigma} := \left(\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma} \right)^\top.$$

Beobachtet werden die Zufallsvariablen

$$Y_{ni} = f^\top(t_{ni})\beta + \sigma(t_{ni})\epsilon_{ni}, \quad i = 1, \dots, n,$$

wobei ϵ_{ni} , $i = 1, \dots, n$, stochastisch unabhängige Zufallsvariablen mit identischer, jetzt aber *unbekannter* Verteilung Q seien.

Wir interessieren uns für den unbekannt Parameter $\beta \in \mathbb{R}^k$, den sogenannten **Hauptparameter**. Die unbekannt Fehlerverteilung Q sei aus einer Menge \mathcal{G} von Wahrscheinlichkeitsmaßen. Da dieser Parameter, der die Komplexität des Modells erhöht, für uns nicht von primärem Interesse ist, wird er als **Nebenparameter** oder **Störparameter** bezeichnet. Insgesamt liegt unseren statistischen Experimenten also der Parameterraum $\mathbb{R}^k \times \mathcal{G}$ zu Grunde. Dieselbe suffiziente Transformation T_n wie in Abschnitt 2.1 führt uns zu den Zufallsvariablen

$$Z_{ni} = \frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})} + \epsilon_{ni}, \quad i = 1, \dots, n,$$

und damit zu den Experimenten

$$E_n = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sigma(t_{ni})}} \mid (\beta, Q) \in \mathbb{R}^k \times \mathcal{G} \right\} \right).$$

Um eine nichtentartete Asymptotik zu erhalten, muss wieder im Parameterraum lokalisiert werden. In Hinblick auf die uns interessierenden Testprobleme werden wir nur im Hauptparameter und dort wie bisher mit Rate $\frac{1}{\sqrt{n}}$ um den Ursprung lokalisieren. Wir erhalten somit die lokalisierten Experimente

$$E_n^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}} \mid (\beta, Q) \in \mathbb{R}^k \times \mathcal{G} \right\} \right), \quad (5.1)$$

d.h. Zufallsvariablen

$$Z_{ni}^{loc} = \frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})} + \epsilon_{ni}, \quad i = 1, \dots, n, \quad (5.2)$$

mit $\epsilon_{ni} \sim Q$.

Ist die Fehlerverteilung Q bekannt, d.h. es liegt das Modell aus (2.24) auf Seite 52 vor, so werden wir künftig $E_n^{loc}(Q)$ schreiben. Für die Fisher-Information J der Fehlerverteilung, vgl. (2.20) auf Seite 48, schreiben wir künftig $J(Q)$.

Da wir nur im Hauptparameter lokalisieren, werden die Experimente E_n^{loc} im Allgemeinen nicht schwach gegen einen nichttrivialen Grenzwert konvergieren, sondern nur die Experimente $E_n^{loc}(Q)$, vgl. Satz 2.16. Für die Untersuchung der in der Einleitung erwähnten Testprobleme ist dies jedoch ausreichend.

5.2 Adaptive Schätzverfahren

In diesem Abschnitt werden wir kurz die Theorie adaptiver Schätzverfahren skizzieren. Um die Definition eines adaptiven Schätzers zu motivieren, gehen wir zunächst auf die klassische Theorie effizienter und asymptotisch effizienter Schätzer ein. Im Mittelpunkt stehen jedoch die Schätzer aus der Arbeit Koul und Susarla (1983), die die Grundlage für die im darauf folgenden Abschnitt 5.3 untersuchten Tests bilden.

5.2.1 Effiziente und asymptotisch effiziente Schätzer

Ausführliche und sehr gute Darstellungen zu diesem Thema finden sich in den Büchern Bickel, Klaassen, Ritov und Wellner (1993), Ibragimov und Has'minskii (1981) und vor

allem Witting und Müller-Funk (1995). Der Rahmen und die Darstellung der folgenden Definitionen unterscheidet sich jedoch von allen drei Büchern, da wir die Kontinuität mit den bisherigen Ausführungen erhalten wollen.

Zuerst führen wir effiziente Schätzer ein. Grob gesprochen sind das Schätzer, für die in der Cramér-Rao-Ungleichung das Gleichheitszeichen gilt.

Definition 5.1 (vgl. Ibragimov und Has'minskii (1981), S. 65) Seien $(H, \langle \cdot, \cdot \rangle)$ ein endlichdimensionaler Hilbertraum und $F = (\mathcal{X}, \mathcal{A}, \{P_h \mid h \in H\})$ ein Experiment, das durch das σ -endliche Maß ν dominiert wird. Mit $p(x; h) := \frac{dP_h}{d\nu}(x)$, $x \in \mathcal{X}$, $h \in H$, bezeichnen wir die zugehörigen Dichten bezüglich ν . Das Experiment F heißt **regulär**, falls die folgenden drei Bedingungen erfüllt sind:

1. Die Abbildung $H \rightarrow \mathbb{R}$, $h \mapsto p(x; h)$ ist für ν -fast alle $x \in \mathcal{X}$ stetig.
2. Die Abbildung $H \rightarrow L^2(\nu)$, $h \mapsto \sqrt{p(\cdot; h)}$ ist in jedem Punkt $h \in H$ Fréchet-differenzierbar. Die Ableitungen $\psi(h)$, $h \in H$, sind dann stetige lineare Operatoren $H \rightarrow L^2(\nu)$. Ein Experiment mit dieser Eigenschaft heißt **L^2 -differenzierbar**.
3. Die Abbildung $H \rightarrow \mathcal{L}(H; L^2(\nu))$, $h \mapsto \psi(h)$ ist stetig. Dabei bezeichne $\mathcal{L}(H; L^2(\nu))$ den Raum aller stetigen linearen Abbildungen von H nach $L^2(\nu)$ versehen mit der kanonischen Abbildungsnorm. Man beachte, dass auf Grund der endlichen Dimension von H alle linearen Operatoren auch stetig sind.

Sei nun F ein reguläres Experiment, und die Dimension von H sei gleich k . Wir geben uns eine Orthonormalbasis (ONB) U von H vor und bezeichnen mit $h_U \in \mathbb{R}^k$ den zu einem $h \in H$ gehörigen Koordinatenvektor bezüglich U . Für festes $h_0 \in H$ existieren dann Funktionen $g_1(h_0), \dots, g_k(h_0) \in L^2(\nu)$, so dass mit der Schreibweise $g(h_0) := (g_1(h_0), \dots, g_k(h_0))^T$ für die Ableitung $\psi(h_0)$ gilt

$$\psi(h_0)(h) = h_U^T g(h_0), \quad h \in H.$$

Jetzt definieren wir eine lineare Abbildung $I(h_0) : H \rightarrow H$ durch die Abbildungsmatrix

$$[I(h_0)]_U := 4 \int_{\mathcal{X}} g(h_0) g(h_0)^T d\nu \tag{5.3}$$

bezüglich der Basis U . Der lineare Operator $I(h_0)$ ist *symmetrisch, positiv semidefinit, unabhängig vom dominierenden Maß ν* , vgl. Ibragimov und Has'minskii (1981), S. 65, und *unabhängig von der gewählten ONB U* , was man mittels einer Basistransformation mit orthogonalen Matrizen leicht einsieht.

Definition 5.2 (Fisher-Informationsoperator) Ist F ein reguläres Experiment, so heißt der durch (5.3) definierte Operator $I(h_0)$ **Fisher-Informationsoperator** von F im Parameter $h_0 \in H$.

Bemerkung 5.3 (vgl. Ibragimov und Has'minskii (1981), S. 64) Ist $H = \mathbb{R}^k$, so stimmt $I(h_0)$ mit der „klassischen“ Fischer-Informationsmatrix überein, d.h.

$$I(h_0) = \int_{\mathcal{X}} \frac{\partial p}{\partial h}(x; h_0) \left(\frac{\partial p}{\partial h}(x; h_0) \right)^\top p^{-2}(x; h_0) P_{h_0}(dx),$$

sofern der Ausdruck auf der rechten Seite existiert.

Beispiel 5.4 (vgl. Ibragimov und Has'minskii (1981), S. 71)

Gegeben sei das Experiment $F = \mathcal{N}(I_H, C)$, vgl. (1.5), mit positiv definiten Kovarianz C und dem endlichdimensionalen Hilbertraum H als Beobachtungsraum und zugleich als Parameterraum. Dieses Experiment ist regulär mit dem Fisher-Informationsoperator

$$I(h) = C^{-1},$$

der für alle $h \in H$ derselbe ist.

Jetzt können wir die berühmte Cramér-Rao-Ungleichung im Rahmen endlichdimensionaler Hilberträume formulieren. Der Beweis verläuft analog zum Fall des Parameterraums \mathbb{R}^k , wie er z.B. in Ibragimov und Has'minskii (1981), S. 73, zu finden ist. Desweiteren beschränken wir uns auf den Fall erwartungstreuer Schätzer. Man vergleiche Definition A.14 für den Erwartungswert und die Kovarianz von hilbertraumwertigen Zufallsvariablen.

Satz 5.5 (Cramér-Rao-Ungleichung)

Seien $F = (\mathcal{X}, \mathcal{A}, \{P_h \mid h \in H\})$ ein reguläres Experiment und $T : \mathcal{X} \rightarrow H$ ein erwartungstreuer Schätzer für $h \in H$. Weiter sei $\|T - h\|$ für jedes $h \in H$ quadratisch integrierbar bezüglich P_h , und für jedes $h \in H$ seien die Integralwerte in einer Umgebung von h beschränkt. Die Fisher-Informationsoperatoren $I(h)$ seien für alle $h \in H$ positiv definit. Dann gilt

$$\text{Cov}_h(T - h) \geq_L I^{-1}(h)$$

für alle $h \in H$, d.h. die linearen symmetrischen Operatoren $\text{Cov}_h(T - h) - I^{-1}(h)$ sind positiv semidefinit.

Bemerkung 5.6 Ist eine Basis U von H gegeben und sind $[\text{Cov}_h(T - h)]_U$ bzw. $[I(h)]_U$ die Abbildungsmatrizen von $\text{Cov}_h(T - h)$ bzw. $I(h)$ bezüglich der Basis U , so lässt sich die Cramér-Rao-Ungleichung mittels der Löwner-Ordnung in der Form

$$[\text{Cov}_h(T - h)]_U \geq_L [I(h)]_U^{-1}, \quad h \in H,$$

schreiben.

Die Cramér-Rao-Ungleichung motiviert nun die folgende Definition eines effizienten Schätzers. Hierbei beschränken wir uns wiederum auf erwartungstreue Schätzer.

Definition 5.7 (Effizienter Schätzer) *Unter den Voraussetzungen der Cramér-Rao-Ungleichung 5.5 nennen wir einen erwartungstreuen Schätzer $T : \mathcal{X} \rightarrow H$ **effizient**, falls gilt*

$$\text{Cov}_h(T - h) = I^{-1}(h) \quad \text{für alle } h \in H.$$

Jetzt werden wir den Begriff der asymptotisch effizienten Schätzfolge einführen, der im Gegensatz zum Begriff des effizienten Schätzers nicht eindeutig definiert ist. In Ibragimov und Has'minskii (1981) wird z.B. asymptotische Effizienz von Schätzern im Sinn von Fisher, Bahadur, Rao und Wolfowitz unterschieden. Der für unsere Belange passende Effizienzbegriff ist der von Fisher. Wir werden diesen in etwas anderer Form definieren als in der zitierten Literatur.

Definition 5.8 (Asymptotisch effiziente Schätzfolge (im Sinn von Fisher))

Wir betrachten wieder den Fall eines endlichdimensionalen Hilbertraums $(H, \langle \cdot, \cdot \rangle)$. Sei $F_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,h} \mid h \in H\}) \in \mathcal{E}(H)$ eine Folge von Experimenten, die schwach gegen das Gaußexperiment $\mathcal{N}(I_H, C)$ mit positiv definiter Covarianz C konvergiert, also $F_n \xrightarrow{w} \mathcal{N}(I_H, C)$.

*Eine Folge von Schätzern $T_n : \mathcal{X}_n \rightarrow H$ heißt **asymptotisch effizient**, falls für alle $h \in H$ gilt*

$$\mathcal{L}(T_n \mid P_{n,h}) \xrightarrow{D} N(h, C), \quad (5.4)$$

falls also die Verteilung der Schätzer T_n unter den Maßen $P_{n,h}$ schwach gegen eine Normalverteilung auf H mit Erwartungswert h und Covarianz C konvergiert.

Bemerkung 5.9

1. Wie wir in vielen Beispielen gesehen haben, muss man i.A. im Parameterraum lokalisieren, um schwache Konvergenz von Experimenten nachweisen zu können. Liegt eine Lokalisierung vor, so sind die $h \in H$ *lokale Parameter*. In diesem Zusammenhang wird die Schätzfolge aus Definition 5.8 auch **lokal asymptotisch effizient** genannt.
2. Versieht man den Hilbertraum H mit dem durch C^{-1} induzierten Innenprodukt $[h_1, h_2] := \langle h_1, C^{-1}h_2 \rangle$, $h_1, h_2 \in H$, so ist $\mathcal{N}(I_H, C)$ äquivalent (im Sinn von Le-Cam's Δ -Metrik) zum Standard-Gaußshft auf $(H, [\cdot, \cdot])$. Man vergleiche dazu nur die Hellingertransformierten aus Formel (1.9) und Beispiel 1.15 sowie Satz 1.7. Ebenso ergibt sich die Aussage, wenn man beachtet, dass die Covarianz einer hilbertraumwertigen Zufallsvariablen vom Innenprodukt abhängt und sich gemäß Lemma A.16 beim Wechsel des Innenproduktes verändert und dass bei Innenproduktwechseln dieser Form die Normalverteilung definitionsgemäß erhalten bleibt.
3. Nach Beispiel 5.4 ist die Covarianz C der Grenzverteilung aus (5.4) gleich der Inversen des Fisher-Informationsoperators im Grenzexperiment $\mathcal{N}(I_H, C)$.

Um die Definition einer asymptotisch effizienten Schätzfolge ähnlich zu motivieren wie die des effizienten Schätzers, bräuchten wir ein asymptotisches Äquivalent zur Cramér-Rao-Ungleichung. Dies erweist sich jedoch als weitaus schwierigeres Problem. Ein Ansatz zu dessen Lösung ist der *Satz von LeCam-Bahadur*, wie man ihn (allerdings nur für den Fall des „Einstichprobenproblems“ mit Parameterraum \mathbb{R}^k) in Witting und Müller-Funk (1995), Satz 6.199, findet. Dieser besagt, dass unter den Voraussetzungen von Definition 5.8 für jede asymptotisch normalverteilte Schätzfolge \tilde{T}_n mit Grenzverteilung $N(h, D(h))$, $h \in H$, die Operatoren $D(h) - C$ positiv semidefinit sind für alle $h \in H$ abgesehen von einer λ_H -Nullmenge. Dabei sei λ_H das Lebesgue-Maß auf dem endlichdimensionalen Hilbertraum H . Asymptotisch normale Schätzer, für die diese λ_H -Nullmenge nicht leer ist, werden **supereffizient** genannt. Hierzu vergleiche man das Beispiel von Hodges, das in Witting und Müller-Funk (1995) in Form von Beispiel 6.32 wiedergegeben ist. Unter geeigneten Regularitätsbedingungen lässt sich die Existenz supereffizienter Schätzer jedoch ausschließen, vgl. z.B. Witting und Müller-Funk (1995), Satz 6.202.

Weitergehende Überlegungen zur Existenz einer asymptotischen Cramér-Rao-Schranke basieren auf dem Faltungssatz von Hájek, auf den hier nicht weiter eingegangen werden soll.

Wir wollen diesen Abschnitt mit einer Bemerkung beschließen.

Bemerkung 5.10

1. Unter geeigneten Regularitätsvoraussetzungen sind Maximum-Likelihood-Schätzer asymptotisch effizient, vgl. Witting und Müller-Funk (1995), Satz 6.203.
2. Auf Grund der endlichen Dimension von H gibt es immer eine asymptotisch effiziente Schätzfolge:

Nach Bemerkung 5.9 konvergieren die F_n schwach gegen den Standard-Gaußshift auf $(H, [\cdot, \cdot])$. Nach Strasser (1985), Korollar 80.6, existiert wegen $\dim(H) < \infty$ immer eine zentrale Folge X_n (vgl. Korollar 1.38). Diese kann nach Strasser (1985), Beispiel 68.5, in der Form $X(h) = [h, \kappa_n]$ mit Zufallsvariablen $\kappa_n : \mathcal{X}_n \rightarrow H$ geschrieben werden. Diese Zufallsvariablen κ_n , $n \in \mathbb{N}$, bilden jetzt nach Strasser (1985), Satz 80.12, eine asymptotisch effiziente Schätzfolge.

5.2.2 Adaptive Schätzer

Bevor wir in einem geeigneten Rahmen auf den Begriff eines adaptiven Schätzers eingehen, wollen wir auf das Buch Bickel et al. (1993) hinweisen, das wohl die bislang umfassendste Arbeit auf dem Gebiet der adaptiven Schätztheorie darstellt. Weiter müssen wir festhalten, dass es, wie schon beim Begriff der asymptotisch effizienten Schätzfolge, auch hier keine einheitliche Definition gibt.

Die Grundidee des adaptiven Schätzens geht auf die Arbeit Stein (1956) zurück. Stein

wollte einen endlichdimensionalen Hauptparameter in Anwesenheit eines unbekanntem unendlichdimensionalen Nebenparameters asymptotisch ebenso gut schätzen, wie wenn dieser Nebenparameter bekannt wäre. Erste adaptive Schätzer für einen Lokationsparameter einer unbekanntem Verteilung werden in Beran (1974) und Stone (1975) vorgestellt. In Bickel (1982) werden adaptive Schätzer für eine größere Klasse von Problemen konstruiert. In Koul und Susarla (1983) werden adaptive Schätzer für Regressionsmodelle, die unsere Modelle aus Abschnitt 5.1 einschließen, untersucht. Auf diese Arbeit werden wir im nächsten Abschnitt 5.2.3 genauer eingehen, da sich der dort verwendete Adaptionsbegriff als der für unsere Ziele passende erweist.

Auch bei unserer Definition eines adaptiven Schätzers werden wir von den üblichen Formulierungen abweichen und sie in LeCam's Rahmen schwach konvergenter Experimente einbetten.

Definition 5.11 (Adaptive Schätzfolge) *Es seien $(H, \langle \cdot, \cdot \rangle)$ ein endlichdimensionaler Hilbertraum und $\Theta \neq \emptyset$ ein beliebiger Parameterraum. Weiter seien $F_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,(h,\theta)} \mid (h,\theta) \in H \times \Theta\}) \in \mathcal{E}(H \times \Theta)$ eine Folge von Experimenten. Für jedes feste $\theta \in \Theta$ sollen die Experimente $F_n(\theta) := (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,(h,\theta)} \mid h \in H\}) \in \mathcal{E}(H)$ schwach gegen die Gaußexperimente $\mathcal{N}(I_H, C(\theta))$ mit positiv definiten Covarianzen $C(\theta)$ konvergieren, also $F_n(\theta) \xrightarrow{w} \mathcal{N}(I_H, C(\theta))$.*

*Eine Folge von Schätzern $T_n : \mathcal{X}_n \rightarrow H$ heißt **(Θ -)adaptiv**, falls für alle $h \in H$ und alle $\theta \in \Theta$ gilt*

$$\mathcal{L}(T_n \mid P_{n,(h,\theta)}) \xrightarrow{\mathcal{D}} N(h, C(\theta)),$$

falls also die Verteilung der Schätzer T_n unter den Maßen $P_{n,(h,\theta)}$ schwach gegen eine Normalverteilung mit Erwartungswert h und Covarianz $C(\theta)$ konvergiert.

Bemerkung 5.12

1. Eine Θ -adaptive Schätzfolge ist für jedes $\theta \in \Theta$ im Sinne von Definition 5.8 *asymptotisch effizient*. Im nächsten Abschnitt 5.2.3 sowie in Satz 5.14 werden wir Schätzer kennenlernen, die für jedes $\theta \in \Theta$ asymptotisch normalverteilt sind, was man auch als eine „Anpassungseigenschaft“ bezeichnen kann, die aber nicht für jedes θ die asymptotische Covarianz $C(\theta)$ besitzen. Wir werden solche Schätzfolgen *nicht* als adaptiv bezeichnen.
2. Die schwache Konvergenz der Experimente $F_n(\theta)$ setzt in Hinblick auf sinnvolle Beispiele eine Lokalisierung im Hauptparameter voraus. Man wird in jedem Experiment $F_n(\theta)$ denselben Lokalisierungspunkt $h_0 \in H$ wählen. Der Parameter $h \in H$, der dann geschätzt werden soll, ist somit ein *lokaler Parameter*. Man könnte einen adaptiven Schätzer dann auch als **in h_0 lokal Θ -adaptiv** bezeichnen.

So gesehen ist unsere Definition eine Spezialisierung der Definition, die in Bickel (1982) gegeben wird. Dort wird zwar auch im Hauptparameter lokalisiert, es wird aber der unbekanntem Lokalisierungspunkt geschätzt, nicht der lokale Parameter.

3. Neuere Adaptionbegriffe setzen nicht nur eine Lokalisierung im Hauptparameter, sondern auch eine im Nebenparameter voraus. Eine solche Definition findet sich z.B. in Begun, Hall, Huang und Wellner (1983). In diesem Zusammenhang sind auch besonders die Arbeiten Schick (1986) und Schick (1987) zu erwähnen. Für unsere Zwecke ist obige Definition jedoch völlig ausreichend, wie wir im Abschnitt 5.3 sehen werden.
4. Die Frage nach der Existenz adaptiver Schätzer ist ein schwieriges Problem, auf das in den zitierten Arbeiten in verschiedenster Weise eingegangen wird. Der Nebenparameterraum Θ darf dabei nicht zu groß sein. Wir wollen hier nicht weiter auf diese Problematik eingehen und lediglich feststellen, dass die in unserem Beispiel der Linearen Regression auftretenden Nebenparameterräume klassisch gewählt sind und nicht wesentlich vergrößert werden können.

Wir sind nun an einem Punkt, wo wir die skizzierte Theorie auf die in Abschnitt 5.1 beschriebenen Linearen Regressionsmodelle E_n^{loc} anwenden wollen. Dazu geben wir zunächst an, wann ein Schätzer für diese Modelle \mathcal{G} -adaptiv ist.

Satz 5.13 (Adaptive Schätzer im Linearen Regressionsmodell)

Gegeben sei eine Menge \mathcal{G} von Fehlerverteilungen Q , die allesamt den Voraussetzungen 2.9 genügen sollen, mit endlicher positiver Fisher-Information $J(Q)$. Weiterhin sollen Designfolge, Grenzdesign, Regressoren und Varianzfunktion die Voraussetzungen von Satz 2.16 erfüllen, und B bezeichne die Matrix aus (2.26) auf Seite 53. Wir betrachten die lokalisierten Experimente

$$E_n^{loc} = \left(\mathbb{R}^n, \mathcal{B}^n, \left\{ \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}} \mid (\beta, Q) \in \mathbb{R}^k \times \mathcal{G} \right\} \right)$$

aus (5.1). Eine Folge von Schätzern $T_n : \mathbb{R}^n \rightarrow \mathbb{R}^k$, $n \in \mathbb{N}$, für $\beta \in \mathbb{R}^k$ ist genau dann \mathcal{G} -adaptiv, wenn gilt

$$\mathcal{L} \left(T_n \mid \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}} \right) \xrightarrow{\mathcal{D}} N \left(\beta, \frac{1}{J(Q)} B^{-1} \right) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}. \quad (5.5)$$

Beweis. Es ist lediglich die Grenzverteilung aus Definition 5.11 zu identifizieren. Nach Satz 2.16 konvergieren für feste Fehlerverteilung $Q \in \mathcal{G}$ die lokalisierten Experimente $E_n^{loc}(Q)$ schwach gegen das Experiment $\mathcal{N} \left(\sqrt{J(Q)} B^{\frac{1}{2}}, I_k \right)$. Dieses Grenzexperiment ist aber äquivalent zu $\mathcal{N} \left(I_k, \frac{1}{J(Q)} B^{-1} \right)$. \square

Es stellt sich die Frage, ob der aus der Theorie Linearer Modelle bekannte Gauß-Markoff-Schätzer in unserem Sinne adaptiv ist. Die Antwort ist bekannt: Der Gauß-Markoff-Schätzer besitzt zwar eine „Adaptionseigenschaft“, wie sie in Bemerkung 5.12 ange-

sprochen wurde, ist aber nicht asymptotisch effizient und somit nicht in unserem Sinne adaptiv.

Satz 5.14 (Asymptotik des Gauß-Markoff-Schätzers) *Gegeben sei eine Menge \mathcal{G} von Fehlerverteilungen Q , die alle zentriert seien und eine endliche Varianz $\tau^2(Q)$ besitzen, d.h.*

$$\int_{\mathbb{R}} x Q(dx) = 0 \quad \text{und} \quad \int_{\mathbb{R}} x^2 Q(dx) = \tau^2(Q) < \infty.$$

Weiterhin sollen Designfolge, Grenzdesign, Regressoren und Varianzfunktion die Voraussetzungen von Satz 2.16 erfüllen, und B bezeichne die Matrix aus (2.26). Wir betrachten die lokalisierten Experimente E_n^{loc} , also die Zufallsvariablen

$$Z_{ni}^{loc} = \frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})} + \epsilon_{ni}, \quad i = 1, \dots, n,$$

aus (5.2) mit $\epsilon_{ni} \sim Q$, die dem Linearen Modell

$$Z_n^{loc} = \frac{1}{\sqrt{n}} X_n \beta + \epsilon_n, \quad \beta \in \mathbb{R}^k, \quad Q \in \mathcal{G},$$

mit $X_n = \Sigma_n^{-1} A_n^\top$ und

$$A_n^\top = \begin{pmatrix} f_1(t_{n1}) & \cdots & f_k(t_{n1}) \\ \vdots & \ddots & \vdots \\ f_1(t_{nn}) & \cdots & f_k(t_{nn}) \end{pmatrix} \quad \text{bzw.} \quad \Sigma_n^2 = \begin{pmatrix} \sigma^2(t_{n1}) & & 0 \\ & \ddots & \\ 0 & & \sigma^2(t_{nn}) \end{pmatrix}$$

wie in (2.8) bzw. (2.9) entsprechen.

Dann gilt für den Gauß-Markoff-Schätzer $\mathbb{R}^n \rightarrow \mathbb{R}^k$, $z \mapsto \sqrt{n} (X_n^\top X_n)^{-1} X_n^\top z$ die schwache Konvergenz

$$\sqrt{n} (X_n^\top X_n)^{-1} X_n^\top Z_n^{loc} \xrightarrow{\mathcal{D}} N(\beta, \tau^2(Q)B^{-1})$$

für alle $\beta \in \mathbb{R}^k$ und alle $Q \in \mathcal{G}$.

Der Beweis dieses Satzes verläuft analog zu dem von Satz 3.4 und wird hier nicht wiederholt.

5.2.3 Die Schätzer von Koul & Susarla

Wir wollen nun einen adaptiven Schätzer für den Hauptparameter in unserem lokalisierten Regressionsmodell E_n^{loc} kennenlernen. In der Literatur finden sich mehrere adaptive Schätzer für Lineare Regressionsmodelle mit nicht zufälligen Covariablen, die jeweils auf verschiedenen Ansätzen beruhen. Exemplarisch seien hier die Arbeiten Jurečková und

Sen (1984) sowie Koul und Susarla (1983) genannt, von denen wir nur auf letztere näher eingehen.

In der Arbeit Koul und Susarla (1983) wird zum einen ein Schätzer vorgestellt, der für die Klasse der um den Ursprung symmetrischen Verteilungen, die den Voraussetzungen 2.9 genügen, adaptiv ist. Zum anderen wird ein Schätzer angegeben, der für eine größere Klasse von Verteilungen asymptotisch normal, aber nicht mehr adaptiv ist.

Die Arbeit von Koul und Susarla (1983) ist allerdings in einem sehr allgemeinen Rahmen gehalten und bleibt praxisrelevante Beispiele schuldig. Wir werden die Schätzer zunächst in ihrer Originalfassung präsentieren und anschließend zeigen, dass die Voraussetzungen für die Adaptivität dieser Schätzer in unserem Modell erfüllt sind.

5.2.3.1 Die Schätzer in ihrer Originalfassung

Zuerst wollen wir festhalten, dass die in der Arbeit Koul und Susarla (1983) konstruierten Schätzer translations- und skaleninvariant sind. Zur Vereinfachung werden wir hier nur die translationsinvariante Form der Schätzer angeben.

Beobachtet werden in dieser Arbeit die Zufallsvektoren

$$\tilde{Y}_n = \tilde{X}_n \beta + \epsilon_n, \quad \beta \in \mathbb{R}^k, \quad (5.6)$$

wobei $\tilde{X}_n \in \mathbb{R}^{n \times k}$ bekannte Matrizen seien und $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nn})^\top$ die Fehlervektoren. Die ϵ_{ni} seien dabei stochastisch unabhängig und besitzen die gemeinsame Verteilung Q aus einer Klasse von Verteilungen \mathcal{G} , deren Elemente alle den Voraussetzungen 2.9 genügen. (In Koul und Susarla (1983) wird auch noch der Fall untersucht, dass die Voraussetzungen 2.9 nicht erfüllt sind. Dies ist aber für unsere Zwecke von geringerem Interesse.)

Mit \tilde{x}_{ni} , $i = 1, \dots, n$, bezeichnen wir die Zeilenvektoren der Matrizen \tilde{X}_n , also

$$\tilde{x}_{ni}^\top \in \mathbb{R}^k \quad \text{und} \quad \tilde{X}_n = (\tilde{x}_{n1}^\top, \dots, \tilde{x}_{nn}^\top)^\top,$$

und mit \tilde{Y}_{ni} die Komponenten des Zufallsvektors \tilde{Y}_n . Weiter seien in diesem Abschnitt

- ψ_r die λ -Dichte der eindimensionalen Normalverteilung $N(0, r^2)$, $r > 0$, also

$$\psi_r(x) = \frac{1}{\sqrt{2\pi r^2}} \exp\left(-\frac{x^2}{2r^2}\right), \quad x \in \mathbb{R},$$

- $g : \mathbb{R} \rightarrow \mathbb{R}$ zweimal stetig differenzierbar mit

$$g(x) = g(-x), \quad g(x) \leq g(0), \quad x \in \mathbb{R}, \quad \text{und} \quad g(x) = 0 \quad \text{für} \quad |x| \geq 1,$$

- $a_n, c_n, r_n > 0$, $n \in \mathbb{N}$, mit

$$a_n \rightarrow 0, \quad r_n \rightarrow 0 \quad \text{und} \quad c_n \rightarrow \infty.$$

In der Arbeit von Koul und Susarla (1983) wird auf die beiden folgenden Klassen von Fehlerverteilungen eingegangen:

1. Die Klasse \mathcal{G}_{sym} der *symmetrischen Fehlerverteilungen*. Damit meinen wir alle Q , die eine Lebesgue-Dichte q mit $q(x) = q(-x)$, $x \in \mathbb{R}$, besitzen und die Voraussetzungen 2.9 erfüllen mit positiver endlicher Fisher-Information $J(Q)$.
2. Die Klasse \mathcal{G}_H aller Fehlerverteilungen, die die Voraussetzungen 2.9, also die Hájek-Bedingungen, erfüllen mit positiver endlicher Fisher-Information $J(Q)$.

Die Schätzer für die Klasse \mathcal{G}_{sym}

Es seien $\bar{\beta}_n^* : \mathbb{R}^n \rightarrow \mathbb{R}^k$, $n \in \mathbb{N}$, translationsinvariante Schätzer mit

$$\left(\tilde{X}_n^\top \tilde{X}_n\right)^{\frac{1}{2}} \left(\bar{\beta}_n^* - \beta\right) = O_P(1) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}_{sym}. \quad (5.7)$$

Dabei bezeichne $O_P(1)$ einen (unter der zu β und Q gehörigen Verteilung) straffen Term. Mit diesen $\bar{\beta}_n^*$ definieren wir für jedes $n \in \mathbb{N}$ die auf der Methode der Kerndichteschätzung beruhenden Schätzer für die Dichte q von Q sowie deren Ableitung q' gemäß

$$\hat{q}_n^*(x) := \frac{1}{2n} \sum_{i=1}^n \left[\psi_{r_n} \left(x - \tilde{Y}_{ni} + \tilde{x}_{ni} \bar{\beta}_n^* \right) + \psi_{r_n} \left(-x - \tilde{Y}_{ni} + \tilde{x}_{ni} \bar{\beta}_n^* \right) \right], \quad x \in \mathbb{R},$$

$$\hat{q}_n^{*'}(x) := \frac{1}{2n} \sum_{i=1}^n \left[\psi'_{r_n} \left(x - \tilde{Y}_{ni} + \tilde{x}_{ni} \bar{\beta}_n^* \right) + \psi'_{r_n} \left(-x - \tilde{Y}_{ni} + \tilde{x}_{ni} \bar{\beta}_n^* \right) \right], \quad x \in \mathbb{R},$$

sowie die Schätzer für den Score $\frac{q'}{q}$

$$\hat{L}_n^*(x) := \frac{\hat{q}_n^{*'}(x)}{\max\{\hat{q}_n^*(x), a_n\}}, \quad x \in \mathbb{R},$$

und die Schätzer für die Fisher-Information $J(Q)$

$$\hat{J}_n^* := \int_{\mathbb{R}} \frac{(\hat{q}_n^{*'}(x))^2}{\max\{\hat{q}_n^*(x), a_n\}} g\left(\frac{x}{c_n}\right) dx. \quad (5.8)$$

Setzen wir

$$\Upsilon_n^*(x) := \left(\psi_{r_n} \left(x - \tilde{Y}_{n1} + \tilde{x}_{n1} \bar{\beta}_n^* \right), \dots, \psi_{r_n} \left(x - \tilde{Y}_{nn} + \tilde{x}_{nn} \bar{\beta}_n^* \right) \right)^\top, \quad x \in \mathbb{R},$$

und setzen wir voraus, dass die Matrizen $\tilde{X}_n^\top \tilde{X}_n$ *regulär* sind, so können wir die uns interessierenden Schätzer $\hat{\beta}_n^*$, $n \in \mathbb{N}$, definieren:

$$\hat{\beta}_n^* := \bar{\beta}_n^* - \frac{1}{\hat{J}_n^*} \int_{\mathbb{R}} \hat{L}_n^*(x) g\left(\frac{x}{c_n}\right) \left(\tilde{X}_n^\top \tilde{X}_n\right)^{-1} \tilde{X}_n^\top \Upsilon_n^*(x) dx. \quad (5.9)$$

Satz 5.15 (vgl. Koul und Susarla (1983), Theorem 4.2)

Die Matrizen $\tilde{X}_n^\top \tilde{X}_n$ seien regulär. Dann gilt unter den Voraussetzungen

$$z_n^2 := \max_{i=1, \dots, n} \tilde{x}_{ni} \left(\tilde{X}_n^\top \tilde{X}_n \right)^{-1} \tilde{x}_{ni}^\top \xrightarrow{n \rightarrow \infty} 0, \quad (5.10)$$

$$\frac{c_n \max \left\{ \frac{1}{\sqrt{n}}, z_n \right\}}{r_n^5 a_n} \xrightarrow{n \rightarrow \infty} 0, \quad (5.11)$$

$$\exists \delta \in \left(0, \frac{1}{2} \right) : \frac{a_n^\delta c_n}{r_n^{\frac{13}{4}}} \xrightarrow{n \rightarrow \infty} 0 \quad (5.12)$$

für die Schätzer $\hat{\beta}_n^*$ aus Formel (5.9)

$$\left(\tilde{X}_n^\top \tilde{X}_n \right)^{\frac{1}{2}} \left(\hat{\beta}_n^* - \beta \right) \xrightarrow{\mathcal{D}} N \left(0, \frac{1}{J(Q)} I_k \right) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}_{sym}, \quad (5.13)$$

wobei die schwache Konvergenz unter dem zu β und Q gehörigen Maß zu verstehen ist.

Die Schätzer für die Klasse \mathcal{G}_H

Es sei \tilde{x}_{ni_j} , $j = 1, \dots, k$, die j -te Koordinate des Zeilenvektors \tilde{x}_{ni} . Mit

$$m_{nj} := \frac{1}{n} \sum_{i=1}^n \tilde{x}_{ni_j}, \quad j = 1, \dots, k,$$

bezeichnen wir das Mittel der j -ten Spalte der Matrix \tilde{X}_n . Weiter seien

$$D_n := \tilde{X}_n - \begin{pmatrix} m_{n1} & \cdots & m_{nk} \\ \vdots & \ddots & \vdots \\ m_{n1} & \cdots & m_{nk} \end{pmatrix}, \quad n \in \mathbb{N}, \quad (5.14)$$

Zentrierungen von \tilde{X}_n . Die Matrizen $D_n^\top D_n$ setzen wir als *regulär* voraus, insbesondere dürfen die Modellmatrizen \tilde{X}_n keine konstanten Spalten besitzen.

Es seien $\bar{\beta}_n : \mathbb{R}^n \rightarrow \mathbb{R}^k$, $n \in \mathbb{N}$, translationsinvariante Schätzer mit

$$\left(D_n^\top D_n \right)^{\frac{1}{2}} \left(\bar{\beta}_n - \beta \right) = O_P(1) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}_H. \quad (5.15)$$

Damit definieren wir für jedes $n \in \mathbb{N}$ die Schätzer für die Dichte q von Q sowie deren Ableitung q'

$$\hat{q}_n(x) := \frac{1}{n} \sum_{i=1}^n \psi_{r_n} \left(x - \tilde{Y}_{ni} + \tilde{x}_{ni} \bar{\beta}_n \right), \quad x \in \mathbb{R},$$

$$\hat{q}'_n(x) := \frac{1}{n} \sum_{i=1}^n \psi'_{r_n} \left(x - \tilde{Y}_{ni} + \tilde{x}_{ni} \bar{\beta}_n \right), \quad x \in \mathbb{R},$$

sowie die Schätzer für den Score $\frac{q'}{q}$

$$\hat{L}_n(x) := \frac{\hat{q}'_n(x)}{\max\{\hat{q}_n(x), a_n\}}, \quad x \in \mathbb{R},$$

und die Schätzer für die Fisher-Information $J(Q)$

$$\hat{J}_n := \int_{\mathbb{R}} \frac{(\hat{q}'_n(x))^2}{\max\{\hat{q}_n(x), a_n\}} g\left(\frac{x}{c_n}\right) dx. \quad (5.16)$$

Setzen wir nun noch

$$\Upsilon_n(x) := \left(\psi_{r_n} \left(x - \tilde{Y}_{n1} + \tilde{x}_{n1} \bar{\beta}_n \right), \dots, \psi_{r_n} \left(x - \tilde{Y}_{nn} + \tilde{x}_{nn} \bar{\beta}_n \right) \right)^\top, \quad x \in \mathbb{R},$$

so können wir die uns interessierenden Schätzer $\hat{\beta}_n$, $n \in \mathbb{N}$, definieren:

$$\hat{\beta}_n := \bar{\beta}_n - \frac{1}{\hat{J}_n} \int_{\mathbb{R}} \hat{L}_n(x) g\left(\frac{x}{c_n}\right) (D_n^\top D_n)^{-1} D_n^\top \Upsilon_n(x) dx. \quad (5.17)$$

Satz 5.16 (vgl. Koul und Susarla (1983), Theorem 4.1)

Die Matrizen $D_n^\top D_n$ seien regulär. Dann gilt unter den Voraussetzungen

$$w_n^2 := \max_{i=1, \dots, n} \tilde{x}_{ni} (D_n^\top D_n)^{-1} \tilde{x}_{ni}^\top \xrightarrow{n \rightarrow \infty} 0, \quad (5.18)$$

$$\exists C > 0 \forall n \in \mathbb{N} : \text{Spur} \left[\left(\tilde{X}_n^\top \tilde{X}_n \right) (D_n^\top D_n)^{-1} \right] \leq C, \quad (5.19)$$

$$\frac{c_n \max \left\{ \frac{1}{\sqrt{n}}, w_n \right\}}{r_n^5 a_n} \xrightarrow{n \rightarrow \infty} 0, \quad (5.20)$$

$$\exists \delta \in \left(0, \frac{1}{2} \right) : \frac{a_n^\delta c_n}{r_n^{\frac{13}{4}}} \xrightarrow{n \rightarrow \infty} 0 \quad (5.21)$$

für die Schätzer $\hat{\beta}_n$ aus Formel (5.17)

$$(D_n^\top D_n)^{\frac{1}{2}} \left(\hat{\beta}_n - \beta \right) \xrightarrow{\mathcal{D}} N \left(0, \frac{1}{J(Q)} I_k \right) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}_H. \quad (5.22)$$

Wir wollen nun die Sätze 5.15 und 5.16 auf unser Modell E_n^{loc} anwenden. Dabei werden wir auch näher auf die Voraussetzungen dieser Sätze eingehen.

5.2.3.2 Die Schätzer für das Modell E_n^{loc}

Im Modell E_n^{loc} beobachten wir die Zufallsvektoren $Z_n^{loc} = \frac{1}{\sqrt{n}}X_n\beta + \epsilon_n$ mit $X_n = \Sigma_n^{-1}A_n^\top$ und A_n bzw. Σ_n^2 aus den Formeln (2.8) bzw. (2.9). Mit den Bezeichnungen aus Formel (5.6) ergibt sich

$$\tilde{Y}_n = Z_n^{loc} \quad \text{und} \quad \tilde{X}_n = \frac{1}{\sqrt{n}}X_n = \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{f_1}{\sigma}(t_{n1}) & \cdots & \frac{f_k}{\sigma}(t_{n1}) \\ \vdots & \ddots & \vdots \\ \frac{f_1}{\sigma}(t_{nn}) & \cdots & \frac{f_k}{\sigma}(t_{nn}) \end{pmatrix}.$$

Satz 5.17 ($\hat{\beta}_n^*$ ist \mathcal{G}_{sym} -adaptiv für das Lineare Regressionsmodell E_n^{loc})

Gegeben seien die lokalisierten Linearen Regressionsmodelle E_n^{loc} mit dem Nebenparameterraum \mathcal{G}_{sym} . Insbesondere sind dann für alle $Q \in \mathcal{G}_{sym}$ die Voraussetzungen 2.9 erfüllt mit endlicher positiver Fisher-Information $J(Q)$. Weiterhin sollen Designfolge, Grenzdesign, Regressoren und Varianzfunktion die Voraussetzungen von Satz 2.16 erfüllen, und B bezeichne die Matrix aus (2.26).

Es sollen die beiden Konvergenzvoraussetzungen (5.11) und (5.12) aus Satz 5.15 gelten. Dann ist die Folge der Schätzer $\hat{\beta}_n^*$, $n \in \mathbb{N}$, aus Formel (5.9) \mathcal{G}_{sym} -adaptiv im Sinn von Definition 5.11, d.h. in Anbetracht von Satz 5.13

$$\mathcal{L} \left(\hat{\beta}_n^* \left| \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}} \right. \right) \xrightarrow{\mathcal{D}} N \left(\beta, \frac{1}{J(Q)}B^{-1} \right) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}_{sym}. \quad (5.23)$$

Beweis. Nach Satz 5.13 ist nur die schwache Konvergenz (5.23) nachzuweisen. Dies werden wir natürlich mit Satz 5.15 bewerkstelligen. Dazu müssen wir zunächst Voraussetzung (5.10) dieses Satzes nachweisen. Wie wir schon mehrfach gesehen haben, gilt die elementweise Konvergenz der Matrizen

$$\tilde{X}_n^\top \tilde{X}_n = \frac{1}{n}X_n^\top X_n \xrightarrow{n \rightarrow \infty} B. \quad (5.24)$$

Die Matrix B ist positiv definit, so dass für hinreichend großes $n \in \mathbb{N}$ auch die Matrizen $\tilde{X}_n^\top \tilde{X}_n$ regulär sind. Ferner gilt mit der abkürzenden Schreibweise aus Abschnitt 5.1 wegen der Voraussetzung (2.25) aus Satz 2.16

$$z_n^2 = \max_{i=1, \dots, n} \tilde{x}_{ni} \left(\tilde{X}_n^\top \tilde{X}_n \right)^{-1} \tilde{x}_{ni}^\top = \frac{1}{n} \max_{i=1, \dots, n} \frac{f^\top}{\sigma}(t_{ni}) \underbrace{\left(\tilde{X}_n^\top \tilde{X}_n \right)^{-1}}_{\rightarrow B^{-1}} \frac{f}{\sigma}(t_{ni}) \xrightarrow{n \rightarrow \infty} 0.$$

Damit sind die Voraussetzungen von Satz 5.15 erfüllt und aus der schwachen Konvergenz (5.13) folgt mit Formel (5.24) sofort die Behauptung (5.23). \square

In Satz 5.15 wird vorausgesetzt, dass alle $\tilde{X}_n^\top \tilde{X}_n$ regulär sind. Dies ist hier erst ab

hinreichend großem n garantiert.

Nun wenden wir uns dem Regressionsmodell mit Nebenparameterraum \mathcal{G}_H zu. Für die Matrizen D_n aus Formel (5.14) gilt dann

$$D_n = \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{f_1}{\sigma}(t_{n1}) - \frac{1}{n} \sum_{i=1}^n \frac{f_1}{\sigma}(t_{ni}) & \cdots & \frac{f_k}{\sigma}(t_{n1}) - \frac{1}{n} \sum_{i=1}^n \frac{f_k}{\sigma}(t_{ni}) \\ \vdots & \ddots & \vdots \\ \frac{f_1}{\sigma}(t_{nn}) - \frac{1}{n} \sum_{i=1}^n \frac{f_1}{\sigma}(t_{ni}) & \cdots & \frac{f_k}{\sigma}(t_{nn}) - \frac{1}{n} \sum_{i=1}^n \frac{f_k}{\sigma}(t_{ni}) \end{pmatrix}.$$

Damit ergibt sich

$$\begin{aligned} D_n^\top D_n &= \frac{1}{n} \left(\sum_{\zeta=1}^n \left[\frac{f_j}{\sigma}(t_{n\zeta}) - \frac{1}{n} \sum_{i=1}^n \frac{f_j}{\sigma}(t_{ni}) \right] \left[\frac{f_l}{\sigma}(t_{n\zeta}) - \frac{1}{n} \sum_{i=1}^n \frac{f_l}{\sigma}(t_{ni}) \right] \right)_{j,l=1}^k \\ &= \left(\int_{[a,b]} \frac{f_j f_l}{\sigma^2} dH_n - \int_{[a,b]} \frac{f_j}{\sigma} dH_n \int_{[a,b]} \frac{f_l}{\sigma} dH_n \right)_{j,l=1}^k \end{aligned}$$

und unter den Voraussetzungen von Satz 2.16 folgt die elementweise Konvergenz

$$D_n^\top D_n \xrightarrow{n \rightarrow \infty} \left(\int_{[a,b]} \frac{f_j f_l}{\sigma^2} dH - \int_{[a,b]} \frac{f_j}{\sigma} dH \int_{[a,b]} \frac{f_l}{\sigma} dH \right)_{j,l=1}^k =: D. \quad (5.25)$$

Lemma 5.18 Die Matrix D aus Formel (5.25) ist positiv semidefinit.

Sind die Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ linear unabhängig in $L^2(H)$, so sind die folgenden Aussagen äquivalent.

1. D ist positiv definit.
2. Keine der Funktionen $\frac{f_j}{\sigma}$, $j = 1, \dots, k$, ist konstant in $L^2(H)$.

Beweis. Die Matrix D ist die Kovarianzmatrix von $\left(\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}\right)^\top$ unter der zu der Verteilungsfunktion H gehörigen Verteilung. Somit folgt die Aussage aus elementaren Sätzen über die Struktur von Kovarianzmatrizen. \square

Satz 5.19 ($\hat{\beta}_n$ ist asymptotisch normal für E_n^{loc} und \mathcal{G}_H)

Gegeben seien die lokalisierten Linearen Regressionsmodelle E_n^{loc} mit dem Nebenparameterraum \mathcal{G}_H . Keine der Funktionen $\frac{f_1}{\sigma}, \dots, \frac{f_k}{\sigma}$ sei konstant in $L^2(H)$. Weiterhin sollen Designfolge, Grenzdesign, Regressoren und Varianzfunktion die Voraussetzungen von Satz 2.16 erfüllen.

Es mögen die beiden Konvergenzvoraussetzungen (5.20) und (5.21) aus Satz 5.16 gelten. Dann gilt für die Folge der Schätzer $\hat{\beta}_n$, $n \in \mathbb{N}$, aus Formel (5.17)

$$\mathcal{L} \left(\hat{\beta}_n \mid \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}} \right) \xrightarrow{\mathcal{D}} N \left(\beta, \frac{1}{J(Q)} D^{-1} \right) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}_H.$$

Beweis. Wir müssen lediglich die Voraussetzungen von Satz 5.16 überprüfen, dann folgt die Behauptung mit (5.25) aus der schwachen Konvergenzaussage (5.22).

Wegen der Konvergenz $D_n^\top D_n \rightarrow D$ und wegen Lemma 5.18 sind die Matrizen $D_n^\top D_n$ ab hinreichend großem $n \in \mathbb{N}$ regulär.

Voraussetzung (5.18) folgt mittels der Konvergenzaussage (5.25) analog wie die entsprechende Aussage im Beweis von Satz 5.17.

Wegen $\tilde{X}_n^\top \tilde{X}_n \rightarrow B$ und $D_n^\top D_n \rightarrow D$ folgt auch Voraussetzung (5.19) direkt. \square

Man beachte, dass auch hier die Matrizen $D_n^\top D_n$ erst ab hinreichend großem $n \in \mathbb{N}$ regulär sind. Da im Allgemeinen $D \neq B$ ist, sind die Schätzer $\hat{\beta}_n$ nicht \mathcal{G}_H -adaptiv im Sinn von Definition 5.11.

Korollar 5.20 *Es gelten die Voraussetzungen von Satz 5.19. Weiter sei*

$$\int_{[a,b]} \frac{f_j}{\sigma} dH = 0 \quad \text{für } j = 1, \dots, k. \quad (5.26)$$

Dann sind die Schätzer $\hat{\beta}_n$ aus (5.17) \mathcal{G}_H -adaptiv.

Wir werden den Abschnitt über adaptive Schätzverfahren mit einigen Bemerkungen abschließen.

Bemerkung 5.21

1. Wir wollen in dieser Arbeit nicht auf notwendige oder hinreichende Bedingungen für die Existenz adaptiver Schätzer eingehen. In der Literatur werden jedoch meist zwei verschiedene „Adaptionsvoraussetzungen“ gemacht. Zum einen ist das die Voraussetzung symmetrischer Fehler, die bei uns in Form der Klasse \mathcal{G}_{sym} gegeben ist, zum anderen ist das eine „Zentrierungseigenschaft“ der Modellmatrizen, die bei uns durch Gleichung (5.26) gegeben ist, wobei in diesem Fall eine größere Klasse von Fehlerverteilungen \mathcal{G}_H zugelassen ist.

Da die Klasse \mathcal{G}_H Fehlerverteilungen zulässt, deren Erwartungswert nicht gleich Null ist, ist es ganz natürlich vorauszusetzen, dass keiner der Regressoren konstant sein darf, dass also kein „Intercept“ auftreten darf. Man könnte sonst nicht mehr zwischen Intercept und Erwartungswert der unbekanntenen Fehlerverteilung unterscheiden. Man vergleiche hierzu auch die Arbeit Portnoy und Koenker (1989), in der ein adaptiver Schätzer für den „slope“ konstruiert wird.

2. Die beiden Schätzer $\hat{\beta}_n^*$ und $\hat{\beta}_n$ sind jeweils noch von der Wahl der Funktion g sowie der Folgen $(a_n)_{n \in \mathbb{N}}$, $(r_n)_{n \in \mathbb{N}}$ und $(c_n)_{n \in \mathbb{N}}$ abhängig. Zu einer geeigneten Wahl dieser Größen wird in der Arbeit Koul und Susarla (1983) keine Aussage gemacht.
3. Ebenfalls zu wählen sind noch die „Ausgangsschätzer“ $\overline{\beta}_n^*$ und $\overline{\beta}_n$, mit denen die Schätzer $\hat{\beta}_n^*$ und $\hat{\beta}_n$ konstruiert werden. $\overline{\beta}_n^*$ und $\overline{\beta}_n$ sollen translationsinvariant sein und die Voraussetzungen (5.7) bzw. (5.15) erfüllen. Die Existenz solcher „Preliminary Estimators“ ist gewährleistet, man vergleiche dazu die Anmerkungen in den Arbeiten Bickel (1982) oder Koul und Susarla (1983). Setzen wir zusätzlich voraus, dass die Fehlerverteilung Erwartungswert Null und endliche Varianz besitzt, so können wir für $\overline{\beta}_n^*$ oder $\overline{\beta}_n$ auch den Gauß-Markoff-Schätzer aus Satz 5.14 verwenden.

5.3 Adaptive Testverfahren

In diesem Abschnitt werden wir den Adaptionsbegriff auf die Testtheorie ausdehnen. Ziel ist es, wie am Anfang von Kapitel 5 erwähnt, *eine* Testfolge für unser Lineares Regressionsmodell anzugeben, die für eine möglichst *große Klasse* von Fehlerverteilungen asymptotisch optimal in der Klasse aller asymptotisch α -ähnlichen Testfolgen ist, d.h. für jede Fehlerverteilung dieser Klasse dieselbe Grenzgüte besitzt wie die auf der Kenntnis der Fehlerverteilung beruhende optimale Testfolge aus Satz 3.1.

Adaptive Testverfahren wurden ausführlich in der Arbeit Choi, Hall und Schick (1996) untersucht. Dort wurde allerdings nicht der Begriff „adaptiv“ sondern „effizient“ verwendet. Wir werden hier aber in Analogie zur Schätztheorie von adaptiven Testfolgen sprechen. In der Arbeit Choi et al. (1996) liegt auch eine Lokalisierung im Nebenparameter vor, was für unsere Zwecke nicht erforderlich ist. Regressionsmodelle werden in dieser Arbeit in Beispiel 3 auf Seite 856 behandelt, dort aber unter sehr abstrakten Voraussetzungen.

Auch im Folgenden betten wir alle Definitionen in unseren Kontext schwach konvergenter Experimente ein, so dass wir eine andere Darstellung und auch eine große Allgemeinheit erhalten.

5.3.1 Definition und Konstruktion mittels adaptiver Schätzer

Zunächst wollen wir eine allgemeine Definition einer adaptiven Testfolge angeben, bevor wir diese dann für unsere Regressionsmodelle E_n^{loc} aus (5.1) präzisieren.

Im allgemeinen Fall haben wir Experimente $F_n = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,(h,\theta)} \mid (h,\theta) \in H \times \Theta\}) \in \mathcal{E}(H \times \Theta)$ betrachtet, wobei $(H, \langle \cdot, \cdot \rangle)$ ein *endlichdimensionaler* Hilbertraum war und $\Theta \neq \emptyset$ ein beliebiger Nebenparameterraum. Weiter haben wir vorausgesetzt, dass für

jedes $\theta \in \Theta$ die Experimente $F_n(\theta) = (\mathcal{X}_n, \mathcal{A}_n, \{P_{n,(h,\theta)} \mid h \in H\}) \in \mathcal{E}(H)$ schwach gegen ein Gaußexperiment konvergieren, genauer

$$F_n(\theta) \xrightarrow{w} \mathcal{N}(I_H, C(\theta))$$

mit positiv definiten Covarianzen $C(\theta)$. Nach Bemerkung 5.9 sind diese Gaußexperimente äquivalent zu den Standardgaußshifts auf $(H, \langle \cdot, \cdot \rangle_{C(\theta)^{-1}})$ mit $\langle h_1, h_2 \rangle_{C(\theta)^{-1}} := \langle h_1, C(\theta)^{-1}h_2 \rangle$, $h_1, h_2 \in H$.

Wir untersuchen wie in Kapitel 3 die „eindimensionalen“ Testprobleme

$$H_\ell = \{h \in H \mid \langle h, \ell \rangle \leq 0\}, \quad K_\ell = \{h \in H \mid \langle h, \ell \rangle > 0\} \quad (5.27)$$

mit einem *fest vorgegebenen* $\ell \in H \setminus \{0\}$.

Halten wir zunächst einmal ein $\theta \in \Theta$ fest und definieren $\gamma(\theta) := C(\theta)\ell$, so ist

$$\langle h, \gamma(\theta) \rangle_{C(\theta)^{-1}} = \langle h, \ell \rangle, \quad \|\gamma(\theta)\|_{C(\theta)^{-1}}^2 = \langle \ell, C(\theta)\ell \rangle,$$

und nach Satz 1.49 existiert eine von $\theta \in \Theta$ abhängige Testfolge $\varphi_{n,\theta}^*$, die im Sinne von Definition 1.41 asymptotisch optimal ist in der Klasse aller asymptotisch α -ähnlichen Testfolgen und die die Grenzgüte

$$\lim_{n \rightarrow \infty} g_{\varphi_{n,\theta}^*}(h) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\langle h, \ell \rangle}{\langle \ell, C(\theta)\ell \rangle} \right), \quad h \in H,$$

besitzt. Man beachte, dass wegen der endlichen Dimension von H nach Bemerkung 5.10 immer eine zentrale Folge existiert, was in Satz 1.49 vorausgesetzt war.

Diese Überlegungen führen uns zu der folgenden

Definition 5.22 (Adaptive Testfolge (zum Niveau α für (H_ℓ, K_ℓ))

Die Experimente F_n , $n \in \mathbb{N}$, sollen die gerade beschriebenen Voraussetzungen erfüllen. Wir betrachten das Testproblem (H_ℓ, K_ℓ) mit $\ell \in H \setminus \{0\}$ aus Formel (5.27). Eine Folge von Tests $\varphi_n : \mathcal{X}_n \rightarrow [0, 1]$, $n \in \mathbb{N}$, heißt **Θ -adaptiv (zum Niveau $\alpha \in (0, 1)$ für das Testproblem (H_ℓ, K_ℓ))**, falls für die Gütefunktionen g_{φ_n} gilt

$$\lim_{n \rightarrow \infty} g_{\varphi_n}(h) = \Phi \left(\Phi^{-1}(\alpha) + \frac{\langle h, \ell \rangle}{\langle \ell, C(\theta)\ell \rangle} \right) \quad \text{für alle } h \in H \text{ und alle } \theta \in \Theta.$$

Bemerkung 5.23

1. Eine Θ -adaptive Testfolge ist für *jedes* feste $\theta \in \Theta$ asymptotisch optimal in der Klasse aller α -ähnlichen Testfolgen für das Problem (H_ℓ, K_ℓ) im Sinne von Definition 1.41. Wäre der Parameter θ bekannt, so könnte man keinen asymptotisch besseren Test angeben. Wir verbinden also auch hier den „Adaptionsbegriff“ mit einer Optimalitätseigenschaft.
2. Da schwache Konvergenz von Experimenten meistens eine Lokalisierung im Parameterraum impliziert, ist es auch in diesem Fall wieder naheliegend, von **lokal adaptiven Testfolgen** zu sprechen.

Nun wollen wir angeben, was Definition 5.22 für unsere Regressionsmodelle bedeutet. Der folgende Satz ist eine direkte Folgerung aus dieser Definition, man vergleiche auch den Beweis von Satz 5.13.

Satz 5.24 (Adaptive Testfolge für die Modelle E_n^{loc})

Gegeben sei eine Menge \mathcal{G} von Fehlerverteilungen Q , die allesamt den Voraussetzungen 2.9 genügen sollen mit endlicher positiver Fisher-Information $J(Q)$. Weiterhin sollen Designfolge, Grenzdesign, Regressoren und Varianzfunktion die Voraussetzungen von Satz 2.16 erfüllen, und B bezeichne die Matrix aus (2.26). Wir betrachten die lokalisierten Experimente E_n^{loc} aus Formel (5.1) und die Testprobleme

$$H_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta \leq 0\}, \quad K_\ell = \{\beta \in \mathbb{R}^k \mid \ell^\top \beta > 0\}, \quad (5.28)$$

mit $\ell \in \mathbb{R}^k \setminus \{0\}$.

Eine Folge von Tests $\varphi_n : \mathbb{R}^n \rightarrow [0, 1]$, $n \in \mathbb{N}$, ist genau dann \mathcal{G} -adaptiv zum Niveau $\alpha \in (0, 1)$ für das Problem (H_ℓ, K_ℓ) , wenn gilt

$$\lim_{n \rightarrow \infty} g_{\varphi_n}(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \sqrt{J(Q)} \frac{\ell^\top \beta}{\sqrt{\ell^\top B^{-1} \ell}} \right) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}.$$

Nun liegt natürlich die Frage nahe, ob man mittels lokal adaptiver Schätzfolgen auch adaptive Testfolgen konstruieren kann. Die Antwort ist sehr einfach. Wir beschränken uns dabei auf die Betrachtung der Linearen Regressionsmodelle E_n^{loc} .

Satz 5.25 (Konstruktion adaptiver Tests aus adaptiven Schätzern)

Gegeben seien die Experimente E_n^{loc} und das Testproblem (H_ℓ, K_ℓ) aus Formel (5.28). Alle Voraussetzungen von Satz 5.24 seien erfüllt.

Weiter seien eine im Sinn von Satz 5.13 \mathcal{G} -adaptive Folge von Schätzern $T_n : \mathbb{R}^n \rightarrow \mathbb{R}^k$, $n \in \mathbb{N}$, gegeben, sowie eine Folge $J_n : \mathbb{R}^n \rightarrow \mathbb{R}$, $n \in \mathbb{N}$, von konsistenten Schätzern für $J(Q)$, d.h.

$$\lim_{n \rightarrow \infty} \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}} (|J_n - J(Q)| > \epsilon) = 0 \quad \text{für alle } \epsilon > 0, \beta \in \mathbb{R}^k \text{ und } Q \in \mathcal{G}. \quad (5.29)$$

Dann ist die Folge von Tests $\varphi_n : \mathbb{R}^n \rightarrow [0, 1]$,

$$\varphi_n(x) := \begin{cases} 1, & \sqrt{\frac{J_n(x)}{\ell^\top B^{-1} \ell}} \ell^\top T_n(x) > \Phi^{-1}(1 - \alpha) \\ 0, & \sqrt{\frac{J_n(x)}{\ell^\top B^{-1} \ell}} \ell^\top T_n(x) < \Phi^{-1}(1 - \alpha) \end{cases}, \quad x \in \mathbb{R}^n, n \in \mathbb{N},$$

\mathcal{G} -adaptiv.

Beweis. Aus (5.5) und (5.29) folgt die schwache Konvergenz

$$\mathcal{L} \left(\sqrt{\frac{J_n}{\ell^\top B^{-1} \ell}} \ell^\top T_n \mid \bigotimes_{i=1}^n Q_{\frac{f^\top(t_{ni})\beta}{\sqrt{n}\sigma(t_{ni})}} \right) \xrightarrow{\mathcal{D}} N \left(\sqrt{J(Q)} \frac{\ell^\top \beta}{\sqrt{\ell^\top B^{-1} \ell}}, 1 \right)$$

für alle $\beta \in \mathbb{R}^k$ und $Q \in \mathcal{G}$ und damit wie im Beweis von Satz 3.5 die Behauptung. \square

5.3.2 Tests mit den Schätzern von Koul & Susarla

Wir wollen nun die auf den Schätzern $\hat{\beta}_n^*$ aus (5.9) und $\hat{\beta}_n$ aus (5.17) basierenden Testfolgen betrachten. Man beachte dabei, dass nach Theorem 5.1 aus Koul und Susarla (1983) die Schätzer \hat{J}_n^* aus (5.8) unter den Voraussetzungen von Satz 5.15 konsistent im Sinne von (5.29) für $J(Q)$, $Q \in \mathcal{G}_{sym}$, sind. Ebenso sind die Schätzer \hat{J}_n aus (5.16) unter den Voraussetzungen von Satz 5.16 konsistent im Sinne von (5.29) für $J(Q)$, $Q \in \mathcal{G}_H$. Der folgende Satz ist eine direkte Folge der Sätze 5.17 und 5.25.

Satz 5.26 (\mathcal{G}_{sym} -adaptive Tests für E_n^{loc} basierend auf $\hat{\beta}_n^*$ und \hat{J}_n^*)

Gegeben seien die lokalisierten Linearen Regressionsmodelle E_n^{loc} mit dem Nebenparameterraum \mathcal{G}_{sym} . Weiterhin sollen Designfolge, Grenzdesign, Regressoren und Varianzfunktion die Voraussetzungen von Satz 2.16 erfüllen, und B bezeichne die Matrix aus (2.26). Es sollen die beiden Konvergenzvoraussetzungen (5.11) und (5.12) aus Satz 5.15 gelten. $\hat{\beta}_n^*$ bzw. \hat{J}_n^* seien die Schätzer aus (5.9) bzw. (5.8).

Dann ist die Folge von Tests $\hat{\varphi}_n^* : \mathbb{R}^n \rightarrow [0, 1]$,

$$\hat{\varphi}_n^*(x) := \begin{cases} 1, & \sqrt{\frac{\hat{J}_n^*(x)}{\ell^\top B^{-1} \ell}} \ell^\top \hat{\beta}_n^*(x) > \Phi^{-1}(1 - \alpha) \\ 0, & \sqrt{\frac{\hat{J}_n^*(x)}{\ell^\top B^{-1} \ell}} \ell^\top \hat{\beta}_n^*(x) < \Phi^{-1}(1 - \alpha) \end{cases}, \quad x \in \mathbb{R}^n, n \in \mathbb{N}, \quad (5.30)$$

\mathcal{G}_{sym} -adaptiv, also

$$\lim_{n \rightarrow \infty} g_{\hat{\varphi}_n^*}(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \sqrt{J(Q)} \frac{\ell^\top \beta}{\sqrt{\ell^\top B^{-1} \ell}} \right) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}_{sym}.$$

Den folgenden Satz erhält man analog zu Satz 5.25 aus Satz 5.19.

Satz 5.27 (Auf $\hat{\beta}_n$ und \hat{J}_n basierende Testfolge für E_n^{loc})

Gegeben seien die lokalisierten Linearen Regressionsmodelle E_n^{loc} mit dem Nebenparameterraum \mathcal{G}_H . Keine der Funktionen f_1, \dots, f_k sei konstant in $L^2(H)$. Weiterhin sollen Designfolge, Grenzdesign, Regressoren und Varianzfunktion die Voraussetzungen von Satz 2.16 erfüllen.

Es mögen die beiden Konvergenzvoraussetzungen (5.20) und (5.21) aus Satz 5.16 gelten. $\hat{\beta}_n$ bzw. \hat{J}_n seien die Schätzer aus (5.17) bzw. (5.16), D sei die Matrix aus (5.25).

Dann hat die Folge von Tests $\hat{\varphi}_n : \mathbb{R}^n \rightarrow [0, 1]$,

$$\hat{\varphi}_n(x) := \begin{cases} 1, & \sqrt{\frac{\hat{J}_n(x)}{\ell^\top D^{-1} \ell}} \ell^\top \hat{\beta}_n(x) > \Phi^{-1}(1 - \alpha) \\ 0, & \sqrt{\frac{\hat{J}_n(x)}{\ell^\top D^{-1} \ell}} \ell^\top \hat{\beta}_n(x) < \Phi^{-1}(1 - \alpha) \end{cases}, \quad x \in \mathbb{R}^n, n \in \mathbb{N}, \quad (5.31)$$

die Grenzgüte

$$\lim_{n \rightarrow \infty} g_{\hat{\varphi}_n}(\beta) = \Phi \left(\Phi^{-1}(\alpha) + \sqrt{J(Q)} \frac{\ell^\top \beta}{\sqrt{\ell^\top D^{-1} \ell}} \right) \quad \text{für alle } \beta \in \mathbb{R}^k \text{ und alle } Q \in \mathcal{G}_H.$$

Diese Tests sind im Allgemeinen nicht \mathcal{G}_H -adaptiv im Sinn von Definition 5.22. Hält man ein $Q \in \mathcal{G}_H$ fest und betrachtet dazu den optimalen Test φ_n^* aus Satz 3.1, so ergibt sich die von Q unabhängige Asymptotisch Relative Effizienz

$$ARE(\varphi_n^* : \hat{\varphi}_n) = \frac{\ell^\top D^{-1} \ell}{\ell^\top B^{-1} \ell} \geq 1.$$

Die Tests $\hat{\varphi}_n$ sind also genau für die Probleme (H_ℓ, K_ℓ) \mathcal{G}_H -adaptiv, für die gilt

$$\ell^\top B^{-1} \ell = \ell^\top D^{-1} \ell. \tag{5.32}$$

Bemerkung 5.28 Die Menge der ℓ , die Bedingung (5.32) erfüllen, bildet eine Hyper-ebene des \mathbb{R}^k .

Beispiel 5.29 Gegeben sei der Versuchsbereich $\mathcal{V} = [0, 1]$. H sei die Verteilungsfunktion des auf \mathcal{V} eingeschränkten Lebesgue-Maßes. Die Designfolge H_n sei H -äquidistant, die Varianzfunktion $\sigma^2 \equiv 1$. Weiter seien Regressoren $f_1(x) = x$ und $f_2(x) = x^2$, $x \in \mathcal{V}$, gegeben. Damit ergeben sich

$$B = \begin{pmatrix} \frac{1}{3} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{5} \end{pmatrix}, \quad B^{-1} = \begin{pmatrix} 48 & -60 \\ -60 & 80 \end{pmatrix}, \quad D = \begin{pmatrix} \frac{1}{12} & \frac{1}{12} \\ \frac{1}{12} & \frac{4}{45} \end{pmatrix}, \quad D^{-1} = \begin{pmatrix} 192 & -180 \\ -180 & 180 \end{pmatrix}.$$

Es ist

$$\ell^\top B^{-1} \ell = \ell^\top D^{-1} \ell \iff \ell \in \left[\begin{pmatrix} 5 \\ 6 \end{pmatrix} \right],$$

und damit ist die Testfolge $\hat{\varphi}_n$ unter den Voraussetzungen von Satz 5.27 genau dann \mathcal{G}_H -adaptiv, wenn ℓ ein Vielfaches des Vektors $(5, 6)^\top$ ist.

Zum Abschluss dieses Abschnitts wollen wir noch die Testfolge $\hat{\varphi}_n$ mit der Folge der Likelihood-Quotienten-Tests φ_n^{LQ} aus (3.8) vergleichen. Da diese jedoch die Varianz $\tau^2(Q)$ der Fehlerverteilung als bekannt voraussetzen, werden wir gemäß Korollar 3.6 diese Varianz durch einen konsistenten Schätzer τ_n^2 ersetzen. Dann folgt aus Satz 3.5 und Satz 5.27 sofort

Satz 5.30 (Vergleich von $\hat{\varphi}_n$ und φ_n^{LQ})

Es gelten die Voraussetzungen von Satz 5.27. Wir schränken aber die Klasse zulässiger Fehlerverteilungen weiter ein. Es sei $\mathcal{G} := \{Q \in \mathcal{G}_H \mid \int x Q(dx) = 0, \int x^2 Q(dx) =: \tau^2(Q) < \infty\}$. $\hat{\varphi}_n$ sei die Testfolge aus (5.31) und φ_n^{LQ} sei die Folge der LQ-Tests mit konsistentem Varianzschätzer τ_n^2 . Dann gilt für festes $Q \in \mathcal{G}$:

$$ARE(\varphi_n^{LQ} : \hat{\varphi}_n) = \frac{\ell^\top D^{-1} \ell}{J(Q) \tau^2(Q) \ell^\top B^{-1} \ell}.$$

Diese Pitman-Effizienz ist abhängig von der unbekanntenen Fehlerverteilung Q und kann alle Werte im Intervall $(0, \infty)$ annehmen. Somit ist im Allgemeinen keiner der Tests asymptotisch „besser“ als der andere.

Beispiel 5.31 (Fortsetzung von Beispiel 5.29)

Wir betrachten das Testproblem (H_ℓ, K_ℓ) mit $\ell = (1, 0)^\top$. Dann ist für festes $Q \in \mathcal{G}$:

$$\text{ARE}(\varphi_n^{LQ} : \hat{\varphi}_n) = \frac{4}{J(Q)\tau^2(Q)}.$$

Betrachtet man als Fehlerverteilung die zweiseitige Exponentialverteilung Q_r mit Parameter $r \in \mathbb{N}$ aus Beispiel 2.10, so erhält man für $r \geq 13$ die Ungleichung $J(Q_r)\tau^2(Q_r) > 4$ und für $r < 13$ die Ungleichung $J(Q_r)\tau^2(Q_r) < 4$. Damit ist für $r \geq 13$ der Test $\hat{\varphi}_n$ besser, für $r < 13$ ist der LQ-Test besser.

5.4 Simulationen

Den Abschluss dieses Kapitels bilden einige Simulationen, die dem Leser eine grobe Vorstellung von der Qualität einiger der in der vorliegenden Arbeit untersuchten Tests bei festen Beobachtungsumfängen n geben sollen, die aber keinerlei Anspruch auf Vollständigkeit erheben. In dieser Studie wurden die LQ-Tests φ_n^{LQ} aus (3.3) auf Seite 76, die asymptotisch optimalen Tests φ_n^* aus Satz 3.1 sowie die adaptiven Tests $\hat{\varphi}_n^*$ aus (5.30) miteinander verglichen.

Der Simulation lag die folgende Situation zu Grunde:

- Versuchsbereich $\mathcal{V} = [0, 1]$,
- Varianzfunktion $\sigma^2 \equiv 1$,
- Regressoren $f_1 \equiv 1$, $f_2 = id_{[0,1]}$,
- Versuchspunkte $t_{ni} = \frac{i}{n}$, $i = 1, \dots, n$, $n \in \mathbb{N}$.

Es handelt sich also um eine Geradenregression

$$Z_n^{loc} = \frac{1}{\sqrt{n}} X_n \beta + \epsilon_n, \quad \beta \in \mathbb{R}^2,$$

mit

$$X_n = \begin{pmatrix} 1 & \frac{1}{n} \\ \vdots & \vdots \\ 1 & \frac{n}{n} \end{pmatrix},$$

wobei, ganz im Kontext der vorliegenden Arbeit, im Parameterraum mit der Rate $\frac{1}{\sqrt{n}}$ lokalisiert wurde. Die folgenden Fehlerverteilungen Q wurden simuliert:

- Eine $N(0, 1)$ -Verteilung mit Fisher-Information $J(Q) = 1$ und Varianz $\tau^2(Q) = 1$,

- eine Laplace-Verteilung, vgl. Beispiel 3.11, mit Fisher-Information $J(Q) = 1$ und Varianz $\tau^2(Q) = 2$ und
- eine Binomialverteilung auf der Menge $\{-2, -1, 0, 1, 2\}$ mit Trefferwahrscheinlichkeit $p = \frac{1}{2}$ und Varianz $\tau^2(Q) = 1$, die den Voraussetzungen 2.9 nicht genügt und deren Fisher-Information nicht existiert.

Das der Simulation zu Grunde liegende Testproblem ist durch die Linearform $\ell = (0, 1)^\top$ in der Form

$$H_\ell = \{\beta = (\beta_1, \beta_2)^\top \in \mathbb{R}^2 \mid \ell^\top \beta = \beta_2 \leq 0\}, \quad K_\ell = \{\beta = (\beta_1, \beta_2)^\top \in \mathbb{R}^2 \mid \ell^\top \beta = \beta_2 > 0\}$$

gegeben, wir interessieren uns also dafür, ob die Steigung der Geraden negativ ist. Als Test-Niveau wurde $\alpha = 0.05$ gewählt.

Die Simulationsergebnisse beruhen auf 10000 Wiederholungen und wurden für die Beobachtungsumfänge $n = 30$, $n = 50$ und $n = 100$ durchgeführt. Als Parameterwerte wurden für β_2 die Werte 0, 1, 5 und 10 gewählt und stets für β_1 der Wert 0.

Bevor wir die Simulationsergebnisse wiedergeben, müssen wir noch auf die Problematik eingehen, die mit den adaptiven Tests $\hat{\varphi}_n^*$ verbunden ist. In der Teststatistik von $\hat{\varphi}_n^*$ tritt der Schätzer $\hat{\beta}_n^*$ aus (5.9) auf, bei dem noch die Größen r_n, z_n, a_n sowie die Funktion g „eingestellt“ werden müssen. Die kritische Größe dabei ist die „Bandbreite“ r_n , die einen sehr starken Einfluss auf den Schätzer hat. Wir haben für r_n jeweils auf Vorstudien basierende Werte genommen, die in den einzelnen Tabellen mit aufgeführt sind. Ferner wurde $a_n = 0.01$ gesetzt, und g bzw. z_n wurden implizit festgelegt, indem wir die in den Schätzern auftretenden Integrationen jeweils im Intervall $[-20, 20]$ durchgeführt haben. An dieser Stelle sei auf die Arbeit Hsieh und Manski (1987) hingewiesen, in der adaptive Schätzer aus Bickel (1982) simuliert wurden. In der dortigen Simulationsstudie wurde genauer auf die Problematik der „einzustellenden Parameter“ eingegangen. Wir wollen auch noch erwähnen, dass die adaptiven Tests $\hat{\varphi}_n^*$ unserer Simulation nicht nur aufwendig zu implementieren waren, sondern auf Grund der auftretenden Integrationen sehr viel Rechenzeit benötigten, was die geringen Beobachtungsumfänge sowie die geringe Anzahl von Wiederholungen erklärt.

Die folgenden Tabellen geben unsere Simulationsergebnisse wieder, wobei in den Spalten mit der Überschrift $n = \infty$ die jeweiligen Grenzgüten der verschiedenen Tests angeführt sind.

Normalverteilte Fehler

Sind die Fehler $N(0, 1)$ -verteilt, so sind alle drei betrachteten Testfolgen äquivalent, d.h. sie besitzen dieselbe Grenzgüte. Bei kleinen Beobachtungsumfängen ergeben sich jedoch große Unterschiede, wie Tabelle 5.1 verdeutlicht. Während der LQ-Test sowie der asymptotisch optimale Test für alle drei Beobachtungsumfänge dem Wert der Grenzgüte

	$n = 30, r_n = 0.25$			$n = 50, r_n = 0.2$			$n = 100, r_n = 0.16$			$n = \infty$
β_2	φ_n^{LQ}	φ_n^*	$\hat{\varphi}_n^*$	φ_n^{LQ}	φ_n^*	$\hat{\varphi}_n^*$	φ_n^{LQ}	φ_n^*	$\hat{\varphi}_n^*$	alle
0	.0524	.0527	.0587	.0506	.0502	.0693	.0488	.0487	.0725	.05
1	.0855	.0901	.0876	.0842	.0879	.1108	.0842	.0860	.1068	.088
5	.4216	.4790	.3339	.4182	.4499	.3952	.4167	.4357	.4075	.421
10	.8989	.9409	.7547	.8900	.9222	.8134	.8883	.9027	.8221	.893

Tabelle 5.1: Simulationsergebnisse für normalverteilte Fehler

bereits sehr nahe kommt, ist der adaptive Test nicht einmal in der Lage, das vorgegebene Niveau von 5% einzuhalten. Die zum Teil höheren Güterwerte für die Parameter $\beta_2 = 1$, $\beta_2 = 5$ und $\beta_2 = 10$ sind damit nicht mit den Güterwerten der anderen beiden Tests vergleichbar.

Laplace-verteilte Fehler

Sind die Fehler Laplace-verteilt, so besitzen der asymptotisch optimale Test und der adaptive Test eine bessere Grenzgüte als der LQ-Test. Dies kommt bei den simulierten Stichprobenumfängen jedoch nicht zum tragen, vgl. Tabelle 5.2. Auch hier zeigt sich,

	$n = 30, r_n = 0.27$			$n = 50, r_n = 0.22$			$n = 100, r_n = 0.19$			$n = \infty$	
β_2	φ_n^{LQ}	φ_n^*	$\hat{\varphi}_n^*$	φ_n^{LQ}	φ_n^*	$\hat{\varphi}_n^*$	φ_n^{LQ}	φ_n^*	$\hat{\varphi}_n^*$	φ_n^{LQ}	$\varphi_n^*, \hat{\varphi}_n^*$
0	.0509	.0510	.1242	.0499	.0528	.1389	.0511	.0494	.1181	.05	.05
1	.0782	.0873	.1614	.0699	.0850	.1865	.0711	.0833	.1620	.075	.088
5	.2692	.2562	.4202	.2779	.2909	.4641	.2612	.3112	.4032	.266	.421
10	.6588	.3784	.7653	.6557	.4801	.7955	.6643	.5985	.7634	.654	.893

Tabelle 5.2: Simulationsergebnisse für Laplace-verteilte Fehler

dass der adaptive Test das vorgegebene Niveau nicht einhalten kann. Betrachtet man den Parameter $\beta = 10$, so erkennt man, dass der LQ-Test eine deutlich größere Güte aufweist als der asymptotisch optimale Test, was andeutet, dass dessen Asymptotik erst für große Beobachtungsumfänge n greift. Um eine Vorstellung von der Asymptotik von φ_n^* zu bekommen, wurden weitere Simulationen für einen Beobachtungsumfang von $n = 1000$ durchgeführt und in Tabelle 5.3 dargestellt. Dabei erkennen wir, dass selbst

	$n = 1000$		$n = \infty$	
β_2	φ_n^{LQ}	φ_n^*	φ_n^{LQ}	φ_n^*
5	.2689	.3832	.266	.421
10	.6524	.8038	.654	.893

Tabelle 5.3: Weitere Simulationsergebnisse für Laplace-verteilte Fehler

dann der Test φ_n^* seine Grenzgüte noch nicht „erreicht“. Allerdings tritt deutlich seine

größere Güte gegenüber dem LQ-Test zu Tage.

Binomialverteilte Fehler

Zum Abschluss unserer Simulation haben wir eine Binomialverteilung, also eine diskrete Fehlerverteilung betrachtet, so dass der Test φ_n^* , der ja im Wesentlichen auf den Lebesgue-Dichten der Fehlerverteilung beruht, nicht existiert. Auch für den adaptiven Test $\hat{\varphi}_n^*$ haben wir bisher die Voraussetzungen 2.9 an die Fehlerverteilung gestellt. Zu Beginn von Abschnitt 5.2.3.1 wurde aber schon erwähnt, dass der Schätzer $\hat{\beta}_n^*$, auf dem der Test $\hat{\varphi}_n^*$ beruht, auch für z.B. binomialverteilte Fehler anwendbar ist, wenn man für die Fisher-Information $J(Q) = \infty$ setzt. Aus dieser Sicht sind die Simulationsergebnisse aus Tabell 5.4 nicht verwunderlich. Der adaptive Test ist für $\beta_2 = 5$ und $\beta_2 = 10$

	$n = 30, r_n = 0.25$		$n = 50, r_n = 0.2$		$n = 100, r_n = 0.15$		$n = \infty$
β_2	φ_n^{LQ}	$\hat{\varphi}_n^*$	φ_n^{LQ}	$\hat{\varphi}_n^*$	φ_n^{LQ}	$\hat{\varphi}_n^*$	φ_n^{LQ}
0	.0513	.0421	.0499	.0266	.0487	.0079	.05
1	.0875	.0547	.0876	.0410	.0858	.0262	.088
5	.4143	.6189	.4218	.7507	.4266	.8211	.421
10	.8916	.8449	.8948	.9159	.8935	.9492	.893

Tabelle 5.4: Simulationsergebnisse für binomialverteilte Fehler

deutlich besser als der LQ-Test, vor allem auf Grund der Tatsache, dass er sogar ein deutlich geringeres Niveau einhält als die vorgegebenen 5%. Betrachtet man die Grenzwertaussage (5.13) für den Schätzer $\hat{\beta}_n^*$ aus Satz 5.15, so erkennt man, dass $\hat{\beta}_n^*$ wegen $J(Q) = \infty$ trotz der Lokalisierungsrate $\frac{1}{\sqrt{n}}$ ein konsistenter Schätzer für β ist, wohingegen der Gauß-Markoff-Schätzer, auf dem der LQ-Test beruht, bei dieser Rate lediglich asymptotisch normalverteilt ist mit positiver Varianz.

Anhang A

Technische Hilfsmittel

A.1 Variationsabstand, Hellingerabstand, Affinität und Hellingertransformierte

Seien (Ω, \mathcal{A}) ein Messraum und $\mathcal{M}(\Omega)$ der Vektorraum aller endlichen signierten Maße auf (Ω, \mathcal{A}) . Ein signiertes Maß $\mu \in \mathcal{M}(\Omega)$ lässt sich eindeutig in zwei zueinander singuläre endliche Maße zerlegen: $\mu = \mu^+ - \mu^-$. Durch

$$\|\mu\|_{TV} := \mu^+(\Omega) + \mu^-(\Omega)$$

wird eine Norm, die **Totalvariationsnorm** auf $\mathcal{M}(\Omega)$ definiert.

Der Raum $(\mathcal{M}(\Omega), \|\cdot\|_{TV})$ ist ein Banach-Verband, vgl. Elstrodt (1996), S. 273.

Definition A.1 (Variationsnorm) Auf $\mathcal{M}(\Omega)$ wird mittels

$$\|\mu\|_V := \sup\{|\mu(A)| : A \in \mathcal{A}\}$$

eine Norm, die **Variationsnorm** definiert.

Offensichtlich sind die beiden Normen äquivalent, denn für ein signiertes Maß μ gilt die Abschätzung

$$\|\mu\|_V \leq \|\mu\|_{TV} \leq 2\|\mu\|_V.$$

Somit ist auch der Raum $(\mathcal{M}(\Omega), \|\cdot\|_V)$ ein Banach-Verband.

Definition A.2 (Hellingerabstand) Seien $P, Q \in \mathcal{M}(\Omega)$ Wahrscheinlichkeitsmaße und $\nu \in \mathcal{M}(\Omega)$ ein beliebiges Maß mit $P \ll \nu$ und $Q \ll \nu$ (zum Beispiel $\nu = P + Q$). Dann heißt

$$d_H(P, Q) := \frac{1}{\sqrt{2}} \left(\int_{\Omega} \left(\sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right)^2 d\nu \right)^{\frac{1}{2}} = \frac{1}{\sqrt{2}} \left\| \sqrt{\frac{dP}{d\nu}} - \sqrt{\frac{dQ}{d\nu}} \right\|_{L^2(\nu)}$$

Hellingerabstand von P und Q . Dabei bezeichnet $\|\cdot\|_{L^2(\nu)}$ die kanonische Norm des $L_2(\Omega, \mathcal{A}, \nu)$.

Bemerkung A.3 Der Hellingerabstand hängt nicht vom dominierenden Maß ν ab. Auf dem Raum aller Wahrscheinlichkeitsmaße auf (Ω, \mathcal{A}) wird durch d_H eine Metrik mit

$$0 \leq d_H(P, Q) \leq 1 \quad (\text{A.1})$$

definiert.

Definition A.4 (Affinität) Für Wahrscheinlichkeitsmaße P und Q wird durch

$$a(P, Q) := \int_{\Omega} \sqrt{\frac{dP}{d\nu}} \sqrt{\frac{dQ}{d\nu}} d\nu = \left\langle \sqrt{\frac{dP}{d\nu}}, \sqrt{\frac{dQ}{d\nu}} \right\rangle_{L^2(\nu)} \quad (\text{A.2})$$

die **Affinität** von P und Q definiert.

Definition A.5 (Hellingertransformierte)

Seien $\Theta \neq \emptyset$ ein Parameterraum und $E = (\mathcal{X}, \mathcal{A}, \{P_{\theta} : \theta \in \Theta\})$ ein Experiment. Sei

$$A(\Theta) := \{\alpha \subseteq \Theta \mid |\alpha| < \infty\}$$

das System aller endlichen Teilmengen von Θ . Für beliebiges $\alpha \in A(\Theta)$ sei

$$S_{\alpha} := \left\{ z \in \mathbb{R}^{|\alpha|} \mid 0 \leq z_{\theta} \leq 1, \theta \in \alpha, \sum_{\theta \in \alpha} z_{\theta} = 1 \right\}$$

der **Standard-Simplex** des $\mathbb{R}^{|\alpha|}$. Für beliebiges $\alpha \in A(\Theta)$ sei

$$E_{\alpha} := (\mathcal{X}, \mathcal{A}, \{P_{\theta} : \theta \in \alpha\})$$

die Einschränkung von E auf den Parameterraum α .

Für $\alpha \in A(\Theta)$ sei ν ein Maß auf $(\mathcal{X}, \mathcal{A})$, das die Menge $\{P_{\theta} : \theta \in \alpha\}$ dominiert. Dann heißt

$$H(E_{\alpha}) : S_{\alpha} \rightarrow [0, 1], \quad H(E_{\alpha})(z) := \int \prod_{\theta \in \alpha} \left(\frac{dP_{\theta}}{d\nu} \right)^{z_{\theta}} d\nu, \quad z = (z_{\theta})_{\theta \in \alpha} \in S_{\alpha},$$

die **Hellingertransformierte** von E_{α} , vgl. Strasser (1985), S. 271.

Die Hellingertransformierte hängt nicht von dem dominierenden Maß ν ab.

Lemma A.6 (Satz von Scheffé) Sind $P, Q \in \mathcal{M}(\Omega)$ Wahrscheinlichkeitsmaße und werden sie dominiert von einem σ -endlichen Maß ν (zum Beispiel $\nu = P + Q$), so gilt für den Variationsabstand der Maße P und Q

$$\|P - Q\|_V = \frac{1}{2} \int_{\Omega} \left| \frac{dP}{d\nu} - \frac{dQ}{d\nu} \right| d\nu = \frac{1}{2} \left\| \frac{dP}{d\nu} - \frac{dQ}{d\nu} \right\|_{L^1(\nu)}.$$

Beweis. Strasser (1985), S. 7.

Eine direkte Folgerung dieses Lemmas sowie der Substitutionsregel ist das folgende

Lemma A.7 *Seien $\Sigma_1, \Sigma_2 \in \mathbb{R}^{k \times k}$ positiv definite Matrizen. Dann gilt für die folgenden Normalverteilungen*

$$\|N(0, c\Sigma_1) - N(0, c\Sigma_2)\|_V = \|N(0, \Sigma_1) - N(0, \Sigma_2)\|_V \quad \text{für alle } c > 0.$$

Lemma A.8 (Hellingerabstand und Variationsabstand) *Es gilt die Ungleichung*

$$d_H^2(P, Q) \leq \|P - Q\|_V \leq \sqrt{d_H^2(P, Q)(2 - d_H^2(P, Q))}. \quad (\text{A.3})$$

Beweis. Strasser (1985), S. 11.

Lemma A.9 (Hellingerabstand und Affinität) *Es gilt die Beziehung*

$$d_H^2(P, Q) = 1 - a(P, Q). \quad (\text{A.4})$$

Beweis. Offensichtlich.

Lemma A.10 (Affinität, Hellingerabstand und Produktmaße)

Seien $P_i, Q_i, i = 1, \dots, n$, Wahrscheinlichkeitsmaße auf den Messräumen $(\Omega_i, \mathcal{A}_i)$. Dann gilt für die Produktmaße

1.

$$a\left(\bigotimes_{i=1}^n P_i, \bigotimes_{i=1}^n Q_i\right) = \prod_{i=1}^n a(P_i, Q_i), \quad (\text{A.5})$$

2.

$$d_H^2\left(\bigotimes_{i=1}^n P_i, \bigotimes_{i=1}^n Q_i\right) = 1 - \prod_{i=1}^n (1 - d_H^2(P_i, Q_i)). \quad (\text{A.6})$$

Beweis. 1. ist offensichtlich, 2. folgt aus 1. und Lemma A.9, vgl. auch Strasser (1985), S. 12. \square

Lemma A.11 (Hellingerabstand und Hellingertransformierte)

Sei $E = (\Omega, \mathcal{A}, \{P_\theta : \theta \in \Theta\})$ ein Experiment, dann gilt

$$a(P_{\theta_1}, P_{\theta_2}) = 1 - d_H^2(P_{\theta_1}, P_{\theta_2}) = H(E_{\{\theta_1, \theta_2\}})\left(\frac{1}{2}, \frac{1}{2}\right)$$

für alle $\theta_1, \theta_2 \in \Theta$.

Beweis. Folgt direkt aus den Definitionen und Lemma A.9. \square

Die folgenden beiden Beispiele erhält man durch elementare Rechnungen.

Beispiel A.12 (Hellingerabstand zweier eindim. Normalverteilungen)

Seien $\mu_1, \mu_2 \in \mathbb{R}$ und $\sigma_1, \sigma_2 > 0$. Dann ist

$$d_H^2(N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)) = 1 - \sqrt{\frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2} \exp\left\{-\frac{(\mu_1 - \mu_2)^2}{4(\sigma_1^2 + \sigma_2^2)}\right\}}.$$

Beispiel A.13 (Hellingerabstand k-dim. Normalverteilungen mit EW 0)

Seien $\Sigma_1, \Sigma_2 \in \mathbb{R}^{k \times k}$ symmetrisch und positiv definit. Dann gilt für k -dimensionale Normalverteilungen mit Erwartungswert $0 \in \mathbb{R}^k$:

$$d_H^2(N(0, \Sigma_1), N(0, \Sigma_2)) = 1 - 2^{\frac{k}{2}} \cdot \frac{(\det \Sigma_1)^{\frac{1}{4}} (\det \Sigma_2)^{\frac{1}{4}}}{(\det(\Sigma_1 + \Sigma_2))^{\frac{1}{2}}}.$$

A.2 Hilbertraumwertige Zufallsvariablen

Wir wollen hier kurz definieren, was wir unter dem Erwartungswert und der Covarianz einer hilbertraumwertigen Zufallsvariablen verstehen. Zu den folgenden Ausführungen vergleiche man Ibragimov und Rozanov (1978), S. 7 ff. oder Kuo (1975), S. 15 ff.

Definition A.14 (EW und VAR von hilbertraumwertigen Zufallsvar.)

Sei X eine Zufallsvariable mit Werten in einem separablen Hilbertraum $(H, \langle \cdot, \cdot \rangle)$.

Ein Hilbertraumelement $\gamma \in H$ mit

$$\langle \gamma, h \rangle = E\langle X, h \rangle \quad \text{für alle } h \in H$$

wird **Erwartungswert** von X genannt; man schreibt auch $EX = \gamma$.

Ein linearer Operator $S : H \rightarrow H$ mit

$$\langle Sh_1, h_2 \rangle = \text{Cov}(\langle X, h_1 \rangle, \langle X, h_2 \rangle) \quad \text{für alle } h_1, h_2 \in H$$

wird **Covarianz** von X genannt; man schreibt auch $\text{Cov}(X) = S$.

Bemerkung A.15 Der Erwartungswert $E(X)$ muss nicht notwendig existieren. Wenn er existiert, ist er eindeutig bestimmt. Ist $E(\|X\|) < \infty$, so existiert der Erwartungswert. Auch die Covarianz $\text{Cov}(X)$ muss nicht notwendig existieren. Wenn sie existiert, ist sie ein eindeutig bestimmter, symmetrischer und positiv semidefiniter Operator. Dabei nennen wir einen Operator $S : H \rightarrow H$ **positiv semidefinit**, falls für alle $h \in H$ gilt $\langle Sh, h \rangle \geq 0$. Der Operator S heißt **positiv definit**, falls für alle $h \neq 0$ gilt $\langle Sh, h \rangle > 0$. Ist $E(\|X\|^2) < \infty$, so existiert die Covarianz $\text{Cov}(X)$.

Lemma A.16

Sei X eine Zufallsvariable mit Werten in einem separablen Hilbertraum $(H, \langle \cdot, \cdot \rangle)$.

1. Existiert der Erwartungswert EX , so ist er in folgendem Sinn unabhängig vom Innenprodukt $\langle \cdot, \cdot \rangle$:

Für jede stetige Bilinearform $s : H \times H \rightarrow \mathbb{R}$, d.h. aus $h_n \rightarrow h$, $\tilde{h}_n \rightarrow \tilde{h}$ in $(H, \langle \cdot, \cdot \rangle)$ folgt $s(h_n, \tilde{h}_n) \rightarrow s(h, \tilde{h})$, gilt

$$Es(X, h) = s(EX, h) \quad \text{für alle } h \in H.$$

Dabei muss s weder symmetrisch noch definit sein.

2. Existiert die Kovarianz $\text{Cov}(X)$, so ist sie in folgender Weise vom Innenprodukt abhängig. Ist $A : H \rightarrow H$ ein symmetrischer, positiv definiten Operator, so wird durch

$$[h_1, h_2] := \langle Ah_1, h_2 \rangle \quad \text{für } h_1, h_2 \in H$$

ein weiteres Innenprodukt auf H definiert, und $(H, [\cdot, \cdot])$ ist wiederum ein Hilbertraum. Dann existiert die Kovarianz von X in $(H, [\cdot, \cdot])$ und es ist

$$\text{Cov}_{[\cdot, \cdot]}(X) = \text{Cov}(X)A.$$

Beweis. Für den Fall eines endlichdimensionalen Hilbertraums findet man den Beweis in Eaton (1983), S. 73 ff. Dieser Beweis lässt sich leicht verallgemeinern.

1. Sei s eine beliebige stetige Bilinearform. Dann existiert nach Heuser (1992), S. 185, ein stetiger linearer Operator $A : H \rightarrow H$ mit $s(h_1, h_2) = \langle Ah_1, h_2 \rangle$ für alle $h_1, h_2 \in H$. Ist A^* der zu A adjungierte Operator, so gilt für $h \in H$

$$Es(X, h) = E\langle AX, h \rangle = E\langle X, A^*h \rangle = \langle EX, A^*h \rangle = \langle AEX, h \rangle = s(EX, h),$$

insbesondere existieren alle auftretenden Erwartungswerte.

2. Für $h_1, h_2 \in H$ gilt

$$\begin{aligned} \text{Cov}([X, h_1], [X, h_2]) &= \text{Cov}(\langle AX, h_1 \rangle, \langle AX, h_2 \rangle) = \text{Cov}(\langle X, Ah_1 \rangle, \langle X, Ah_2 \rangle) \\ &= \langle \text{Cov}(X)Ah_1, Ah_2 \rangle = \langle A\text{Cov}(X)Ah_1, h_2 \rangle = \\ &= [\text{Cov}(X)Ah_1, h_2], \end{aligned}$$

insbesondere existieren alle auftretenden Kovarianzen. \square

A.3 Ein Transformationsatz

Wir betrachten eine beliebige Verteilungsfunktion H auf $[a, b]$ mit zugehörigem Wahrscheinlichkeitsmaß P_H . Durch

$$H^{-1}(s) := \sup\{t \in [a, b] \mid H(t) \leq s\}, \quad s \in [0, 1], \quad (\text{A.7})$$

wird die *rechtsseitig stetige* Quantilfunktion von H definiert. Wir setzen hierbei $\sup\{\emptyset\} := a$. H^{-1} wächst monoton, somit ist diese Funktion auch *von oben halbstetig*.

Satz A.17 (Transformationsatz) Sei $f \in L^p(H)$, $p \in [1, \infty)$, und λ bezeichne das Lebesgue-Maß auf $[0, 1]$. Dann gelten

1. $f \circ H^{-1} \in L^p(\lambda)$ und

2.

$$\int_{[a,t]} f^p dH = \int_{[0,H(t)]} (f \circ H^{-1})^p d\lambda \quad \text{für alle } t \in [a, b]. \quad (\text{A.8})$$

Beweis. Wir zeigen Teil 2. Die erste Behauptung wird dabei implizit mitbewiesen. Definieren wir für beliebiges, aber festes $t \in [a, b]$ die maßdefinierende Funktion

$$H_t : [a, b] \rightarrow [0, 1], \quad H_t(x) := \min\{H(x), H(t)\}, \quad x \in [a, b],$$

so ist das zu H_t gehörige Maß P_{H_t} das Bildmaß von $\lambda_t := \lambda|_{[0,H(t)]}$ unter der Abbildung H^{-1} , also

$$P_{H_t} = \lambda_t^{H^{-1}},$$

denn für alle $x \in [a, b]$ gilt

$$\begin{aligned} \lambda_t^{H^{-1}}([a, x]) &= \lambda_{[0,H(t)]}(\{s \in [0, 1] \mid H^{-1}(s) \in [a, x]\}) \\ &= \begin{cases} H(x), & \text{falls } H(x) \leq H(t) \\ H(t), & \text{falls } H(x) > H(t) \end{cases} \\ &= \min\{H(x), H(t)\} = H_t(x) = P_{H_t}([a, x]). \end{aligned}$$

Jetzt folgt mit dem Transformationsatz für Bildmaße

$$\begin{aligned} \int_{[a,t]} f^p(x) P_H(dx) &= \int_{[a,b]} f^p(x) P_{H_t}(dx) = \int_{[a,b]} f^p(x) \lambda_t^{H^{-1}}(dx) = \\ &= \int_{[0,1]} (f \circ H^{-1}(x))^p \lambda_t(dx) = \int_{[0,H(t)]} (f \circ H^{-1}(x))^p \lambda(dx). \end{aligned}$$

□

A.4 Sätze über Funktionen von beschränkter Variation

Satz A.18 Seien G_n , $n \in \mathbb{N}$, und G Verteilungsfunktionen auf dem Intervall $[a, b] \subset \mathbb{R}$ mit

$$\sup_{t \in [a,b]} |G_n(t) - G(t)| \rightarrow 0 \quad \text{für } n \rightarrow \infty, \quad (\text{A.9})$$

und sei $g : [a, b] \rightarrow \mathbb{R}$ von beschränkter Variation auf $[a, b]$. Dann gilt

$$\int_{[a,b]} g dG_n \rightarrow \int_{[a,b]} g dG \quad \text{für } n \rightarrow \infty.$$

Die Integrale sind hierbei nicht als Riemann-Stieltjes Integrale, sondern als Integrale bezüglich der durch G_n bzw. G eindeutig bestimmten Maße zu verstehen.

Beweis. Es genügt, die Aussage für monotone Funktionen g zu beweisen. Seien P_n, P die zu G_n, G gehörigen Wahrscheinlichkeitsmaße, und sei g eine beliebige monotone Funktion. Dann gilt unter der Voraussetzung (A.9)

$$P_n(g^{-1}((\alpha, \beta))) \xrightarrow{n \rightarrow \infty} P(g^{-1}((\alpha, \beta))) \quad \text{für alle reellen Zahlen } \alpha < \beta.$$

Man beachte dabei, dass $g^{-1}((\alpha, \beta))$ wieder ein (nicht notwendig offenes) Intervall ist. Mit Billingsley (1968), Theorem 2.2 folgt jetzt die schwache Konvergenz der Verteilungen

$$P_n^g \xrightarrow{\mathcal{D}} P^g.$$

Die Funktion g ist beschränkt, also ist $id_{g([a,b])}$ eine stetige und beschränkte Funktion auf dem Bild $g([a, b])$ von g . Jetzt folgt mit der Definition der schwachen Konvergenz

$$\int_{[a,b]} g dP_n = \int_{g([a,b])} id_{g([a,b])} dP_n^g \xrightarrow{n \rightarrow \infty} \int_{g([a,b])} id_{g([a,b])} dP^g = \int_{[a,b]} g dP.$$

□

Satz A.19 Sei $g \in BV[0, 1]$ mit Totalvariation $V(g)$. Für die Diskretisierung

$$g_n(t) := \sum_{i=1}^{n-1} g\left(\frac{i}{n}\right) 1_{[\frac{i-1}{n}, \frac{i}{n})}(t) + g(1) 1_{[\frac{n-1}{n}, 1]}(t), \quad t \in [0, 1],$$

von g gilt die Abschätzung

$$\int_0^1 (g(x) - g_n(x))^2 dx \leq \frac{V^3(g) + V(g)}{n}.$$

Beweis. Sei

$$C_{ni} := \sup \left\{ \left| g(x) - g\left(\frac{i}{n}\right) \right| : x \in \left[\frac{i-1}{n}, \frac{i}{n} \right) \right\}, \quad i = 1, \dots, n.$$

Dann ist $\sum_{i=1}^n C_{ni} \leq V(g)$. Zudem sind maximal $[V(g)]$ der Summanden C_{ni} größer als 1. Damit folgt

$$\begin{aligned} \int_0^1 (g(x) - g_n(x))^2 dx &= \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} \underbrace{\left(g(x) - g_n\left(\frac{i}{n}\right) \right)^2}_{\leq C_{ni}^2} dx \leq \sum_{i=1}^n \frac{1}{n} C_{ni}^2 \\ &= \frac{1}{n} \sum_{\substack{i=1 \\ C_{ni} > 1}}^n C_{ni}^2 + \frac{1}{n} \sum_{\substack{i=1 \\ C_{ni} \leq 1}}^n C_{ni}^2 \leq \frac{1}{n} \sum_{\substack{i=1 \\ C_{ni} > 1}}^n V^2(g) + \frac{1}{n} \sum_{i=1}^n C_{ni} \\ &\leq \frac{1}{n} V(g) V^2(g) + \frac{1}{n} V(g) = \frac{V^3(g) + V(g)}{n}. \end{aligned}$$

□

Satz A.20 Seien $f \in BV[a, b]$ und $g : [c, d] \rightarrow [a, b]$ monoton wachsend. Dann ist $f \circ g \in BV[c, d]$ und es gilt

$$V(f \circ g) \leq V(f).$$

Beweis. Die Aussage folgt direkt aus der Definition der beschränkten Variation. □

Satz A.21 (vgl. Hildebrandt (1963), S.52) Seien α von beschränkter Variation auf $[a, b]$ und $u \in C[a, b]$. Dann ist die Funktion

$$t \mapsto \int_a^t \alpha(x) u(dx), \quad t \in [a, b],$$

stetig auf $[a, b]$.

A.5 Stochastische Integrale nichtzufälliger Funktionen

Wir tragen hier kurz einige bekannte Tatsachen zum Thema „Stochastische Integrale nichtzufälliger Funktionen“ zusammen. In der Literatur finden sich die Aussagen zum Beispiel in Wentzell (1981), S. 38-41.

Sei $[a, b] \subset \mathbb{R}$ ein abgeschlossenes Intervall. Wir betrachten einen *zentrierten* stochastischen Prozess $\xi(t)$, $t \in [a, b]$, auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit *existierenden zweiten Momenten* und *unkorrelierten Zuwächsen*. Dann existiert eine bis auf eine additive Konstante eindeutig bestimmte monoton wachsende Funktion

$F : [a, b] \rightarrow \mathbb{R}$ mit $E(\xi(t) - \xi(s))^2 = F(t) - F(s)$ für $s, t \in [a, b]$ mit $s < t$. Wir betrachten im Folgenden nur solche Prozesse, für die die zugehörigen Funktionen F *rechtsseitig stetig* sind. Man beachte, dass jede Funktion F mittels der Gleichung

$$\mu((s, t]) := F(t) - F(s) \quad \text{für } s, t \in [a, b], s < t,$$

ein endliches Maß μ auf $((a, b], \mathcal{B}((a, b]))$ definiert. Wir werden im Folgenden die Funktion F mit dem zugehörigen Maß identifizieren. Der Punkt a spielt in den Betrachtungen keine Rolle, so dass wir die nicht festgelegte Konstante von F vernachlässigen und von „der“ Funktion F mit $F(a) = 0$ sprechen können.

Beispiel A.22

1. Für die Brownsche Bewegung $W(t)$, $t \in [0, 1]$, ist die zugehörige Funktion F die Identität auf $[0, 1]$.
2. Ist H eine beliebige Verteilungsfunktion auf $[a, b]$, so wird durch $\xi(t) := W(H(t))$ ein zentrierter Gaußprozess mit unabhängigen Zuwächsen definiert, und es ist

$$\text{Cov}(W(H(t)), W(H(s))) = \min\{H(t), H(s)\} \quad \text{für } t, s \in [a, b].$$

Die zugehörige Funktion ist $F = H$.

Unter obigen Voraussetzungen existiert das stochastische Integral

$$I(f) := \int_a^b f(t) \xi(dt) \quad \text{für alle } f \in L^2(F) \tag{A.10}$$

als Grenzwert in $L^2(P)$ von Summen der Form

$$\sum_{i=0}^n f_{n,i} [\xi(t_{n,i+1}) - \xi(t_{n,i})], \tag{A.11}$$

wobei $a = t_{n,0} < t_{n,1} < \dots < t_{n,n} < t_{n,n+1} = b$ eine Zerlegungsnullfolge sei, und f „geeignet“ durch die $f_{n,i}$ approximiert werde, vgl. Wentzell (1981). $I(f)$ ist somit eine quadratisch integrierbare Zufallsvariable auf (Ω, \mathcal{A}, P) .

Wir fassen einige Eigenschaften des stochastischen Integrals in einem Lemma zusammen.

Lemma A.23 *Unter obigen Voraussetzungen gilt für das stochastische Integral $I(f)$ aus (A.10)*

1. $E(I(f)) = 0$ für $f \in L^2(F)$.
2. $\text{Var}(I(f)) = \int_{(a,b]} f^2 dF$ für $f \in L^2(F)$.

$$3. \operatorname{Cov}(I(f), I(g)) = \int_{(a,b)} fg \, dF \quad \text{für } f, g \in L^2(F).$$

Lemma A.24 Seien $\xi(t)$, $t \in [a, b]$, und $f \in L^2(F)$ wie oben. Wir definieren den Prozess $\zeta(t)$ für $t \in [a, b]$ durch

$$\zeta(t) := \int_a^t f(s) \xi(ds).$$

Dann gilt für $g \in L^2(F)$ fast sicher die Gleichheit

$$\int_a^b g(t) \zeta(dt) = \int_a^b g(t) f(t) \xi(dt).$$

A.6 Kernreproduzierende Hilberträume

In diesem Abschnitt fassen wir die wichtigsten Hilfsmittel aus der Theorie der kernreproduzierenden Hilberträume zusammen. Die Aussagen finden sich in größerer Allgemeinheit zum Beispiel in Parzen (1967) oder Luschgy (1991b).

Sei $Z(s)$, $s \in [0, 1]$, ein *zentrierter Gaußprozess* mit stetigen Pfaden und Covarianzkern $R : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$, und sei (Ω, \mathcal{A}, P) der zu Grunde liegende Wahrscheinlichkeitsraum. Der Prozess entspricht einem Wahrscheinlichkeitsmaß P^Z (dem Bildmaß) auf $(C[0, 1], \mathcal{B}(C[0, 1]))$, dem Raum aller stetigen Funktionen von $[0, 1]$ nach \mathbb{R} , versehen mit der Supremumsnorm und der zugehörigen Borelschen σ -Algebra.

Durch den Covarianzkern R wird der zugehörige **kernreproduzierende Hilbertraum** $(H(R), \langle \cdot, \cdot \rangle_{H(R)})$ mittels der beiden Forderungen

$$R(\cdot, s) \in H(R) \quad \text{für alle } s \in [0, 1], \quad (\text{A.12})$$

$$\langle h, R(\cdot, s) \rangle_{H(R)} = h(s) \quad \text{für alle } s \in [0, 1] \text{ und alle } h \in H(R) \quad (\text{A.13})$$

eindeutig definiert, d.h. sowohl $H(R)$ wie auch das Innenprodukt sind eindeutig festgelegt. Es gilt weiterhin

$$H(R) \subset C[0, 1].$$

Seien $h \in H(R)$ und $Z_h(s) := h(s) + Z(s)$, $s \in [0, 1]$, der Prozess mit Erwartungswertfunktion h und gleichem Covarianzkern R . Sei P^{Z_h} das zugehörige Wahrscheinlichkeitsmaß auf $C[0, 1]$.

Dann sind die Maße P^{Z_h} und P^Z äquivalent, und für die Dichte $\frac{dP^{Z_h}}{dP^Z}$ gilt

$$\frac{dP^{Z_h}}{dP^Z}(u) = \exp \left\{ L(h)(u) - \frac{1}{2} \|h\|_{H(R)}^2 \right\} \quad \text{für } u \in C[0, 1], \quad (\text{A.14})$$

wobei die lineare Isometrie $L : H(R) \rightarrow L^2(C[0, 1], \mathcal{B}(C[0, 1]), P^Z)$ dicht definiert ist durch die Gleichung

$$L \left(\int_{[0,1]} R(\cdot, s) \mu(ds) \right) (u) := \int_{[0,1]} u d\mu \quad (\text{A.15})$$

für $u \in C[0, 1]$ und alle endlichen signierten Maße μ auf $[0, 1]$. Man beachte, dass die Menge $\left\{ \int_{[0,1]} R(\cdot, s) \mu(ds) \mid \mu \text{ endliches signiertes Maß} \right\}$ dicht in $H(R)$ liegt und die endlichen signierten Maße auf $[0, 1]$ mit dem topologischen Dual von $C[0, 1]$ identifiziert werden können.

Der stochastische Prozess

$$(C[0, 1], \mathcal{B}(C[0, 1]), P^Z, \{L(h) \mid h \in H(R)\})$$

ist ein *Gaußprozess*. Für alle $h, g \in H(R)$ gelten die Formeln

$$\mathcal{L}(L(h) \mid P^Z) = N(0, \|h\|_{H(R)}^2), \quad (\text{A.16})$$

$$\mathcal{L}(L(g) \mid P^{Z_h}) = N(\langle g, h \rangle_{H(R)}, \|g\|_{H(R)}^2) \quad (\text{A.17})$$

und

$$\text{Cov}_{P^Z}(L(h), L(g)) = \langle h, g \rangle_{H(R)}. \quad (\text{A.18})$$

Ein wichtiges Beispiel soll diese Begriffe verdeutlichen. Alle Aussagen des Beispiels finden sich in der anfangs zitierten Literatur.

Beispiel A.25 Es sei W eine Version der Brownschen Bewegung mit stetigen Pfaden auf dem Intervall $[0, 1]$. Der zugehörige Covarianzkern ist dann

$$R(s, t) = \min\{s, t\}, \quad s, t \in [0, 1].$$

Damit erhält man als zugehörigen kernreproduzierenden Hilbertraum

$$H(R) = \left\{ h : [0, 1] \rightarrow \mathbb{R} \mid \exists h' \in L^2(\lambda) : h(s) = \int_{[0,s]} h'(x) \lambda(dx), s \in [0, 1] \right\}.$$

Für $h_1, h_2 \in H(R)$ mit zugehörigen h'_1, h'_2 (diese sind λ -fast sicher eindeutig bestimmt) berechnet sich das Innenprodukt durch

$$\langle h_1, h_2 \rangle_{H(R)} = \int_{[0,1]} h'_1 h'_2 d\lambda = \langle h'_1, h'_2 \rangle_{L^2(\lambda)}.$$

Ist $h \in H(R)$ und ist „das“ zugehörige h' rechtsseitig stetig und von beschränkter Variation, so kann man die lineare Isometrie L für dieses h explizit berechnen:

$$L(h)(u) = \int_0^1 h'(x) u(dx), \quad u \in C[0, 1],$$

wobei das Integral als Riemann-Stieltjes Integral zu verstehen ist.

A.7 Absolutstetige Funktionen

Definition A.26 (Absolutstetige Funktionen) Eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ heißt **absolutstetig** (genauer: **auf endlichen Intervallen absolutstetig**), wenn für alle $C > 0$ und alle $\epsilon > 0$ ein $\delta > 0$ existiert, so dass für jeweils endlich viele disjunkte Intervalle $(a_1, b_1), \dots, (a_n, b_n) \subset (-C, C)$, $n \in \mathbb{N}$, mit $\sum_{k=1}^n |b_k - a_k| < \delta$ gilt

$$\sum_{k=1}^n |g(b_k) - g(a_k)| < \epsilon.$$

Jede absolutstetige Funktion ist natürlich auch stetig. Es gilt sogar

Satz A.27 (vgl. Rudin (1987), S. 146) Eine Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ ist genau dann absolutstetig, wenn eine Funktion g' mit

$$g(b) - g(a) = \int_{[a,b]} g' d\lambda, \quad \text{für alle } -\infty < a < b < \infty$$

existiert.

In diesem Fall ist g λ -fast überall differenzierbar mit Ableitung g' . In diesem Sinne werden wir künftig nur noch von der Ableitung einer absolutstetigen Funktion sprechen.

Aus diesem Satz folgt direkt

Korollar A.28 Ist $g : \mathbb{R} \rightarrow \mathbb{R}$ absolutstetig mit $\lim_{x \rightarrow \pm\infty} g(x) = 0$, und ist die Ableitung (im Sinne von Satz A.27) Lebesgue-integrierbar, so gilt

$$\int g' d\lambda = 0.$$

Schließlich benötigen wir noch

Lemma A.29 Ist $g : \mathbb{R} \rightarrow \mathbb{R}$ absolutstetig mit Ableitung g' , so gilt

$$\lambda(\{x \in \mathbb{R} \mid g(x) = 0, g'(x) \neq 0\}) = 0.$$

Beweis. Sei $A := \{x \in \mathbb{R} \mid g(x) = 0, g'(x) \neq 0\}$. Da g nach Satz A.27 λ -fast sicher differenzierbar ist, können wir g für den Beweis dieser Aussage o.B.d.A. als differenzierbar voraussetzen. Nehmen wir weiter $\lambda(A) > 0$ an, so muss ein $x_0 \in A$ existieren, so dass in jeder Umgebung von x_0 Punkte aus A liegen. Sonst wäre A abzählbar und somit eine Lebesgue-Nullmenge. Also existieren Punkte $x_n \in A \setminus \{x_0\}$ mit $x_n \rightarrow x_0$. Es folgt

$$g'(x_0) = \lim_{n \rightarrow \infty} \frac{g(x_n) - g(x_0)}{x_n - x_0} = 0,$$

was im Widerspruch zu $x_0 \in A$ steht. Also muss $\lambda(A) = 0$ gelten. \square

Lemma A.30 *Ist Q ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$, das den Voraussetzungen 2.9 genügt, d.h. Q besitzt eine absolutstetige Lebesgue-Dichte q mit Ableitung q' gemäß Satz A.27 und*

$$0 < \int \left(\frac{q'}{q} \right)^2 dQ < \infty,$$

so folgt

$$\lim_{x \rightarrow \pm\infty} q(x) = 0.$$

Beweis. Aus $\frac{q'}{q} \in L^2(Q)$ folgt $\frac{q'}{q} \in L^1(Q)$ und damit $q' \in L^1(\lambda)$.

Nehmen wir an, es sei $\lim_{x \rightarrow \infty} q(x) \neq 0$, so existieren ein $C > 0$ und eine Folge $(x_n)_{n \in \mathbb{N}}$ mit $x_n \rightarrow \infty$ und $q(x_n) > C$. Da q eine Wahrscheinlichkeitsdichte ist, müssen aber auch ein $0 < c < C$ und eine Folge $(y_n)_{n \in \mathbb{N}}$ mit $y_n \rightarrow \infty$ und $q(y_n) < c$ existieren. Nach Satz A.27 gilt aber

$$q(x_n) - q(0) = \int_{[0, x_n]} q' d\lambda \quad \text{und} \quad q(y_n) - q(0) = \int_{[0, y_n]} q' d\lambda,$$

was auf Grund von $q' \in L^1(\lambda)$ beim Grenzübergang $n \rightarrow \infty$ zum Widerspruch führt. Folglich muss $\lim_{x \rightarrow \infty} q(x) = 0$ gelten, und analog zeigt man $\lim_{x \rightarrow -\infty} q(x) = 0$. \square

Index

- absolutstetige Funktion, 148
- Affinität, 138
- approximative Designtheorie, 90
- Asymptotisch Relative Effizienz, 80
- Asymptotische Äquivalenz
 - schwache, 36
 - starke, 36
- Change-Point Problem, 82
- Defizienz, 13
- Design
 - E -optimales, 105
 - approximatives, 90
 - äquivalentes, 91
 - exaktes, 88
 - konkretes, 88
 - mindestens so informatives, 89, 90
 - stetiges, 90
- Designfolge, 41
 - H -äquidistante, 42
 - asymptotisch mindestens so informativ, 93
 - asymptotisch zulässige, 93
 - lokal asymptotisch mindestens so informativ, 98, 103
 - lokal asymptotisch zulässige, 98
 - schwach lokal asymptotisch mindestens so informativ, 100, 103
- Entscheidungsfunktion
 - randomisierte, 9
 - verallgemeinerte, 11
- Entscheidungspaar, 12
- Entscheidungsproblem, 9
- Entscheidungsraum, 9
- Experiment, 9
 - ϵ -defizientes, 12
 - äquivalentes, 13
 - differenzierbares, 47, 114
 - dominiertes, 11
 - gleich informatives, 13
 - homogenes, 18, 24
 - mindestens so informatives, 12, 13
 - nicht informativeres, 12, 13
 - polnisches, 14
 - reguläres, 114
 - stetiges, 25
- Experimentfolge
 - asymptotisch homogene, 24
 - asymptotisch stetige, 25
- Fisher-Information, 48
- Fisher-Informationsoperator, 114
- Gaußexperiment, 18
- Gaußprozess, 18
- Gaußshift, 18
- Grenzdesign, 41
- H_n - H -Konvergenzklasse, 42
- Hájek-Bedingungen, 48
- Hauptparameter, 112
- Hellingerabstand, 137
- Hellingertransformierte (Experiment), 138
- Informationsabstand
 - von Designs, 91
 - von Experimenten, 13
- Informationsmatrix, 89
- Kernreproduzierender Hilbertraum, 146
- L-Raum, 10
- Löwner-Ordnung, 16

-
- LeCam-Bahadur, Satz von, 117
 - Likelihood-Quotient, 25
 - Likelihood-Quotienten-Test, 76
 - Lindae, Satz von, 34
 - Linearer Prozess, 26
 - Lokalisierung, 46

 - Minimax-Schranke von Hájek-LeCam, 30
 - Modellmatrix, 15

 - Nebenparameter, 112
 - Niveau eines Tests, 27
 - Normalverteilung
 - auf separablem Hilbertraum, 21

 - Operator
 - positiv definit, 140
 - positiv semidefinit, 140

 - Pitman-Effizienz, 80
 - präkompakt, 34

 - Regressionsmodell
 - Allgemeines, 40
 - Lineares, 40
 - Risikofunktion, 9, 11

 - Schätzer
 - effizient, 116
 - Schätzfolge
 - adaptive, 118
 - asymptotisch effizient, 116
 - lokal asymptotisch effizient, 116
 - supereffizient, 117
 - Schwache Konvergenz
 - von Experimenten, 24
 - von Maßen, 26
 - Schwache Topologie (Experiment), 24
 - Störparameter, 112
 - Standard-Gaußshift, 20
 - Standard-Gaußshift auf
 - endlichdimensionalem Hilbertraum, 21
 - unendlichdimensionalem Hilbertraum, 21
 - Standard-Simplex, 138
 - Starke Konvergenz von Experimenten, 24
 - Starke Topologie (Experiment), 13
 - Stochastische Konvergenz, 25

 - Test
 - α -ähnlicher, 27
 - äquivalenter, 28
 - optimaler, 28
 - unverfälschter, 27
 - verteilungsfreier, 75
 - zulässiger, 28
 - Testfolge
 - adaptive, 129
 - asymptotisch α -ähnliche, 28
 - asymptotisch optimale, 28
 - asymptotisch unverfälschte, 28
 - totalbeschränkt, 34
 - Totalvariationsnorm, 137

 - Variationsnorm, 137
 - Verlustfunktion, 9
 - Verteilungskonvergenz
 - von Entscheidungsfunktionen, 29
 - von Maßen, 26

 - Wurzel einer Matrix, 22

 - Zentrale Folge, 27

Symbolverzeichnis

(D, W) , 12	ARE, 80
(E, D, W) , 9	$\hat{\beta}_n$, 124
$A(\Theta)$, 138	$\hat{\beta}_n^*$, 122
$BV[a, b]$, 43	$\delta_D(H_{1n_0}, H_{2n_0})$, 91
$B^{\frac{1}{2}}$, 53	$\delta(\cdot, \cdot)$, 13
$C_b(D)$, 11	$\frac{dP}{dQ}$, 25
$C_c(D)$, 11	\geq_L , 16
$E \sim F$, 13	$\hat{\varphi}_n$, 131
$E \stackrel{\epsilon}{\supseteq} F$, 12	$\hat{\varphi}_n^*$, 131
$E \supseteq F$, 13	\xrightarrow{D} , 26
$E \underset{(D, W)}{\supseteq} F$, 12	\subseteq_q , 17
$E \underset{(D, W)}{\stackrel{\epsilon}{\supseteq}} F$, 12	φ_n^{LQ} , 76
E_Q , 38	$\mathcal{L}(X P)$, 14
E_α , 138	$a(\cdot, \cdot)$, 138
$E_n \xrightarrow{w} E$, 24	$d_H(\cdot, \cdot)$, 137
$E_n^{loc}(H_n)$, 97	g^{LQ} , 79
$E_n^{loc}(Q)$, 113	$o_{P_n}(1)$, 25
E_n^{loc, f_0} , 47	$p(x; h)$, 114
E_n^{loc} , 47	$\mathcal{E}(\cdot)$, 9
G_H , 20, 21	$\mathcal{E}(\cdot) / \sim$, 13
$H(\cdot)$, 138	$\mathcal{F}(\mathcal{X}, \mathcal{A})$, 28
$I(\cdot)$, 114	$\mathcal{F}_\alpha(H, K)$, 28
J , 48	$\mathcal{F}_\alpha^s(H, K)$, 28
$L(E)$, 10	\mathcal{G} , 112
$O_P(1)$, 122	$\mathcal{L}(A, C)$, 15
Q_s , 38	$\mathcal{L}(H; L^2(\nu))$, 114
S_α , 138	$\mathcal{M}(\mathcal{X})$, 10
$V(\cdot)$, 66	$\mathcal{N}(A, C)$, 16
$[\cdot]$, 44	\mathcal{Q} , 43
$\Delta_D(H_{1n_0}, H_{2n_0})$, 91	$\mathcal{R}(E, D)$, 9
$\Delta(\cdot, \cdot)$, 13	$\mathcal{R}_a(E, D)$, 11
\mathcal{G}_H , 122	
\mathcal{G}_{sym} , 122	
$\ \cdot\ _V$, 137	

Literaturverzeichnis

- Bauer, H. (1991). *Wahrscheinlichkeitstheorie* (4. Aufl.). Berlin, New York: de Gruyter.
- Bauer, H. (1992). *Maß- und Integrationstheorie* (2. Aufl.). Berlin, New York: de Gruyter.
- Becker, C. (1983). Schwache asymptotische Normalität von statistischen Experimenten bei unabhängigen, nicht notwendig identisch verteilten Beobachtungen. *Bayreuther Mathematische Schriften*, 13, 1-153.
- Begun, J., Hall, W., Huang, W.-M. und Wellner, J. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.*, 11, 432-452.
- Beran, R. (1974). Asymptotically efficient adaptive rank estimates in location models. *Ann. Statist.*, 2, 63-74.
- Bickel, P. (1982). On adaptive estimation. *Ann. Statist.*, 10, 647-671.
- Bickel, P., Klaassen, C., Ritov, Y. und Wellner, J. (1993). *Efficient and adaptive estimation for semiparametric models*. New York, Heidelberg, Berlin: Springer. (urspr. John Hopkins University Press)
- Billingsley, P. (1968). *Convergence of probability measures*. New York, Chichester, Brisbane, Toronto: Wiley.
- Bischoff, W. (1998). A functional central limit theorem for regression models. *Ann. Statist.*, 26, 1398-1410.
- Bischoff, W. und Miller, F. (2000a). Asymptotically optimal tests and optimal designs for testing the mean in regression models with applications to change-point problems. *Ann. Inst. Statist. Math.*, 52, 658-679.
- Bischoff, W. und Miller, F. (2000b). *In linear regression models with unknown distribution of the error the F-test is asymptotically optimal in the class of test statistics based on linear statistics* (Preprint). Fakultät für Mathematik der Universität Karlsruhe.

- Brown, L. und Low, M. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, *24*, 2384-2398.
- Choi, S., Hall, W. und Schick, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Ann. Statist.*, *24*, 841-861.
- Eaton, M. (1983). *Multivariate statistics*. New York, Chichester, Brisbane, Toronto: Wiley.
- Elstrodt, J. (1996). *Maß- und Integrationstheorie*. Berlin: Springer.
- Farrell, R. (1967). Weak limits of sequences of Bayes procedures in estimation theory. Proc. 5th Berkeley Symp. *Math. Statistics Prob.*, *1*, 83-111.
- Grama, I. und Nussbaum, M. (1998). Asymptotic equivalence of nonparametric generalized linear models. *Probability Theory and Related Fields*, *111*, 167-214.
- Hájek, J. und Šidák, Z. (1967). *Theory of rank tests*. New York, London: Academic Press.
- Heuser, H. (1992). *Funktionalanalysis* (3. Aufl.). Stuttgart: Teubner.
- Hildebrandt, T. (1963). *Introduction to the theory of integration*. New York: Academic Press.
- Hsieh, D. und Manski, C. (1987). Monte Carlo evidence on adaptive maximum likelihood estimation of a regression. *Ann. Statist.*, *15*, 541-551.
- Ibragimov, I. und Has'minskii, R. (1981). *Statistical Estimation - Asymptotic Theory*. New York, Heidelberg, Berlin: Springer.
- Ibragimov, I. und Rozanov, Y. (1978). *Gaussian random processes*. New York, Heidelberg, Berlin: Springer.
- Janssen, A., Milbrodt, H. und Strasser, H. (1985). Infinitely divisible statistical experiments. In *Lecture notes in statistics* (Vol. 27). Berlin, Heidelberg, New York, Tokyo: Springer.
- Jurečková, J. und Sen, P. (1984). On adaptive scale-equivalent M-estimators in linear models. *Statistics and Decisions*, *1*, 31-46.
- Koul, H. L. und Susarla, V. (1983). Adaptive estimation in linear regression. *Statistics and Decisions*, *1*, 379-400.
- Kuo, H. (1975). Gaussian measures in Banach spaces. In *Lecture notes in mathematics* (Vol. 463). Berlin, Heidelberg, New York: Springer.
- LeCam, L. (1979). On a theorem of J. Hájek. In J. Jurečková (Ed.), *Contributions to statistics - Hájek memorial volume* (p. 119-135). Dordrecht: D. Reidel.

-
- LeCam, L. (1985). Sur l'approximation de familles de mesures par des familles gaussiennes. *Ann. Inst. Henri Poincaré*, 21, 225-287.
- LeCam, L. (1986). *Asymptotic methods in statistical decision theory*. New York, Berlin, London: Springer.
- LeCam, L. und Yang, G. (1990). *Asymptotics in statistics*. New York, Berlin, London: Springer.
- Lindae, D. (1972). *Distributions of likelihood ratios and convergence of experiments*. Ph.D. Thesis. Univ. of California. Berkeley.
- Luschgy, H. (1991a). Ordering regression models of Gaussian processes. In *Stochastic orders and decision under risk. IMS Lect. Notes, Monogr. Ser., Proc. Int. Workshop 1989* (Vol. 19). Hamburg.
- Luschgy, H. (1991b). Testing one-sided hypothesis for the mean of a Gaussian process. *Metrika*, 38, 179-194.
- Luschgy, H. (1992). Comparison of location models for stochastic processes. *Probability Theory and Related Fields*, 93, 39-66.
- Milstein, G. und Nussbaum, M. (1998). Diffusion approximation for nonparametric autoregression. *Probability Theory and Related Fields*, 112, 535-543.
- Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, 24, 2399-2430.
- Parzen, E. (1967). *Time series analysis papers*. San Francisco: Holden-Day.
- Portnoy, S. und Koenker, R. (1989). Adaptive L -estimation for linear models. *Ann. Statist.*, 17, 362-381.
- Pukelsheim, F. (1993). *Optimal design of experiments*. New York, Chichester, Brisbane, Toronto: Wiley.
- Rudin, W. (1987). *Real and complex analysis* (3. Aufl.). New York, Hamburg, Toronto: McGraw-Hill.
- Schick, A. (1986). On asymptotically efficient estimation in semiparametric models. *Ann. Statist.*, 14, 1139-1151.
- Schick, A. (1987). A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inf.*, 16, 89-105. (Correction (1989) 22, 269-270.)
- Silvey, S. (1980). *Optimal design*. London, New York: Chapman and Hall.
- Stapleton, J. (1995). *Linear statistical models*. New York, Chichester, Brisbane, Toronto, Singapore: Wiley.

- Stein, C. (1956). Efficient nonparametric testing and estimation. In *Proc. Third Berkeley Symp. Math. Statist. Prob.* (Vol. 1, p. 187-196). Berkeley: Univ. of California Press.
- Stone, C. (1975). Adaptive maximum likelihood estimators of a location parameter. *Ann. Statist.*, 3, 267-284.
- Strasser, H. (1985). *Mathematical theory of statistics*. Berlin, New York.: de Gruyter.
- Torgersen, E. (1991). *Comparison of statistical experiments*. Cambridge: Cambridge University Press.
- Wentzell, A. (1981). *A course in the theory of stochastic processes*. McGraw-Hill.
- Witting, H. (1985). *Mathematische Statistik I*. Stuttgart: Teubner.
- Witting, H. und Müller-Funk, U. (1995). *Mathematische Statistik II*. Stuttgart: Teubner.

Lebenslauf

Persönliche Daten:

Name: Guido Bader
Geboren am: 12.08.1972 in Offenburg
Eltern: Heinrich und Rita Bader

Schulbildung:

1978 – 1982 Grundschole in Rheinau-Helmlingen
1982 – 1991 Windeck-Gymnasium in Bühl
Juni 1991 Abiturprüfung

Studium:

Okt. 1992 – Sept. 1994 Grundstudium in Wirtschaftsmathematik an der Universität Karlsruhe
Sept. 1994 – Dez. 1997 Hauptstudium in Wirtschaftsmathematik an der Universität Karlsruhe

Tätigkeiten während des Studiums:

Okt. 1994 – Sept. 1996 Wissenschaftliche Hilfskraft am Mathematischen Institut I der Universität Karlsruhe
Okt. 1996 – Dez. 1997 Wissenschaftliche Hilfskraft am Institut für Mathematische Stochastik der Universität Karlsruhe

Berufstätigkeit:

Juli 1991 – Juni 1992 Grundwehrdienst in der Graf-Stauffenberg-Kaserne in Sigmaringen
Jan. 1998 – April 1998 Projektarbeit bei der UHU GmbH, Bühl
April 1998 – März 2001 Wissenschaftlicher Angestellter am Institut für Mathematische Stochastik der Universität Karlsruhe
seit April 2001 Wissenschaftlicher Angestellter am Institut für Industrielle Bauproduktion der Universität Karlsruhe