Asymptotisch effizientes Schätzen von Funktionalen im nichtparametrischen Fall

I n a u g u r a l - D i s s e r t a t i o n

zur
Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Andreas Knoch

aus Düsseldorf

Düsseldorf, Oktober 2013

Aus dem Mathematischen Institut
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent:	Prof. Dr. A. Janssen
Koreferent:	Prof. Dr. A. Meister
Tag der mündlichen Prüfung:	22.01.2014

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der asymptotischen Güte von Schätzern $T_n : \Omega^n \to \mathbb{R}$ für Funktionale $\kappa : \mathcal{P} \to \mathbb{R}$ auf nichtparametrischen Familien von Wahrscheinlichkeitsmaßen $\mathcal{P}$. Betrachtet werden Schätzer T_n zum Stichprobenumfang n, die für $n \to \infty$ eine Verteilungskonvergenz aufweisen in Form von

$$\sqrt{n}(T_n - \kappa(P_0)) \xrightarrow[n\to\infty]{\mathcal{D}} Y \sim L$$

unter festem, aber beliebigem $P_0 \in \mathcal{P}$ und zudem eine Regularitätsbedingung erfüllen. Ist die Varianz der Grenzverteilung minimal, so heißt der Schätzer asymptotisch effizient. Ziel der Arbeit ist es, für einen gegebenen Schätzer T_n die asymptotische Effizienz nachzuweisen. Zur Annäherung an dieses Problem werden in den Kapiteln eins und zwei zunächst L_2-differenzierbare parametrische Teilmodelle mit Tangente g betrachtet. Ist κ differenzierbar, so setzt sich die Verteilung L für reguläre Schätzer auf diesen Teilmodellen nach dem Faltungssatz von Hájek-Le Cam aus dem Faltungsprodukt einer Normalverteilung $\mathcal{N}(0, \sigma^2)$ und eines weiteren Wahrscheinlichkeitsmaßes ν zusammen. $\sigma^2 = \frac{\kappa'(P_0)^2}{I_0}$ mit der Fisher-Information $I_0 = \int g^2 dP_0$ hängt hierbei nur von den lokalen Eigenschaften des speziellen Teilmodells, genauer der Tangente g in P_0, und der Ableitung des Funktionals, nicht aber vom Schätzer T_n ab. Die Varianz ist somit dann minimal, wenn ν verschwindet, und σ^2 bildet eine untere Schranke. Da die parametrischen Teilmodelle im nichtparametrischen Modell $\mathcal{P}$ enthalten sind, lässt sich dieses Konzept in Kapitel drei auf den nichtparametrischen Fall erweitern, indem eine große Anzahl solcher Modelle bzw. Tangenten g aus dem Tangentialraum $T(P_0, \mathcal{P})$ mit entsprechenden Teilmodellen, die g als Tangente besitzen, gleichzeitig betrachtet wird. Im Laufe dieser Arbeit wird dabei gezeigt, dass es bereits ausreicht eine dichte Teilmenge des Tangentialraums und zugehörige Teilmodelle heranzuziehen, was eine erhebliche Vereinfachung darstellt, um für gegebene Schätzer die asymptotische Effizienz nachzuweisen. So kann in Kapitel vier in großer Allgemeinheit die asymptotische Effizienz von L-Statistiken, den kanonischen Schätzern der L-Funktionale, nachgewiesen werden. In Kapitel fünf kann mit diesen Methoden zudem die asymptotische Effizienz von Schätzern spezieller Funktionale aus der Survival Analysis und insbesondere erneut sehr schnell jene des bekannten Nelson-Aalen- und Kaplan-Meier-Schätzers sowohl für unzensierte als auch für rechtszensierte Daten gezeigt werden.

Abstract

The present thesis deals with the asymptotic quality of estimators $T_n : \Omega^n \to \mathbb{R}$ for functionals $\kappa : \mathcal{P} \to \mathbb{R}$ on nonparametric families of probability measures $\mathcal{P}$. Treated are estimators T_n for sample size n, that show convergence in distribution in the form of

$$\sqrt{n}(T_n - \kappa(P_0)) \xrightarrow[n\to\infty]{\mathcal{D}} Y \sim L$$

under fixed, but arbitrary $P_0 \in \mathcal{P}$ and meet some regularity condition. If the variance of the limit distribution is minimal, the estimator is called asymptotically efficient. To approach this problem in chapter one and two first L_2-differentiable parametric submodels with tangent g are considered. If κ is differentiable, for regular estimators on these submodels the distribution L is composed by the convolution of a normal distribution $\mathcal{N}(0, \sigma^2)$ and another probability measure ν. Here $\sigma^2 = \frac{\kappa'(P_0)^2}{I_0}$ with the Fisher-Information $I_0 = \int g^2 dP_0$ only depends on the local characteristics of the special submodel, more precisely on the tangent g in P_0, and on the derivative of the functional, but not on the estimator T_n. Therefore, the variance is minimal, if ν vanishes, and σ^2 provides a lower bound. Since the parametric submodels are contained in the nonparametric model $\mathcal{P}$, in chapter three this concept can be extended to the nonparametric case by simultaneously taking a big amount of such submodels respectively tangents g from the tangent space $T(P_0, \mathcal{P})$ with corresponding submodels, which have tangent g, into account. In the course of the thesis it will be shown, that it suffices to use a dense subset of the tangent space and corresponding submodels, which provides a significant simplification for proving asymptotic efficiency. This way, in chapter four the asymptotic efficiency of L-statistics, the canonical estimators for L-functionals, can be proved in big generality. Moreover, in chapter five the asymptotic efficiency of estimators for special functionals from the Survival Analysis, especially of the well-known Nelson-Aalen- und Kaplan-Meier-Estimator for both uncensored and right censored data can be shown very fast by these methods.

Inhaltsverzeichnis

Einleitung

Seien $x_1, ..., x_n$ Beobachtungen eines n-fach unabhängig wiederholten Experiments, also n Realisierungen von i.i.d. Zufallsvariablen $X_1, ..., X_n : \Omega' \to \Omega$ mit unbekannter Verteilung $P \in \mathcal{P}$ aus einer Familie von Wahrscheinlichkeitsmaßen $\mathcal{P}$ auf Ω. Je mehr über diese Verteilung bekannt ist, desto kleiner ist die Menge der in Frage kommenden Verteilungen $\mathcal{P}$. In dieser Arbeit betrachten wir Fälle, in denen von vornherein nur sehr wenig über die Verteilung bekannt ist, so dass $\mathcal{P}$ eine nichtparametrische Familie von Wahrscheinlichkeitsmaßen darstellt. Das heißt $P \in \mathcal{P}$ ist nicht durch einen Parameter $\vartheta \in \mathbb{R}^k, k \in \mathbb{N}$ identifizierbar. Ist beispielsweise überhaupt nichts über P bekannt, so ergibt sich $\mathcal{P}$ erst einmal als Menge aller Wahrscheinlichkeitsmaße.

Ein statistisches Funktional ist eine Abbildung $\kappa : \mathcal{P} \to \mathbb{R}$. Sei nun κ ein solches Funktional. Anhand der vorliegenden Beobachtungen soll dieses Funktional geschätzt werden. Ist $T_n : \Omega^n \to \mathbb{R}$ ein Schätzer für κ, so stellt sich die Frage nach der Güte desselben. In dieser Arbeit sollen Kriterien für ein optimales asymptotisches Verhalten des Schätzers, also für ein optimales Verhalten bei gegen unendlich wachsendem Stichprobenumfang n, hergeleitet werden. Zeigt ein Schätzer ein noch näher zu spezifizierendes optimales Grenzverhalten, so ist dies nicht unmittelbar auf einen finiten Stichprobenumfang n zu übertragen, es bietet aber dennoch einen Anhaltspunkt für die in der Praxis auftretende finite Situation, um den Schätzer bei festem n einem anderen vorzuziehen.

Was ist nun das optimale asymptotische Verhalten eines solchen Schätzers? Zunächst einmal setzen wir eine Verteilungskonvergenz voraus. So soll T_n normiert und zentriert am Funktional unter P_0 für alle $P_0 \in \mathcal{P}$ in Verteilung gegen eine Zufallsvariable Y konvergieren in Form von

$$\sqrt{n}(T_n - \kappa(P_0)) \xrightarrow[n\to\infty]{\mathcal{D}} Y \sim L.$$

Diese Voraussetzung erscheint sinnvoll, wenn wir einen asymptotisch besten Schätzer finden wollen. Für den Fall, dass diese Bedingung nicht vorausgesetzt wird, siehe beispielsweise Pfanzagl [32]. Wünschenswert für einen solchen Schätzer wäre, dass der Erwartungswert der Grenzvariablen $E(Y) = 0$ beträgt, der Schätzer also in gewissem Sinne (und mit einer gewissen Rate) "asymptotisch erwartungstreu" ist, und die Varianz $Var(Y)$ unter allen zugelassenen Schätzern minimal ist, was zumindest in der Grenzverteilung interpretierbar ist als eine möglichst geringe erwartete Abweichung vom zu schätzenden Wert des Funktionals. Zugelassen werden hier aber nicht alle Schätzer, die obige Verteilungskonvergenz aufweisen mit $E(Y) = 0$, sondern wir fordern zusätzlich die Regularität der Schätzer, das heißt eine Verteilungskonvergenz auch unter Alternativen. Ziel ist es nun, für solche Schätzer eine asymptotische untere Schranke für die Varianz $Var(Y)$, die sogenannte Informationsschranke, herzuleiten. Haben wir eine solche Schranke ermittelt und finden einen regulären Schätzer, der diese Schranke annimmt, so ist er in diesem Sinne offenbar asymptotisch optimal. Wir sprechen dann von asymptotischer Effizienz.

Fixieren wir einen Fußpunkt $P_0 \in \mathcal{P}$, so nennen wir eine einparametrische Familie von Wahrscheinlichkeitsmaßen $\{P_\vartheta : \vartheta \in \Theta\} \subset \mathcal{P}, \Theta \subset \mathbb{R}$, die durch P_0 verläuft, also $0 \in \mathring{\Theta}$, auch Kurve (lokal) in $\mathcal{P}$. Da diese parametrischen Modelle also im nichtparametrischen Modell $\mathcal{P}$ als Teilmodelle enthalten sind, besteht das Konzept nun darin, zunächst untere Schranken für die Varianz der Grenzverteilung L in den Teilmodellen zu finden und hieraus anschließend eine für das gesamte Modell $\mathcal{P}$ zu bilden. Dabei betrachten wir hier nur die in Kapitel eins definierten von Le Cam eingeführten L_2-differenzierbaren Kurven in $\mathcal{P}$ mit Tangente $g \in L_2(P_0)$. Viele Eigenschaften dieser Kurven hängen nur von dieser Tangente ab, was zum Tangentialraum, dem Abschluss der linearen Hülle der Menge aller Tangenten, überleitet. Dieser Tangentialraum bildet einen Hilbertraum, der für unsere Untersuchungen eine schöne wohl verstandene Struktur bietet. Zudem wird in Kapitel eins das volle Modell eingeführt, das Standard-nichtparametrische Modell, das dieser Arbeit zugrunde liegt.

In Kapitel zwei untersuchen wir nun also L_2-differenzierbare einparametrische Kurven und folgen den Arbeiten von Le Cam [27] und Hájek [13].
Es werden zunächst einige Aspekte der Le Cam'schen Theorie der lokalen asymptotischen Normalität wiederholt und an unsere Zwecke angepasst. Mit Hilfe dieser

Theorie liefert der Hájek-Le Cam-Faltungssatz für solche L_2-differenzierbaren Kurven unter der zusätzlichen Bedingung der Regularität des Schätzers, das heißt einer Verteilungskonvergenz auch unter Alternativen, das zentrale Hilfsmittel, um eine untere Schranke für die Varianz unserer Grenzverteilung L zu finden. Dieser Satz besagt, dass die Grenzverteilung sich aus dem Faltungsprodukt einer Normalverteilung $\mathcal{N}(0, \sigma^2)$ und eines weiteren Wahrscheinlichkeitsmaßes zusammensetzt, Y sich also als Summe $Y = Z + W$ aus einer normalverteilten Zufallsvariablen $Z \sim \mathcal{N}(0, \sigma^2)$ und einer von Z stochastisch unabhängigen Zufallsvariablen W schreiben lässt. σ^2 ist dabei gegeben durch die Tangente g und die Ableitung des Funktionals κ und entspricht der Cramér-Rao-Schranke

$$\sigma^2 = \frac{\kappa'(P_0)^2}{I_0}$$

mit der Fisher-Information $I_0 = \int g^2 dP_0$. Der Faltungssatz liefert allerdings mehr als nur eine asymptotische Cramér-Rao-Schranke, da er auch die optimale Gestalt der Grenzverteilung angibt. So ist $Var(Y) = Var(Z) + Var(W)$ nun minimal, falls W verschwindet. Die optimale Grenzverteilung in diesem Sinne ergibt sich also zu $\mathcal{N}(0, \sigma^2)$ und σ^2 bildet die gesuchte untere Schranke. Besitzt ein Schätzer T_n diese Grenzverteilung und nimmt die Schranke somit an, dann nennen wir ihn asymptotisch effizient.

Kapitel drei enthält sowohl die Rückführung auf die Nichtparametrik als auch eine nähere Betrachtung der Funktionale. Dabei folgen wir den Arbeiten von Pfanzagl, Wefelmeyer [30], [31], Strasser [35] sowie Van der Vaart [36] bzw. [37], Bickel, Klaasen, Ritov, Wellner (von nun an abgekürzt durch BKRW) [7] und Janssen [15], [16], [19]. Eine gute Einführung in das Thema und eine hilfreiche Zusammenfassung der gerade für diese Arbeit wichtigen Aspekte bietet auch die Dissertation von Vladimir Ostrovski [29].

Ziel ist es hier, den Faltungsatz aus Kapitel zwei auf das nichtparametrische Modell zu erweitern. Zunächst erweitern wir die Differenzierbarkeit des Funktionals κ auf den nichtparametrischen Fall, indem wir die Differenzierbarkeit von $\kappa : \Theta \to \mathbb{R}$ entlang einparametrischer Kurven $\{P_\vartheta : \vartheta \in \Theta\}$ fordern. Die Ableitung des Funktionals ist dann darstellbar als inneres Produkt aus der jeweiligen Tangente und dem eindeutig bestimmten kanonischen Gradienten $\tilde{\kappa}$ des Funktionals, das heißt

$$\lim_{\vartheta \to 0} \frac{\kappa(P_\vartheta) - \kappa(P_0)}{\vartheta} = \int \tilde{\kappa} g \, dP_0.$$

Eine zentrale Bedeutung gewinnt hier der Begriff der dichten Differenzierbarkeit. Ein Funktional heißt dicht differenzierbar, wenn eine dichte Teilmenge des Tangentialraums existiert, so dass es zu jeder Tangente in dieser Teilmenge eine L_2-differenzierbare Kurve lokal in $\mathcal{P}$ gibt, entlang derer das Funktional differenzierbar ist. Auch in diesem Fall ist $\tilde{\kappa}$ eindeutig bestimmt und wir können zeigen, dass die dichte Differenzierbarkeit bereits ausreicht, um eine maximale untere Schranke für $Var(Y)$ zu erhalten. Diese ergibt sich dann aus der quadratischen Norm des kanonischen Gradienten $\|\tilde{\kappa}\|^2$. Gilt für einen Schätzer T_n eines Funktionals κ im nichtparametrischen Modell also die Verteilungskonvergenz

$$\sqrt{n}(T_n - \kappa(P_0)) \xrightarrow[n\to\infty]{\mathcal{D}} Y \sim \mathcal{N}(0, \|\tilde{\kappa}\|^2),$$

so ist er asymptotisch effizient. Setzen wir die Voraussetzungen noch weiter herab und fordern nur eine Differenzierbarkeit des Funktionals bezüglich einer beliebigen Teilmenge des Tangentialraums, so erhalten wir immer noch eine untere Schranke für $Var(Y)$. Nimmt ein Schätzer T_n (asymptotisch, zentriert und normiert) diese an, so ist der Schätzer bereits asymptotisch effizient und es gibt offensichtlich keine größere Schranke.

Zudem beleuchten wir in diesem Kapitel den Zusammenhang von asymptotischer Effizienz und asymptotischer Linearität. Ein Schätzer T_n heißt asymptotisch linear in χ, wenn er die Darstellung

$$\sqrt{n}(T_n - \kappa(P_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \chi(X_i) + R_n$$

für eine Einflussfunktion $\chi \in L_2(P_0)$ und einen asymptotisch vernachlässigbaren Restterm R_n besitzt. Ist der Schätzer asymptotisch linear im kanonischen Gradienten $\tilde{\kappa}$ des Funktionals, so ist er asymptotisch effizient. Umgekehrt impliziert die asymptotische Effizienz des Schätzers eine solche Entwicklung von T_n.

In den beiden folgenden Kapiteln wird die Theorie aus den ersten drei Kapiteln auf spezielle Situationen angewendet. In Kapitel vier beschäftigen wir uns mit der Klasse der L-Funktionale. An dieser Stelle sei auch die wertvolle Vorarbeit von Felix von Müller erwähnt, die er in seiner Diplomarbeit [38] auf diesem Gebiet geleistet hat. Für ein Wahrscheinlichkeitsmaß P mit der inversen Verteilungsfunktion F^{-1} sind L-Funktionale Funktionale der Form

$$\kappa(P) = \int h(F^{-1}(u))d\nu(u)$$

zur Bewertungsfunktion h und Gewichtungsmaß ν. Wir besprechen zunächst gewisse Eigenschaften dieser Funktionale, betrachten einige Beispiele und ihre kanonischen Schätzer. Ist die Gewichtung ν ein diskretes Maß, so sprechen wir vom diskreten L-Funktional. Da der diskrete Fall in der Literatur bereits ausführlich behandelt ist, vgl. beispielsweise ausführlich von Müller [38], S.46 ff., und die Theorie, insbesondere die neuen Ansätze aus Kapitel drei hier kaum oder nur triviale Anwendung finden, legen wir besonderes Augenmerk auf den stetigen Fall mit stetiger Gewichtung ν. Hier können wir mit Hilfe der Theorie aus Kapitel drei die dichte Differenzierbarkeit des Funktionals auf speziellen (ausreichend großen) Familien von Wahrscheinlichkeitsmaßen unter gewissen Voraussetzungen zeigen und somit die optimale Grenzverteilung eines Schätzers im Sinne der asymptotischen Effizienz ermitteln. Da die kanonischen Schätzer der L-Funktionale dieses Kriterium erfüllen, haben wir in diesem Fall den Nachweis der asymptotischen Effizienz dieser Schätzer erbracht.

In Kapitel fünf widmen wir uns Funktionalen im Rahmen des Survival Analysis. Nach einer kurzen Einführung in die wichtigsten Begriffe und Zusammenfassung relevanter bekannter Resultate, leiten wir auch ein Modell für rechtszensierte Daten her, vgl. hierzu ausführlich beispielsweise Janssen, Werft [20], und führen die Standard-Schätzer für Survivalfunktion und kumulative Hazardfunktion, den Kaplan-Meier- und Nelson-Aalen-Schätzer, ein. Die Hazard-Funktionale, mit denen wir uns anschließend befassen, haben die Form

$$\kappa(P) := \int h \, d\Lambda = \int \frac{h}{1-F} \, dP$$

für eine messbare reellwertige Funktion h, wobei F die zu P zugehörige Verteilungsfunktion bezeichne. Da diese Funktionale über das Hazardmaß Λ definiert sind, beweisen wir einen Satz, der für eine gegebene Tangente g eine L_2-differenzierbare Kurve mit Tangente g liefert, die durch einen Dichtequotienten des Hazardmaßes gegeben ist und in dieser Form eine im Folgenden einfach zu behandelnde Struktur hat. Mit Hilfe dieser Kurven können wir eine Differenzierbarkeit des Hazard-Funktionals sowohl im unzensierten als auch im rechtszensierten Fall zeigen und erhalten somit in beiden Fällen eine untere Schranke für die Varianz der Grenzverteilung eines regulären Schätzers für dieses Funktional.
Ein Spezialfall des Hazard-Funktionals ist die kumulative Hazardfunktion. Ihr Standard-Schätzer, der Nelson-Aalen-Schätzer, nimmt die so hergeleitete Schranke an, wodurch wir auf diesem Wege erneut die asymptotische Effizienz des Nelson-Aalen-

Schätzers beweisen können. Auch die asymptotische Effizienz des Kaplan-Meier-Schätzers für die Survivalfunktion können wir in diesem Zuge mit Hilfe einer Kettenregel erneut verifizieren, da die Survivalfunktion sich als Funktion der kumulativen Hazardfunktion darstellen lässt.

Kapitel 1

Grundlagen

1.1 Motivation

In dieser Arbeit behandeln wir Funktionale $\kappa : \mathcal{P} \to \mathbb{R}$ auf nichtparametrischen Familien von Wahrscheinlichkeitsmaßen $\mathcal{P}$. Da solche Familien die parametrischen Modelle als Teilmodelle enthalten, führt uns dies für einen beliebigen Fußpunkt $P_0 \in \mathcal{P}$ zurück auf die Untersuchung parametrischer Familien von Wahrscheinlichkeitsmaßen $\mathcal{P}_1 = \{P_\vartheta : \vartheta \in \Theta\}$ mit $\Theta \subset \mathbb{R}, 0 \in \mathring{\Theta}$, sogenannten Kurven (lokal) in $\mathcal{P}$, die durch P_0 verlaufen.
Wir wollen uns nun damit beschäftigen, asymptotisch beste Schätzer für κ im Sinne einer asymptotisch möglichst kleinen Streuung um das zu schätzende Funktional zu finden. Für eine solche parametrische Familie von Wahrscheinlichkeitsmaßen $\mathcal{P}_1 \subset \mathcal{P}$ liefert die Cramér-Rao-Ungleichung

$$\frac{\kappa'(P_0)^2}{I_0} \leq Var_{P_0}(T)$$

unter gewissen Voraussetzungen eine untere Schranke für die Varianz eines erwartungstreuen Schätzers $T : \Omega \to \mathbb{R}$ in P_0, wobei κ' die Ableitung der Abbildung $\vartheta \mapsto \kappa(P_\vartheta)$ und I_0 die Fisher-Information in P_0 bezeichnet. Ist $T_n : \Omega^n \to \mathbb{R}$ eine erwartungstreue Schätzfolge zum Stichprobenumfang n, so hat die Cramér-Rao-Ungleichung die Form

$$\frac{\kappa'(P_0)^2}{I_0} \leq Var_{P_0^n}(\sqrt{n}T_n).$$

Eine Schätzfolge T_n heißt daher Fisher-effizient in P_0, falls die Varianz $Var_{P_0^n}(\sqrt{n}T_n)$ für $n \to \infty$ gegen diese Schranke konvergiert.

Da im nichtparametrischen Modell $\mathcal{P}$ mit Sicherheit nicht effizienter geschätzt werden kann als in jedem einzelnen parametrischen Teilmodell, ist eine Schranke für den nichtparametrischen Fall damit durch das Supremum über die Cramér-Rao-Schranken einer möglichst großen Anzahl geeigneter Teilmodelle motiviert. Der Begriff der asymptotischen Effizienz, um den sich diese Arbeit dreht, wird allerdings weit mehr enthalten als lediglich eine asymptotische Cramér-Rao-Schranke. Asymptotisch effizient wird ein Schätzer sein, wenn er eine optimale Grenzverteilung besitzt. Das prinzipielle Vorgehen bleibt jedoch dasselbe: Wir betrachten Kriterien für optimale Schätzer in parametrischen Teilmodellen und führen diese zusammen zu einem Kriterium für einen optimalen (asymptotisch effizienten) Schätzer im nichtparametrischen Modell $\mathcal{P}$.
Im parametrischen Fall setzt man klassischerweise die L_1-Differenzierbarkeit des Modells voraus. Im nichtparametrischen Fall beschränken wir uns auf die Analyse der von Le Cam eingeführten L_2-differenzierbaren Kurven, die in diesem Kapitel definiert werden. Solche Kurven sind ausgezeichnet durch eine Tangente $g \in L_2(P_0)$ und viele Eigenschaften hängen nicht von der genauen Gestalt der Kurve, sondern nur von der Tangente g ab. Somit erhalten wir eine Einbettung unserer Modelle oder genauer ihrer Tangenten in den Hilbertraum $L_2(P_0)$ und können seine Struktur nutzen, um die Gestalt des nichtparametrischen Modells $\mathcal{P}$ zu verstehen.

1.2 L_2-Differenzierbare Kurven und Tangentialräume

Sei $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen und $P_{\vartheta_0} \in \mathcal{P}$. Sei $\Theta \subset \mathbb{R}$. Wir sagen, dass eine Kurve $\vartheta \mapsto P_\vartheta$ lokal in $\mathcal{P}$ liegt, falls es ein $\varepsilon > 0$ gibt, so dass $P_{\vartheta_0+t} \in \mathcal{P}$ für alle $|t| < \varepsilon$ gilt. Die Kurve heißt $L_2(P_{\vartheta_0})$-differenzierbar, wenn simpel formuliert der Differenzenquotient der Wurzel des Dichtequotienten von P_ϑ bezüglich einem dominierenden Maß in L_2 konvergiert.

Definition 1.1 (L_2-Differenzierbarkeit)
Sei $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen. Eine durch μ dominierte Kurve $\vartheta \mapsto P_\vartheta$ in $\mathcal{P}$ heißt L_2-differenzierbar in ϑ_0 (kurz $L_2(P_{\vartheta_0})$-differenzierbar) mit Tangente (oder auch score function) $g \in L_2(P_{\vartheta_0})$, falls gilt:

$$\left\| \frac{2}{t} \left(\left(\frac{dP_{\vartheta_0+t}}{d\mu} \right)^{\frac{1}{2}} - \left(\frac{dP_{\vartheta_0}}{d\mu} \right)^{\frac{1}{2}} \right) - g \left(\frac{dP_{\vartheta_0}}{d\mu} \right)^{\frac{1}{2}} \right\|_{L_2(\mu)} \xrightarrow[t \to 0]{} 0.$$

Lemma 1.2 (Konsequenzen)

a) g ist P_{ϑ_0}-f.s. eindeutig.

b) Aus der L_2-Differenzierbarkeit in P_{ϑ_0} folgt die L_1-Differenzierbarkeit in P_{ϑ_0}, also

$$\left\| \frac{1}{t}\left(\frac{dP_{\vartheta_0+t}}{d\mu} - \frac{dP_{\vartheta_0}}{d\mu}\right) - g\frac{dP_{\vartheta_0}}{d\mu}\right\|_{L_1(\mu)} \xrightarrow[t\to 0]{} 0$$

und die L_1-Stetigkeit in P_{ϑ_0}, das heißt

$$\left\| \frac{dP_{\vartheta_0+t}}{d\mu} - \frac{dP_{\vartheta_0}}{d\mu}\right\|_{L_1(\mu)} \xrightarrow[t\to 0]{} 0.$$

Insbesondere folgt hieraus die schwache Konvergenz

$$P_{\vartheta_0+t} \xrightarrow[t\to 0]{} P_{\vartheta_0}.$$

c) Sei $f : \Omega \to \mathbb{R}$ messbar und beschränkt. Dann gilt im Falle der $L_1(P_{\vartheta_0})$-Differenzierbarkeit

$$\frac{d}{d\vartheta}\int f\, dP_{\vartheta|\vartheta=\vartheta_0} = \int fg\, dP_{\vartheta_0}.$$

d) Für eine Tangente g gilt

$$\int g\, dP_{\vartheta_0} = 0.$$

Daher wird der Raum

$$L_2^0(P_{\vartheta_0}) := \{g \in L_2 : \int g\, dP_{\vartheta_0} = 0\}$$

als Tangentenraum bezeichnet.

e) Für $x \in \mathbb{R}$ gilt

$$P_{\vartheta_0+t}((-\infty, x]) = P_{\vartheta_0}((-\infty, x]) + t \int\limits_{(-\infty,x]} g\, dP_{\vartheta_0} + o(t)$$

gleichmäßig in x.

Beweis.

a) Aus der L_2-Konvergenz folgt die stochastische Konvergenz und die Grenzfunktion ist P-f.s. eindeutig.

b) Zum Beweis der L_1-Differenzierbarkeit siehe Witting [41], S.175, Satz 1.190. Die L_1-Stetigkeit ist dann trivialerweise notwendig. Sei nun $f : \mathbb{R} \to \mathbb{R}$ stetig und beschränkt mit $|f| \le c \in \mathbb{R}$. Dann folgt

$$\begin{aligned} &\left| \int f \, dP_{\vartheta_0+t} - \int f \, dP_{\vartheta_0} \right| \le \int |f| \left| \frac{dP_{\vartheta_0+t}}{d\mu} - \frac{dP_{\vartheta_0}}{d\mu} \right| d\mu \\ &\le c \int \left| \frac{dP_{\vartheta_0+t}}{d\mu} - \frac{dP_{\vartheta_0}}{d\mu} \right| d\mu \xrightarrow[t\to 0]{} 0 \end{aligned}$$

und damit die schwache Konvergenz.

c) Sei $|f| \le c$. Dann gilt

$$\begin{aligned} &\left| \frac{1}{t} \left(\int f \, dP_{\vartheta_0+t} - \int f \, dP_{\vartheta_0} \right) - \int fg \, dP_{\vartheta_0} \right| \\ &\le \int |f| \left| \left(\frac{1}{t} \left(\frac{dP_{\vartheta_0+t}}{d\mu} - \frac{dP_{\vartheta_0}}{d\mu} \right) - g \frac{dP_{\vartheta_0}}{d\mu} \right) \right| d\mu \\ &\le c \left\| \frac{1}{t} \left(\frac{dP_{\vartheta_0+t}}{d\mu} - \frac{dP_{\vartheta_0}}{d\mu} \right) - g \frac{dP_{\vartheta_0}}{d\mu} \right\|_{L_1(\mu)} \xrightarrow[t\to 0]{} 0. \end{aligned}$$

d) Aus c) folgt

$$0 = \frac{d}{d\vartheta} 1_{|\vartheta=\vartheta_0} = \frac{d}{d\vartheta} \int 1 \, dP_{\vartheta|\vartheta=\vartheta_0} = \int g \, dP_{\vartheta_0}.$$

e) Setze in c) $f = 1_{(-\infty,x]}$. Dann folgt

$$\begin{aligned} &\frac{1}{t} \left(\int\limits_{(-\infty,x]} dP_{\vartheta_0+t} - \int\limits_{(-\infty,x]} dP_{\vartheta_0} \right) = \int\limits_{(-\infty,x]} g \, dP_{\vartheta_0} + o(1) \\ \Leftrightarrow &\frac{1}{t} \int\limits_{(-\infty,x]} dP_{\vartheta_0+t} = \frac{1}{t} \int\limits_{(-\infty,x]} dP_{\vartheta_0} + \int\limits_{(-\infty,x]} g \, dP_{\vartheta_0} + o(1) \\ \Leftrightarrow &\int\limits_{(-\infty,x]} dP_{\vartheta_0+t} = \int\limits_{(-\infty,x]} dP_{\vartheta_0} + t \int\limits_{(-\infty,x]} g \, dP_{\vartheta_0} + o(t). \end{aligned}$$

Die Gleichmäßigkeit folgt sofort aus der Gleichmäßigkeit der Abschätzung in c).

■

Bemerkung 1.3 (praktische Berechnung der score function)

Da die Grenzfunktion bei fast sicherer bzw. L_p-Konvergenz fast sicher eindeutig ist und die Grenzfunktionen somit fast sicher übereinstimmen, folgt unter L_1- bzw. L_2-Differenzierbarkeit mit score function g aus einer angenommenen punktweisen Konvergenz

$$\frac{1}{t}\left(\frac{dP_{\vartheta_0+t}}{d\mu}-\frac{dP_{\vartheta_0}}{d\mu}\right)\xrightarrow[t\to 0]{}\tilde{g}\frac{dP_{\vartheta_0}}{d\mu},$$

dass $\tilde{g}$ bereits die score function g ist. Also ist $g\frac{dP_{\vartheta_0}}{d\mu}$ die formale Ableitung des Dichtequotienten $\frac{dP_\vartheta}{d\mu}$ in $\vartheta=\vartheta_0$ und g lässt sich für $\frac{dP_{\vartheta_0}}{d\mu}(x)>0$ durch

$$\frac{d}{d\vartheta}\log\frac{dP_\vartheta}{d\mu}(x)_{|\vartheta=\vartheta_0}=\left(\frac{dP_\vartheta}{d\mu}(x)\right)^{-1}_{|\vartheta=\vartheta_0}\cdot g(x)\frac{dP_{\vartheta_0}}{d\mu}(x)=g(x)$$

berechnen, falls der Dichtequotient differenzierbar ist. Im Sinne der P_{ϑ_0}-stochastischen Konvergenz existiert die Ableitung unter L_1- bzw. L_2-Differenzierbarkeit immer, da aus der L_p-Konvergenz die stochastische Konvergenz folgt.

Da wir in dieser Arbeit Eigenschaften von $\mathcal{P}$ für alle Fußpunkte $P\in\mathcal{P}$ untersuchen wollen, ersetzen wir ϑ_0 von nun an durch 0 und betrachten beliebige, aber feste Fußpunkte $P_0\in\mathcal{P}$, was uns die Notation erleichtert.

Satz 1.4 (lineare Transformation, vgl. Ostrovski [29], S.7, Satz 1.8)
Sei $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen und $\vartheta\mapsto P_\vartheta$ eine durch μ dominierte $L_2(P_0)$-differenzierbare Kurve in $\mathcal{P}$ mit Tangente g. Dann ist die Kurve $\vartheta\mapsto P_{\lambda\vartheta}$ für alle $\lambda\in\mathbb{R}$ ebenfalls $L_2(P_0)$-differenzierbar mit der Tangente $\lambda g\in L_2^0(P_0)$.

Beweis. Für $\lambda=0$ ist $\vartheta\mapsto P_0$ offensichtlich $L_2(P_0)$-differenzierbar mit der Tangente $\lambda g=0\in L_2^0(P_0)$.
Sei nun $\lambda\neq 0$. Dann gilt mit $s:=\lambda t$

$$\begin{aligned}&\lim_{t\to 0}\left\|\frac{2}{t}\left(\left(\frac{dP_{\lambda t}}{d\mu}\right)^{\frac12}-\left(\frac{dP_0}{d\mu}\right)^{\frac12}\right)-\lambda g\left(\frac{dP_0}{d\mu}\right)^{\frac12}\right\|_{L_2(\mu)}\\
=&\lim_{t\to 0}|\lambda|\left\|\frac{2}{\lambda t}\left(\left(\frac{dP_{\lambda t}}{d\mu}\right)^{\frac12}-\left(\frac{dP_0}{d\mu}\right)^{\frac12}\right)-g\left(\frac{dP_0}{d\mu}\right)^{\frac12}\right\|_{L_2(\mu)}\\
=&\lim_{s\to 0}|\lambda|\left\|\frac{2}{s}\left(\left(\frac{dP_s}{d\mu}\right)^{\frac12}-\left(\frac{dP_0}{d\mu}\right)^{\frac12}\right)-g\left(\frac{dP_0}{d\mu}\right)^{\frac12}\right\|_{L_2(\mu)}=0.\end{aligned}$$

■

Definition 1.5 (Tangentenkegel und Tangentialraum)
Sei $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen und $P_0 \in \mathcal{P}$. Die Menge aller Tangenten an P_0 in $\mathcal{P}$ im Sinne von Definition 1.1 wird definiert als

$$K(P_0, \mathcal{P}) := \{g \in L_2^0(P_0) : \exists\, L_2(P_0)\text{-differenzierbare Kurve in } \mathcal{P} \text{ mit Tangente } g\}.$$

Aus Satz 1.4 folgt, dass $K(P_0, \mathcal{P})$ ein Kegel ist. Daher wird $K(P_0, \mathcal{P})$ als Tangentenkegel bezeichnet. Der $L_2(P_0)$-Abschluss der linearen Hülle des Tangentenkegels wird als Tangentialraum bezeichnet und mit

$$T(P_0, \mathcal{P}) := \overline{lin\, K(P_0, \mathcal{P})}$$

notiert.

Bemerkung 1.6
Der Raum $L_2(P_0)$ ist ein Hilbertraum mit dem Skalarprodukt $\langle f, g\rangle := \int fg\, dP_0$, vgl. etwa Bauer [4], S.99, Bemerkung 1, und nach Lemma A.9 ist jeder abgeschlossene Unterraum eines Hilbertraums selbst ein Hilbertraum. $T(P_0, \mathcal{P})$ ist per Definition abgeschlossen und somit ebenfalls ein Hilbertraum. Seien $g_n \in L_2^0(P_0)$ mit

$$g_n \xrightarrow[n\to\infty]{} g \text{ in } L_2(P_0).$$

Dann folgt

$$\|g_n\|_{L_2(P_0)} \xrightarrow[n\to\infty]{} \|g\|_{L_2(P_0)},$$

also $g \in L_2^0(P_0)$ und $L_2^0(P_0)$ ist abgeschlossen. Also ist auch $L_2^0(P_0)$ ein Hilbertraum.

1.3 Das volle Modell

Sei $\mathcal{M}(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ die Menge aller Verteilungen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
Die Sätze, Lemmata etc. sind in dieser Arbeit weitestgehend allgemein gehalten für allgemeine Wahrscheinlichkeitsräume (Ω, P). Mit dem vollen Modell bezeichnen wir jedoch stets folgende Teilmenge von $\mathcal{M}(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Definiere

$$\mathcal{P}^* := \left\{ P \in \mathcal{M}(\mathbb{R}, \mathcal{B}(\mathbb{R})) \;\middle|\; \begin{array}{l} P \text{ besitzt eine Verteilungsfunktion } F \in \mathcal{C}^1 \\ \text{mit Dichte } f = F' > 0\ \lambda\text{-f.ü. auf dem} \\ \text{zusammenhängenden Träger von } P \end{array} \right\}. \quad (1.1)$$

Bemerkung 1.7

Da $\mathcal{P}^*$ nur stetige Verteilungen mit streng monoton steigenden Verteilungsfunktionen F umfasst, ist die inverse Verteilungsfunktion F^{-1}, vgl. Definition A.1, die Umkehrabbildung von F, das heißt es gilt $F \circ F^{-1} = F^{-1} \circ F = id$. Zudem ist F^{-1} gemäß Lemma A.4 stetig und differenzierbar.

Das volle Modell umfasst also nicht alle Wahrscheinlichkeitsmaße auf $\mathbb{R}$, es ist aber vollständig in dem Sinne, dass der Tangentialraum maximal ist.

Lemma 1.8

Es gilt $T(P_0, \mathcal{P}^*) = L_2^0(P_0)$.

Für den Beweis benötigen wir zuvor noch ein weiteres Lemma.

Lemma 1.9

Sei $\mathcal{P}$ *eine Familie von Wahrscheinlichkeitsmaßen und* $P_0 \in \mathcal{P}$. *Die Menge*

$$D^* = D^*(\mathcal{P}) := \{g \in L_2^0(P_0) : \exists\, c \in \mathbb{R} \textit{ mit } |g| \leq c\}$$

der beschränkten Funktionen in $L_2(P_0)$ *mit* $E_{P_0}(g) = 0$ *liegt dicht in* $L_2^0(P_0)$.

Beweis. Sei $g \in L_2^0(P_0)$ und $g_n := g1_{\{|g|<n\}}$ für $n \in \mathbb{N}$ und $h_n := g_n - \int g1_{\{|g|<n\}}\, dP_0$. Dann ist h_n eine Folge in D^*, denn für alle n gilt $\int h_n\, dP_0 = 0$, also $h_n \in L_2^0(P_0)$, und h_n ist beschränkt.

Ferner gilt $g_n \xrightarrow[n\to\infty]{} g$ P_0-f.s. und nach dem Satz von der dominierten Konvergenz $\int g1_{\{|g|<n\}}\, dP_0 \xrightarrow[n\to\infty]{} \int g\, dP_0 = 0$ wegen $|g1_{\{|g|<n\}}| \leq |g|$ und $g \in L_1(P_0)$. Also gilt $h_n \xrightarrow[n\to\infty]{} g$ P_0-f.s. und

$$|h_n| = \left| g1_{\{|g|<n\}} - \int g1_{\{|g|<n\}}\, dP_0 \right| \leq |g| - \int |g| dP_0 \in L_2(P_0).$$

Mit dem Satz von der dominierten Konvergenz folgt $h_n \to g$ in $L_2(P_0)$. ∎

Beweis. (von Lemma 1.8)

Zur Beweisidee vgl. BKRW [7], S.52.

$T(P_0, \mathcal{P}^*) \subset L_2^0(P_0)$ ist trivial, da $L_2^0(P_0)$ abgeschlossen ist.

Zeige jetzt $D^* \subset T(P_0, \mathcal{P}^*)$. Denn D^* liegt nach Lemma 1.9 dicht in $L_2^0(P_0)$ und da $T(P_0, \mathcal{P}^*)$ per Definition abgeschlossen ist, folgt dann $T(P_0, \mathcal{P}^*) = L_2^0(P_0)$. Sei also $g \in D^*$, $|g| \leq c$. Konstruiere nun eine Exponentialfamilie in $\mathcal{P}^*$ mit score function

(Tangente) g. Dazu definiere für $|\vartheta| \leq \delta$

$$f_\vartheta = \frac{dP_\vartheta}{d\lambda\!\!\!\lambda} := \exp(\vartheta g) C(\vartheta) f_0 \text{ mit}$$
$$C(\vartheta) := \left(\int \exp(\vartheta g)\, dP_0 \right)^{-1}.$$

Nach Konstruktion gilt $\int f_\vartheta(x)\, dx = 1$ und $f_\vartheta > 0$, also ist durch diese Exponentialfamilie eine Kurve in $\mathcal{P}^*$ gegeben. Nach BKRW [7], S.14, Beispiel 1, ist eine Exponentialfamilie $L_2(P_0)$-differenzierbar. Nun gilt wegen

$$\left| \frac{d}{d\vartheta} \exp(\vartheta g) \right| = |g \exp(\vartheta g)| \leq c \exp(\delta c)$$

mit Lemma A.12

$$\begin{aligned}
&\frac{d}{d\vartheta} C(\vartheta) = \frac{d}{d\vartheta} \left(\int \exp(\vartheta g)\, dP_0 \right)^{-1} \\
&= - \left(\int \exp(\vartheta g)\, dP_0 \right)^{-2} \frac{d}{d\vartheta} \int \exp(\vartheta g)\, dP_0 = -C(\vartheta)^2 \int \frac{d}{d\vartheta} \exp(\vartheta g)\, dP_0 \\
&= -C(\vartheta)^2 \int g \exp(\vartheta g)\, dP_0 = -C(\vartheta)^2 C(\vartheta)^{-1} \int g \exp(\vartheta g) C(\vartheta)\, dP_0 \\
&= -C(\vartheta) \int g\, dP_\vartheta = -C(\vartheta) E_{P_\vartheta}(g).
\end{aligned}$$

Gemäß Bemerkung 1.3 ergibt sich als score function

$$\begin{aligned}
\frac{d}{d\vartheta} (\log f_\vartheta(x))_{|\vartheta=0} &= \frac{d}{d\vartheta} (\vartheta g(x) + \log C(\vartheta) + \log f_0(x))_{|\vartheta=0} \\
&= \left(g(x) + \frac{\frac{d}{d\vartheta} C(\vartheta)}{C(\vartheta)} \right)_{|\vartheta=0} = g(x) + \left(\frac{-C(\vartheta) E_{P_\vartheta}(g)}{C(\vartheta)} \right)_{|\vartheta=0} \\
&= g(x) - E_{P_0}(g) = g(x).
\end{aligned}$$

Somit ist eine $L_2(P_0)$-differenzierbare Kurve in $\mathcal{P}^*$ mit Tangente g gefunden. Also gilt $g \in K(P_0, \mathcal{P}^*) \subset T(P_0, \mathcal{P}^*)$ und die Behauptung ist gezeigt. ■

Bemerkung 1.10

a) Aus dem Beweis folgt, dass $K(P_0, \mathcal{P}^*)$ dicht in $T(P_0, \mathcal{P}^*)$ liegt, $K(P_0, \mathcal{P}^*)$ also bereits linear abgeschlossen ist.

b) Für das volle Modell $\mathcal{P} = \mathcal{P}^*$ ist D^* also eine dichte Teilmenge des Tangentialraums $T(P_0, \mathcal{P}^*)$.

Letztere Bemerkung sei noch einmal hervorgehoben. Da wir im weiteren Verlauf der Arbeit bestimmte Eigenschaften nur für eine solche dichte Teilmenge nachweisen müssen, wie wir in Kapitel 2.2 sehen werden, wird es je nach vorliegender Problemstellung einfacher, je kleiner diese dichte Teilmenge gewählt wird. Aus diesem Grund bestimmen wir noch eine weitere, kleinere dichte Teilmenge $D^{**} \subset T(P_0, \mathcal{P}^*)$.

Lemma 1.11
Sei $\mathcal{P}^*$ das volle Modell (1.1). $P_0 \in \mathcal{P}^*$ besitze den zusammenhängenden Träger $\Omega_0 \subset \mathbb{R}$. Dann liegt die Menge

$$D^{**} := D^*(\mathcal{P}^*) \cap \left\{ g \in L_2^0(P_0) \;\middle|\; \begin{array}{l} \exists A \subsetneq \Omega_0, A = [a_1, a_2] \text{ kompakt mit} \\ a_1, a_2 \in \mathring{\Omega}_0 \text{ und } g(x) = 0 \;\forall x \in A^c \end{array} \right\}.$$

der beschränkten Tangenten mit kompaktem Träger, der vollständig im Inneren des Trägers von P_0 liegt, dicht in $L_2^0(P_0) = T(P_0, \mathcal{P}^*)$.

Beweis. Der Beweis verläuft analog zum Beweis von Lemma 1.9.
Sei $g \in L_2^0(P_0)$. Da $\Omega_0 \subset \mathbb{R}$ zusammenhängend ist, gibt es $a, b \in \mathbb{R}$ mit $\Omega_0 = [a, b]$, $\Omega_0 = (-\infty, b]$, $\Omega_0 = [a, \infty)$ oder $\Omega_0 = \mathbb{R}$.
Je nach Gestalt von Ω_0 definieren wir Folgen $a_n, b_n \in \mathbb{R}, n \in \mathbb{N}$. Ist Ω_0 nach unten offen, so setze $a_n := -n$, und ist Ω_0 nach unten offen, so setze $b_n := n$. Andernfalls setze $a_n := a + \frac{1}{n}$ bzw. $b_n := b - \frac{1}{n}$.
Definiere nun $A_n := [a_n, b_n]$ und

$$g_n := g 1_{A_n} 1_{\{|g|<n\}} - \int\limits_{A_n} g 1_{\{|g|<n\}} \, dP_0.$$

Dann ist g_n eine Folge in D^{**} mit $g_n \xrightarrow[n\to\infty]{} g$ P_0-f.s. und es gilt

$$|g_n| = \left| g 1_{A_n} 1_{\{|g|<n\}} - \int\limits_{A_n} g 1_{\{|g|<n\}} \, dP_0 \right| \leq |g| + \int |g| dP_0 \in L_2(P_0).$$

Nach dem Satz von der dominierten Konvergenz folgt $\|g_n - g\|_{L_2(P_0)} \xrightarrow[n\to\infty]{} 0$. ■

Bemerkung 1.12
Da für den kompakten Träger $A = [a_1, a_2]$ von $g \in D^{**}$ gilt $a_1, a_2 \in \mathring{\Omega}_0$, gilt $0 < F_0(a_1) \leq F_0(x) \leq F_0(a_2) < 1$ für alle $x \in A$.

Wir konstruieren nun für das volle Modell $\mathcal{P} = \mathcal{P}^*$ noch zu jedem $g \in D^*$ (und damit auch zu jedem $g \in D^{**} \subset D^*(\mathcal{P}^*)$) eine L_2-differenzierbare Kurve, die g gerade als Tangente besitzt.

Lemma 1.13 (vgl. Strasser [35], S.385)
Zu $g \in D^*$ gibt es ein $\delta > 0$, so dass für $|\vartheta| < \delta$ eine $L_2(P_\vartheta)$-differenzierbare Kurve lokal in $\mathcal{P}^*$ mit Tangente $g_\vartheta := \frac{g}{1+\vartheta g}$ gegeben ist durch $f_\vartheta := \frac{dP_\vartheta}{d\lambda\!\!\!\lambda} = (1+\vartheta g) f_0$ mit $f_0 := \frac{dP_0}{d\lambda\!\!\!\lambda}$. Insbesondere ist die Kurve $L_2(P_0)$-differenzierbar mit Tangente $g_0 = g$.

Beweis. Sei $|g| \leq c$. Nun gibt es ein $\delta > 0$, so dass $1 + \vartheta g > 0$ und somit $f_\vartheta > 0$ für $|\vartheta| < \delta$ gilt. Weiterhin gilt

$$\int f_\vartheta(x)\, dx = \int 1 + \vartheta g\, dP_0 = 1 + \vartheta \int g\, dP_0 = 1$$

und f_ϑ ist eine Wahrscheinlichkeitsdichte und $\vartheta \mapsto P_\vartheta$ eine Kurve in $\mathcal{P}^*$.
Nun gilt

$$\frac{d}{d\vartheta} 2\sqrt{f_\vartheta} = \frac{d}{d\vartheta} 2\sqrt{(1+\vartheta g) f_\vartheta} = \frac{g}{\sqrt{1+\vartheta g}} \sqrt{f_\vartheta} = g_\vartheta \sqrt{f_\vartheta},$$

also punktweise Konvergenz des Differenzenquotienten. Weiter gilt nach Lemma A.13

$$\left|\sqrt{y+z} - \sqrt{y}\right| \leq |z|$$

für $y + z \geq \frac{1}{4}$ und $y \geq \frac{1}{4}$. Daher gilt mit $|g| \leq c$

$$\begin{aligned} &\left| \frac{2}{t} \left(\sqrt{f_{\vartheta+t}} - \sqrt{f_\vartheta} \right) \right| = \left| \frac{2}{t} \left(\sqrt{1 + (\vartheta + t) g} - \sqrt{1 + \vartheta g} \right) \sqrt{f_0} \right| \\ &\leq \left| \frac{2}{t} t g \sqrt{f_0} \right| \leq 2c\sqrt{f_0} \in L_2(\lambda\!\!\!\lambda) \end{aligned}$$

für t und δ ausreichend klein und $|\vartheta| \leq \delta$. Hieraus folgt nach dem Satz von der dominierten Konvergenz die L_2-Konvergenz, also

$$\lim_{t \to 0} \left\| \frac{2}{t} \left(\left(\frac{dP_{\vartheta+t}}{d\lambda\!\!\!\lambda} \right)^{\frac{1}{2}} - \left(\frac{dP_\vartheta}{d\lambda\!\!\!\lambda} \right)^{\frac{1}{2}} \right) - g_\vartheta \left(\frac{dP_\vartheta}{d\lambda\!\!\!\lambda} \right)^{\frac{1}{2}} \right\|_{L_2(\lambda\!\!\!\lambda)} = 0$$

und somit die L_2-Differenzierbarkeit in P_ϑ mit Tangente g_ϑ. ∎

Kapitel 2

Lokale asymptotische Normalität und der Faltungssatz

In Abschnitt 2.1 sollen die wichtigsten Resultate der LAN-Theorie (lokale asymptotische Normalität) von Le Cam kurz zusammengefasst werden. Dabei werden einige bekannte Sätze und Definitionen an unsere Zwecke angepasst und teilweise für Spezialfälle vereinfacht. Den Grundstein für die hier behandelte Theorie legte Le Cam bereits 1960 in seinen Arbeiten [26], die er 1986 [27] noch einmal zusammenfasste. In Abschnitt 2.2 führen wir den Faltungssatz von Hájek-Le Cam [13] ein, der unser zentrales Hilfsmittel darstellt, um ein Kriterium für die asymptotische Effizienz eines Schätzers herzuleiten.

2.1 Lokale asymptotische Normalität

In dieser Arbeit beschäftigen wir uns mit dem Schätzen von Funktionalen in nichtparametrischen Familien von Wahrscheinlichkeitsmaßen $\mathcal{P}$. Diese Untersuchung führt uns zurück auf die parametrischen Teilmodelle. Wir betrachten daher Zufallsvariablen $X_1, ..., X_n$, die unabhängig identisch nach $P_\vartheta \in \mathcal{P}_1 = \{P_\vartheta : \vartheta \in \Theta\}$ mit $\Theta \subset \mathbb{R}, 0 \in \mathring{\Theta}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P_\vartheta)$ verteilt sind. Wir haben also eine Beobachtung des Produktmaßes P_ϑ^n auf dem Wahrscheinlichkeitsraum $(\Omega^n, \mathcal{A}^n, P_\vartheta^n)$ vorliegen. Man nennt $(\Omega^n, \mathcal{A}^n, \{P_\vartheta^n : \vartheta \in \Theta\})$ auch statistisches Experiment, wobei $\{P_\vartheta^n : \vartheta \in \Theta\}$ gerade ein einparametrisches Teilmodell von $\mathcal{P}^n := \{P^n : P \in \mathcal{P}\}$ darstellt. In diesem Experiment wollen wir das asymptotische Verhalten eines Schätzers für ein Funktional $\kappa(P_\vartheta)$ untersuchen. Dabei folgen wir nun dem Ansatz von Le Cam [26]. Dazu parametrisieren wir zunächst um und betrachten für große n die lokale Familie

$$\{P_{n,\vartheta} := P^n_{\frac{\vartheta}{\sqrt{n}}} : \vartheta \in \mathbb{R}\}.$$

Wegen $0 \in \mathring{\Theta}$ gilt $P_{n,\vartheta} \in \mathcal{P}_1$ für n ausreichend groß. Man bezeichnet dann ϑ als lokalen Parameter und $P_{n,\vartheta}$ als lokale Alternative. Der Ansatz von Le Cam beruht nun darauf die Konvergenz des Experiments

$$E_n := (\Omega^n, \mathcal{A}^n, \{P_{n,\vartheta} : \vartheta \in \mathbb{R}\}) \tag{2.1}$$

im Sinne der schwachen Konvergenz von Experimenten zu untersuchen. Ohne auf diesen Begriff hier näher einzugehen - in der Tat werden wir uns mit der Defintion der schwachen Konvergenz von Experimenten gar nicht näher befassen, sondern lediglich die Konvergenz der zugehörigen Likelihoodprozesse betrachten - zeigt sich, dass das Experiment E_n unter gewissen Voraussetzungen schwach gegen den sogenannten eindimensionalen Gauß-Shift

$$G := (\mathbb{R}, \mathcal{B}, \{\mathcal{N}(\sigma^2\vartheta, \sigma^2) : \vartheta \in \mathbb{R}\})$$

für ein $\sigma^2 > 0$ konvergiert. Ist dies der Fall, so approximieren wir in einem nächsten Schritt einen Schätzer T_n für $\kappa(P_\vartheta)$ durch einen Schätzer T im Limesexperiment G und können diesen auf Optimalität untersuchen.
Wegen

$$\frac{d\mathcal{N}(\sigma^2\vartheta, \sigma^2)}{d\lambda\!\!\!\lambda}(x) = \frac{1}{\sqrt{\pi\sigma^2}} \exp\left(-\frac{(x-\sigma^2\vartheta)^2}{2\sigma^2}\right)$$

gilt für den Log-Likelihoodprozess des Gauß-Shifts

$$\log \frac{d\mathcal{N}(\sigma^2\vartheta, \sigma^2)}{d\mathcal{N}(0, \sigma^2)}(x) = x\vartheta - \frac{1}{2}\vartheta^2\sigma^2.$$

Lokale asymptotische Normalität liegt nun vor, wenn der Log-Likelihoodprozess des Ausgangsexperiments asymptotisch die Gestalt des Gauß-Shifts besitzt.

Definition 2.1 (LAN)
Das Experiment E_n (2.1) heißt lokal asymptotisch normal (LAN), wenn es Folgen von Zufallsvariablen $Z_n : \Omega^n \to \mathbb{R}$, $R_{n,\vartheta} : \Omega^n \to [-\infty, \infty]$ und ein $\sigma^2 > 0$ gibt, so dass gilt

$$\log \frac{dP_{n,\vartheta}}{dP_{n,0}} = \vartheta Z_n - \frac{1}{2}\vartheta^2\sigma^2 + R_{n,\vartheta}$$

mit

$$\mathcal{L}(Z_n|P_{n,0}) \xrightarrow[n\to\infty]{} \mathcal{N}(0,\sigma^2)$$

und

$$R_{n,\vartheta} \xrightarrow[n\to\infty]{P_{n,0}} 0 \quad \forall \vartheta \in \mathbb{R}.$$

$\sigma^{-2}Z_n$ heißt dann eine zentrale Folge.

Eine hinreichende Bedingung für LAN ist die ebenfalls von Le Cam eingeführte L_2-Differenzierbarkeit, siehe Definition 1.1.

Satz 2.2 (Le Cam)
Sei $\mathcal{P}_1 = \{P_\vartheta : \vartheta \in \Theta\}$, $\Theta \subset \mathbb{R}, 0 \in \mathring{\Theta}$ eine L_2-differenzierbare Kurve in $\vartheta = 0$ mit der Tangente $g \in L_2^0(P_0)$. Dann erfüllt E_n (2.1) die LAN-Bedingung mit

$$\sigma^2 = I_0 = \int g^2 \, dP_0$$

und der zentralen Folge

$$\sigma^{-2}Z_n(x_1, ..., x_n) = \frac{1}{\sqrt{n}\sigma^2}\sum_{i=1}^{n} g(x_i).$$

Das heißt es gilt

$$\log \frac{dP_{n,\vartheta}}{dP_{n,0}}(x_1, ..., x_n) = \frac{\vartheta}{\sqrt{n}}\sum_{i=1}^{n} g(x_i) - \frac{1}{2}\vartheta^2\sigma^2 + R_{n,\vartheta}.$$

Beweis. Siehe Strasser [35], S.386, Theorem 75.8. ■

Wir kennen damit also bereits das Grenzverhalten der zu unseren einparametrischen Teilmodellen zugehörigen Experimente oder genauer das Grenzverhalten der Log-Likelihoodprozesse unter $P_{n,0}$. Wie eingangs erwähnt widmen wir uns nun Statistiken in solchen Ausgangsexperimenten, also Statistiken unter Alternativen und betrachten deren Grenzverhalten unter $P_{n,\vartheta}$. Dazu führen wir zunächst das Konzept der Benachbartheit ein.

Definition 2.3
Seien $P_n, Q_n, n \in \mathbb{N}$ Folgen von Wahrscheinlichkeitsmaßen auf $(\Omega_n, \mathcal{A}_n)$. $(Q_n)_n$ heißt zu $(P_n)_n$ benachbart, in Zeichen $Q_n \triangleleft P_n$, wenn für alle Folgen $A_n \in \mathcal{A}_n$ mit $P_n(A_n) \xrightarrow[n\to\infty]{} 0$ auch $Q_n(A_n) \xrightarrow[n\to\infty]{} 0$ gilt.

Es folgt die Wiederholung zweier wichtiger Lemmata von Le Cam über benachbarte Wahrscheinlichkeitsmaße. Sie stehen in ihrer ursprünglichen Form bereits in Le Cam [26] und wurden als separate Lemmata von Hájek und Sidák [14] verwendet, um die asymptotische Güte von Rangtests zu vergleichen, und wurden so als Lemmata von Le Cam bekannt.

Lemma 2.4 (erstes Lemma von Le Cam)
Seien $P_n, Q_n, n \in \mathbb{N}$ Folgen von Wahrscheinlichkeitsmaßen auf $(\Omega_n, \mathcal{A}_n)$ und P, Q Wahrscheinlichkeitsmaße auf $(\Omega, \mathcal{A})$ mit

$$\mathcal{L}\left(\log \frac{dQ_n}{dP_n} | P_n\right) \xrightarrow[n\to\infty]{} \mathcal{L}\left(\log \frac{dQ}{dP} | P\right) =: \mu_1$$

schwach. Dann sind äquivalent:

$$Q_n \lhd P_n,$$

$$Q << P \text{ und}$$

$$\int \exp(x)\, d\mu_1(x) = 1.$$

Beweis. Siehe Hájek, Sidák [14], S.251. ∎

Beispiel 2.5
Sei $E_n = (\Omega^n, \mathcal{A}^n, \{P_{n,\vartheta} : \vartheta \in \mathbb{R}\})$ das Experiment (2.1) mit einer L_2-differenzierbaren Kurve $\mathcal{P}_1 = \{P_\vartheta : \vartheta \in \Theta\}$, $\Theta \subset \mathbb{R}, 0 \in \mathring{\Theta}$ in $\vartheta = 0$.
Nach Satz 2.2 erfüllt E_n die LAN-Bedingung, der Log-Likelihoodprozess besitzt asymptotisch die des Gauß-Shifts. Setze also

$$Q := \mathcal{N}(\sigma^2\vartheta, \sigma^2), \quad P := \mathcal{N}(0, \sigma^2).$$

Dann gilt

$$\log \frac{dQ}{dP}(x) = \vartheta x - \frac{\vartheta^2\sigma^2}{2} \text{ und}$$

$$\mu_1 := \mathcal{L}\left(\log \frac{dQ}{dP} | P\right) = \mathcal{N}\left(-\frac{\vartheta^2\sigma^2}{2}, \vartheta^2\sigma^2\right).$$

LAN impliziert nun gerade

$$\mathcal{L}\left(\log \frac{dP_{n,\vartheta}}{dP_{n,0}} | P_{n,0}\right) \xrightarrow[n\to\infty]{} \mathcal{N}\left(-\frac{\vartheta^2\sigma^2}{2}, \vartheta^2\sigma^2\right) = \mu_1$$

schwach.
Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ normalverteilt mit $\mu \in \mathbb{R}, \sigma^2 > 0$. Dann gilt

$$\begin{aligned}
E(e^X) &= \int e^x \, d\mathcal{N}(\mu,\sigma^2)(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp(x) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(-\frac{x^2 - 2x(\mu+\sigma^2) + (\mu+\sigma^2)^2 - \mu^2 - 2\mu\sigma^2 - \sigma^4 + \mu^2}{2\sigma^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int \exp\left(-\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2}\right) \exp\left(-\frac{-2\mu\sigma^2 - \sigma^4}{2\sigma^2}\right) dx \\
&= \exp\left(\mu + \frac{\sigma^2}{2}\right).
\end{aligned}$$

Also ist $E(e^X) = 1$ äquivalent zu $\mu = -\frac{\sigma^2}{2}$. Daher gilt

$$\int \exp(x) \, d\mu_1(x) = 1$$

und nach Lemma 2.4 folgt $P_{n,\vartheta} \lhd P_{n,0}$.

Das dritte Lemma von Le Cam ist in seiner allgemeinen Form gültig für benachbarte Folgen von Wahrscheinlichkeitsmaßen. Da wir uns in dieser Arbeit auf Folgen der Form $P_{n,\vartheta}$ beschränken, für die die Benachbartheitsbedingung $P_{n,\vartheta} \lhd P_{n,0}$ nach Beispiel 2.5 erfüllt ist, formulieren wir eine vereinfachte Version des Lemmas.

Lemma 2.6 (drittes Lemma von Le Cam)
Seien $S_n : (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}, \mathcal{B})$ Statistiken und Q_ϑ ein Wahrscheinlichkeitsmaß auf $\mathbb{R} \times [-\infty, \infty]$ mit

$$\mathcal{L}\left(\left(S_n, \log \frac{dP_{n,\vartheta}}{dP_{n,0}}\right) | P_{n,0}\right) \to Q_\vartheta$$

schwach. Dann gilt

$$\mathcal{L}\left(\left(S_n, \log \frac{dP_{n,\vartheta}}{dP_{n,0}}\right) | P_{n,\vartheta}\right) \to Q'_\vartheta$$

schwach für ein Wahrscheinlichkeitsmaß $Q'_\vartheta << Q_\vartheta$ mit

$$\frac{dQ'_\vartheta}{dQ_\vartheta}(x,y) = \exp(y).$$

Weiterhin gilt

$$\mathcal{L}\left(S_n | P_{n,\vartheta}\right) \to L_\vartheta$$

für ein Wahrscheinlichkeitsmaß L_ϑ definiert durch

$$L_\vartheta(A) = \int 1_A(x) \exp(y) \, dQ_\vartheta(x,y) = \int 1_A(x) \, dQ'_\vartheta(x,y)$$

für alle $A \in \mathcal{B}$.

Beweis. Für einen ausführlichen Beweis einer allgemeineren Version des ersten Teils vgl. Hájek, Sidák [14], S.257.
Es folgt sofort, dass L_ϑ ein Wahrscheinlichkeitsmaß definiert. Die Konvergenz der Randverteilung $\mathcal{L}\left(S_n|P_{n,\vartheta}\right)$ gegen L_ϑ folgt ebenfalls unmittelbar aus dem ersten Teil: Sei $f:\mathbb{R}\to\mathbb{R}$ stetig und beschränkt. Dann gilt

$$\int f(x)\,d\mathcal{L}\left(S_n|P_{n,\vartheta}\right)(x) = \int f(x)\,d\mathcal{L}\left(\left(S_n, \log\frac{dP_{n,\vartheta}}{dP_{n,0}}\right)|P_{n,\vartheta}\right)(x,y)$$

$$\xrightarrow[n\to\infty]{} \int f(x)\,dQ'_\vartheta(x,y) = \int f(x)\exp(y)\,dQ_\vartheta(x,y) = \int f(x)\,dL_\vartheta(x).$$

■

Korollar 2.7
Seien $S_n : (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}, \mathcal{B})$ *Statistiken mit*

$$\mathcal{L}\left(\left(S_n, \log\frac{dP_{n,\vartheta}}{dP_{n,0}}\right)|P_{n,0}\right) \to \left(\begin{pmatrix} 0 \\ -\frac{\sigma_2^2}{2}\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right)$$

schwach. Dann gilt

$$\mathcal{L}\left(S_n|P_{n,\vartheta}\right) \to \mathcal{N}(\sigma_{12}, \sigma_1^2)$$

schwach.

Beweis. Siehe Janssen [19], S.145, Lemma 15.1. ■

Mit Hilfe des dritten Lemmas von Le Cam können wir jetzt eine Statistik unter Alternativen bzw. einen Schätzer im Ausgangsexperiment E_n asymptotisch durch eine Statistik im Limesexperiment ersetzen.

Satz 2.8
Sei $\mathcal{P}_1 = \{P_\vartheta : \vartheta \in \Theta\}$, $\Theta \subset \mathbb{R}, 0 \in \mathring{\Theta}$ *eine* L_2*-differenzierbare Kurve in* $\vartheta = 0$ *mit der Tangente* $g \in L_2^0(P_0)$ *und* $\sigma^2 := \int g^2\,dP_0$. *Sei* $S_n : (\Omega_n, \mathcal{A}_n) \to (\mathbb{R}, \mathcal{B})$ *eine Statistik, die für alle* $\vartheta \in \mathbb{R}$ *unter* $P_{n,\vartheta} = P^n_{\frac{\vartheta}{\sqrt{n}}}$ *in Verteilung konvergiere. Dann gibt es eine Statistik* $S = S(X_\vartheta)$ *mit* $X_\vartheta \sim \mathcal{N}(\vartheta, \sigma^{-2})$ *im Limesexperiment, sodass gilt*

$$S_n \xrightarrow[n\to\infty]{\mathcal{D}} S$$

unter $P_{n,\vartheta}$ *für alle* $\vartheta \in \mathbb{R}$.

Beweis. Vgl. auch Van der Vaart [37], S.98, Theorem 7.10.
Nach Satz 2.2 erfüllt $\mathcal{P}_1$ die LAN-Bedingung mit der zentralen Folge

$$\sigma^{-2}Y_n := \frac{1}{\sqrt{n}\sigma^2}\sum_{i=1}^{n} g(x_i)$$

mit

$$Y_n \xrightarrow[n\to\infty]{\mathcal{D}} Y \sim \mathcal{N}(0,\sigma^2)$$

unter $P_{n,0}$, also gilt

$$\log\frac{dP_{n,\vartheta}}{dP_{n,0}} \xrightarrow[n\to\infty]{\mathcal{D}} \vartheta Y - \frac{\vartheta^2\sigma^2}{2}$$

unter $P_{n,0}$.
Nach Voraussetzung konvergiert auch S_n in Verteilung unter $P_{n,0}$. Aus der Verteilungskonvergenz folgt Straffheit der Randverteilungen und man sieht leicht, dass hieraus auch die gemeinsame Straffheit von $\left(S_n, \log\frac{dP_{n,\vartheta}}{dP_{n,0}}\right)$ folgt. Mit dem Satz von Prohorov, vgl. Billingsley [6], S.37, Theorem 6.1, folgt damit die Verteilungskonvergenz

$$\mathcal{L}\left(\left(S_n, \log\frac{dP_{n,\vartheta}}{dP_{n,0}}\right)|P_{n,0}\right) \xrightarrow[n\to\infty]{} Q_0 = P^{(S',\Lambda)}$$

entlang Teilfolgen von $\{n\}$ für ein Wahrscheinlichkeitsmaß Q_0 mit der Randverteilung $\mathcal{L}(\Lambda) = \mathcal{L}(\vartheta Y - \frac{\vartheta^2\sigma^2}{2})$.
Nach Lemma 2.6 gibt es nun ein Wahrscheinlichkeitsmaß L_ϑ definiert durch

$$L_\vartheta(A) = \int 1_A(x)\exp(y)\,dQ_0(x,y) = \int 1_A(x)\exp\left(\vartheta y - \frac{\vartheta^2\sigma^2}{2}\right)\,dP^{(S',Y)}$$

für alle $A \in \mathcal{B}$ mit

$$\mathcal{L}(S_n|P_{n,\vartheta}) \xrightarrow[n\to\infty]{} L_\vartheta$$

entlang der entsprechenden Teilfolgen. Da S_n nach Voraussetzung für alle $\vartheta \in \mathbb{R}$ unter $P_{n,\vartheta}$ unabhängig von der Wahl der speziellen Teilfolge in Verteilung konvergiert, ist L_ϑ eindeutig.
Gesucht ist eine Statistik S, so dass $S(X_\vartheta)$ nach L_ϑ verteilt ist.
Wir konstruieren zunächst eine messbare Abbildung T auf $\mathbb{R}\times(0,1)$ mit

$$(T(Y,U),Y) \stackrel{\mathcal{D}}{=} (S',Y)$$

für eine von Y stochastisch unabhängige gleichverteilte Zufallsvariable $U \sim \lambda\!\!\!\lambda_{|(0,1)}$.
Für beliebiges, aber festes $y \in \mathbb{R}$ sei

$$F(x|y) := P(S' \le x|Y = y)$$

die Verteilungsfunktion von $P^{S'|Y=y}$ und

$$Q(u|y) := F^{-1}(u|y)$$

die zugehörige inverse Verteilungsfunktion. Nach Satz A.5 gilt

$$T(y,U) := Q(U|y) \sim P^{S'|Y=y}$$

P^Y-f.s.. Wegen der Unabhängigkeit von Y und U folgt

$$P((T(Y,U),Y) \in A \times B) = \int\limits_B P^{T(Y,U)|Y=y}(A)\,dP^Y(y) = \int\limits_B P^{T(y,U)}(A)\,dP^Y(y)$$

$$= \int\limits_B P^{S'|Y=y}(A)\,dP^Y(y) = P((S',Y) \in A \times B).$$

für alle Borel-messbaren Mengen $A, B \in \mathcal{B}$, also die gewünschte Gleichheit der Verteilungen $(T(Y,U),Y) \stackrel{\mathcal{D}}{=} (S',Y)$.
Sei jetzt $X_\vartheta \sim \mathcal{N}(\vartheta,\sigma^2)$ ebenfalls unabhängig von U.
Dann ist $\sigma^2 X_0 \sim \mathcal{N}(0,\sigma^2) = P^Y$ und es gilt

$$P^{(T(\sigma^2 X_0,U),\sigma^2 X_0)} = P^{(T(Y,U),Y)} = P^{(S',Y)}.$$

Mit dem Satz von Fubini gilt nun aufgrund der Unabhängigkeit von X_ϑ und U für alle $\vartheta \in \mathbb{R}$:

$$\begin{aligned}
P(T(\sigma^2 X_\vartheta, U) \in A) &= \int P(T(\sigma^2 x, U) \in A)\,dP^{X_\vartheta}(x) \\
&= \int P(T(\sigma^2 x, U) \in A)\sqrt{\frac{\sigma^2}{2\pi}}\exp\left(-\frac{(x-\vartheta)^2\sigma^2}{2}\right)dx \\
&= \int P(T(\sigma^2 x, U) \in A)\exp\left(\vartheta\sigma^2 x - \frac{\vartheta^2\sigma^2}{2}\right)\sqrt{\frac{\sigma^2}{2\pi}}\exp\left(-\frac{x^2\sigma^2}{2}\right)dx \\
&= \int\int 1_A(T(\sigma^2 x, u))\,dP^U(u)\exp\left(\vartheta\sigma^2 x - \frac{\vartheta^2\sigma^2}{2}\right)dP^{X_0}(x) \\
&= \int 1_A(T(\sigma^2 x, u))\exp\left(\vartheta\sigma^2 x - \frac{\vartheta^2\sigma^2}{2}\right)dP^{(X_0,U)}(x,u) \\
&= \int 1_A(T(\sigma^2 X_0, U))\exp\left(\vartheta\sigma^2 X_0 - \frac{\vartheta^2\sigma^2}{2}\right)dP \\
&= \int 1_A(x)\exp\left(\vartheta y - \frac{\vartheta^2\sigma^2}{2}\right)dP^{(T(\sigma^2 X_0,U),\sigma^2 X_0)}(x,y)
\end{aligned}$$

$$= \int 1_A(x) \exp\left(\vartheta y - \frac{\vartheta^2\sigma^2}{2}\right) dP^{(S',Y)}(x,y) = L_\vartheta(A)$$

für alle $A \in \mathcal{A}$ und die Behauptung ist erfüllt für $S(X_\vartheta) := T(\sigma^2 X_\vartheta, U)$. ■

2.2 Der Faltungssatz

Im parametrischen Fall heißt eine Schätzfolge T_n Fisher-effizient in P_0, falls die Varianz $Var_{P_0^n}(\sqrt{n}T_n)$ für $n \to \infty$ gegen die Cramér-Rao-Schranke konvergiert. Wie das folgende Beispiel zeigen wird, reicht es jedoch nicht aus, punktweise Konvergenz gegen diese Schranke für alle $P_0 \in \mathcal{P}$ zu fordern.

Beispiel 2.9 (Hodges Estimator)
Betrachte das Experiment $\{\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{P}\}$ mit $\mathcal{P} = \{P_\vartheta = \mathcal{N}(\vartheta, 1) : \vartheta \in \mathbb{R}\}$ und dem zu schätzenden Funktional $\kappa(\vartheta) = \vartheta$. Bekanntlich ist der gleichmäßig beste erwartungstreue Schätzer für κ gegeben durch

$$T_n := \frac{1}{n}\sum_{i=1}^n X_i.$$

Nach dem zentralen Grenzwertsatz gilt

$$\sqrt{n}(T_n - \vartheta) \xrightarrow[n\to\infty]{\mathcal{D}} Z \sim \mathcal{N}(0,1)$$

und die Cramér-Rao-Schranke 1 wird auch im Limes angenommen.
Definiere nun

$$S_n := T_n 1_{\{|T_n| \geq n^{-1/4}\}}.$$

Dann gilt für $\vartheta \neq 0$

$$P_\vartheta(S_n \neq T_n) = P_\vartheta(|T_n| < n^{-1/4}) - \underbrace{P_\vartheta(|T_n| = 0)}_{=0}$$

$$= P_\vartheta(-n^{-1/4} < T_n < n^{-1/4})$$

$$= P_\vartheta(-n^{1/4} - \vartheta\sqrt{n} < \sqrt{n}(T_n - \vartheta) < n^{1/4} - \vartheta\sqrt{n})$$

$$= P_\vartheta(\sqrt{n}(T_n - \vartheta) < n^{1/4} - \vartheta\sqrt{n}) - P_\vartheta(\sqrt{n}(T_n - \vartheta) \leq -n^{1/4} - \vartheta\sqrt{n}).$$

Für $\vartheta > 0$ gilt $\pm n^{1/4} - \vartheta\sqrt{n} \to -\infty$ und für $\vartheta < 0$ gilt $\pm n^{1/4} - \vartheta\sqrt{n} \to \infty$ für $n \to \infty$. In beiden Fällen folgt mit Lemma A.7

$$P_\vartheta(\sqrt{n}(S_n - \vartheta) \neq \sqrt{n}(T_n - \vartheta)) = P_\vartheta(S_n \neq T_n) \xrightarrow[n\to\infty]{} 0.$$

Insbesondere folgt damit die stochastische Konvergenz

$$\sqrt{n}(S_n - \vartheta) - \sqrt{n}(T_n - \vartheta) \xrightarrow[n\to\infty]{} 0$$

und mit der allgemeinen Version des Lemmas von Slutsky, vgl. Klenke [24], S.243, Satz 13.18,

$$\sqrt{n}(S_n - \vartheta) \xrightarrow[n\to\infty]{\mathcal{D}} Z \sim \mathcal{N}(0,1).$$

S_n verhält sich also für $\vartheta \neq 0$ asymptotisch genauso wie T_n.
Für $\vartheta = 0$ gilt jedoch wiederum nach Lemma A.7

$$\begin{aligned}
&P_0(S_n = 0) = P_0\left(|T_n| < n^{-1/4}\right) \\
&= P_0\left(-n^{1/4} < \sqrt{n}T_n < n^{1/4}\right) \\
&= P_0\left(\sqrt{n}T_n < n^{1/4}\right) - P_0\left(\sqrt{n}T_n \leq -n^{1/4}\right) \xrightarrow[n\to\infty]{} 1 - 0 = 1,
\end{aligned}$$

was sehr viel stärker ist als Verteilungskonvergenz gegen $\mathcal{N}(0,1)$. Die Varianz konvergiert hier gegen 0 und S_n unterbietet somit die Cramér-Rao-Schranke. Diese scheinbare Verbesserung von T_n geschieht jedoch auf Kosten des Verhaltens in der Umgebung von 0 für festes n, vgl. hierzu Van der Vaart [37], S.110. Dieses Phänomen nennt man auch Supereffizienz.

Schon in den parametrischen (Teil-)Modellen können die Cramér-Rao-Schranken also und damit auch ihr Supremum asymptotisch unterboten werden. Es stellt sich hier nun also die Frage, in welcher Klasse von Schätzern sich dennoch ein bester Schätzer finden lässt, oder anders formuliert, welche (nicht allzu einschränkende) Bedingung an die Schätzer gestellt werden muss, um Supereffizienz auszuschließen. Im parametrischen Fall ist die Antwort bekannt. Das Beispiel des in 0 supereffizienten Hodges Estimators zeigt, dass es nicht ausreicht das Grenzverhalten von Schätzern lokal für jedes $P \in \mathcal{P}$ zu betrachten, sondern für das entsprechende Teilmodell verschiedene Parameter gleichzeitig betrachtet werden müssen. Dies führt zum Begriff der Regularität.

Definition 2.10
Sei $\mathcal{P}_1 = \{P_\vartheta : \vartheta \in \Theta\}$, $\Theta \subset \mathbb{R}$, $0 \in \mathring{\Theta}$, eine parametrische in P_0 L_2-differenzierbare Familie von Wahrscheinlichkeitsmaßen mit Tangente g, sei $\kappa : \mathcal{P}_1 \to \mathbb{R}$ ein in P_0 differenzierbares Funktional und sei $T_n : \Omega^n \to \mathbb{R}$ eine Schätzfolge für κ. Dann heißt T_n

regulär in P_0, falls es für alle $\vartheta \in \Theta$ ein von ϑ unabhängiges Wahrscheinlichkeitsmaß L gibt mit

$$\mathcal{L}\left(\sqrt{n}\left(T_n - \kappa\left(P_{\frac{\vartheta}{\sqrt{n}}}\right)\right)|P^n_{\frac{\vartheta}{\sqrt{n}}}\right) \to L$$

schwach für $n \to \infty$.

Das Grenzverhalten soll sich also unter lokalen Alternativen, die mit der Rate $\frac{1}{\sqrt{n}}$ gegen 0 konvergieren, nicht ändern. Im parametrischen Fall werden wir sehen, dass diese Bedingung ausreicht, um Supereffizienz auszuschließen, also gerade in der Klasse der regulären Schätzer ein asymptotisch bester gefunden werden kann.

Um dies zu zeigen, benötigen wir zuvor noch ein Hilfslemma zu Normalverteilungsmodellen bei finitem Stichprobenumfang. Für mehrdimensionale Normalverteilungsmodelle vgl. Van der Vaart [37], Proposition 8.4, S.113.
Sei

$$P_\vartheta = \mathcal{N}(\vartheta, \sigma^2), \vartheta \in \Theta \subset \mathbb{R}$$

für ein $\sigma^2 > 0$, $X \sim P_\vartheta$ und sei $T = T(X)$ ein Schätzer in diesem Modell für $\kappa(P_\vartheta) = a\vartheta, a \in \mathbb{R}$.

Lemma 2.11
Ist die Verteilung

$$\mathcal{L}(T - a\vartheta|P_\vartheta) = L$$

unabhängig von $\vartheta \in \Theta$, dann besitzt $L = P_0^T$ die Darstellung

$$L = \mathcal{N}(0, a^2\sigma^2) * \nu$$

für ein Wahrscheinlichkeitsmaß ν.

Beweis. Siehe Van der Vaart [37], S.113. ■

Ein solcher Schätzer T wird auch "äquivariant in Verteilung" genannt. Das einfachste Beispiel hierfür ist der gleichmäßig beste erwartungstreue (UMVU) Schätzer $T = aX$. Die Verteilung von T unter P_ϑ ist hier trivialerweise $\mathcal{N}(0, a^2\sigma^2)$ für alle ϑ. Ein in Verteilung äquivarianter Schätzer ist also verteilt wie die Summe aus aX und einer davon unabhängigen Zufallsvariablen.

Mit Hilfe dieses Lemmas und der LAN-Theorie, können wir nun eine Version des Faltungssatzes von Hájek-Le Cam beweisen, vgl. Van der Vaart [37], Theorem 8.8, S.115.

Satz 2.12 (Faltungssatz)
Sei $\mathcal{P}_1 = \{P_\vartheta : \vartheta \in \Theta\}$ eine parametrische in $P_0 \in \mathcal{P}_1$ L_2-differenzierbare Familie von Wahrscheinlichkeitsmaßen mit Tangente g, sei $\kappa : \mathcal{P} \to \mathbb{R}$ ein in P_0 differenzierbares Funktional mit Ableitung $\kappa'(P_0)$ und sei $T_n : \Omega^n \to \mathbb{R}$ eine reguläre Schätzfolge für κ mit Grenzverteilung L. Dann gibt es ein Wahrscheinlichkeitsmaß ν mit

$$L = \mathcal{N}\left(0, \frac{\kappa'(P_0)^2}{I_0}\right) * \nu,$$

wobei $I_0 := \int g^2 \, dP_0$ die Fisher-Information in P_0 bezeichnet.

Beweis. Wir schreiben

$$S_n := \sqrt{n}\,(T_n - \kappa\,(P_0)) = \sqrt{n}\left(T_n - \kappa\left(P_{\frac{\vartheta}{\sqrt{n}}}\right)\right) + \sqrt{n}\left(\kappa\left(P_{\frac{\vartheta}{\sqrt{n}}}\right) - \kappa\,(P_0)\right)$$

Da T_n regulär ist, konvergiert der erste Summand unter $P_{n,\vartheta} = P^n_{\frac{\vartheta}{\sqrt{n}}}$ in Verteilung gegen eine Zufallsvariable mit der Verteilung L. Der zweite Summand konvergiert aufgrund der Differenzierbarkeit des Funktionals gegen $\vartheta\kappa'(P_0)$. Insgesamt gilt daher nach Slutsky

$$S_n \xrightarrow[n\to\infty]{} Z + \vartheta\kappa'(P_0)$$

in Verteilung unter $P_{n,\vartheta}$ mit $Z \sim L$.
Nach Satz 2.8 gibt es somit eine Statistik $S = S(X)$ mit

$$S_n \xrightarrow{D, P_{n,\vartheta}} S(X)$$

für ein $X \sim \mathcal{N}(\vartheta, \sigma^{-2})$, also

$$S - \vartheta\kappa'(P_0) \sim L.$$

Da L unabhängig ist von ϑ, ist S somit äquivariant in Verteilung und nach Lemma 2.11 folgt die Behauptung für $\vartheta = 0$. ■

Somit bildet $\frac{\kappa'(P_0)^2}{I_0}$ in dieser Klasse von Schätzern eine untere Schranke für die Varianz der Grenzverteilung, denn es gilt $Var(L) = \frac{\kappa'(P_0)^2}{I_0} + Var(\nu)$. Kann also ein regulärer Schätzer gefunden werden mit

$$\mathcal{L}(\sqrt{n}(T_n - \kappa(P_0))|P_0) \to \mathcal{N}\left(0, \frac{\kappa'(P_0)^2}{I_0}\right),$$

das heißt mit $\nu = \varepsilon_0$, so ist er in diesem (Teil-)Modell der Beste in der Klasse der regulären Schätzer.

Kapitel 3

Asymptotische Effizienz

3.1 Asymptotische Effizienz für Funktionale

Nachdem wir im zweiten Kapitel die parametrischen Teilmodelle gemäß der Theorie von Le Cam untersucht haben, kommen wir nun zurück zu unserer nichtparametrischen Familie von Wahrscheinlichkeitsmaßen $\mathcal{P}$ und legen unser Augenmerk in diesem Kapitel zusätzlich vermehrt auf die zu schätzenden Funktionale $\kappa : \mathcal{P} \to \mathbb{R}$, wobei wir den Arbeiten von Pfanzagl und Wefelmeyer [30], [30], Strasser [35], Van der Vaart [37], BKRW [7] und Janssen [15], [16], [19] folgen.
Dabei untersuchen wir einen Schätzer zunächst auf asymptotische Effizienz lokal in einem festen Fußpunkt $P_0 \in \mathcal{P}$ und nennen ihn anschließend asymptotisch effizient, wenn er diese Eigenschaft lokal für alle $P_0 \in \mathcal{P}$ erfüllt.

Die einparametrischen Kurven sind im nichtparametrischen Modell $\mathcal{P}$ nach wie vor enthalten. Ein Funktional κ ist im nichtparametrischen Modell offensichtlich „schwerer" zu schätzen als in jeder einzelnen Kurve, da mehr Verteilungen in Frage kommen. Daher liefert jede L_2-differenzierbare Kurve, entlang derer κ differenzierbar ist, eine untere Schranke für die „Varianz" der Grenzverteilung regulärer Schätzer (falls diese existiert).

Zunächst erweitern wir den Begriff der Differenzierbarkeit des Funktionals κ auf den nichtparametrischen Fall.

Definition 3.1 (differenzierbares Funktional)
Sei $\mathcal{P}$ eine nichtparametrische Familie von Wahrscheinlichkeitsmaßen. Ein

Funktional $\kappa : \mathcal{P} \to \mathbb{R}$ heißt differenzierbar in $P_0 \in \mathcal{P}$ mit einem Gradienten $\dot{\kappa} \in L_2^0(P_0)$, falls für jede $L_2(P_0)$-differenzierbare Kurve $\vartheta \mapsto P_\vartheta$ lokal in $\mathcal{P}$ mit Tangente g gilt:

$$\lim_{\vartheta \to 0} \frac{\kappa(P_\vartheta) - \kappa(P_0)}{\vartheta} = \int \dot{\kappa} g \, dP_0. \tag{3.1}$$

Dieser klassische Differenzierbarkeitsbegriff lässt sich für unsere Zwecke noch stark vereinfachen. Wir werden sehen, dass es vollkommen ausreicht, zu jeder Tangente nur eine Kurve zu finden, die (3.1) erfüllt und benötigen nicht einmal alle Tangenten des Tangentialraums.

Definition 3.2 (dicht differenzierbares Funktional)
Sei $\mathcal{P}$ eine nichtparametrische Familie von Wahrscheinlichkeitsmaßen. Ein Funktional $\kappa : \mathcal{P} \to \mathbb{R}$ heißt dicht differenzierbar in $P_0 \in \mathcal{P}$ mit einem Gradienten $\dot{\kappa} \in L_2^0(P_0)$, falls es eine konvexe dichte Teilmenge $D \subset T(P_0, \mathcal{P})$ gibt, für die gilt: Für alle $g \in D$ existiert eine $L_2(P_0)$-differenzierbare Kurve $\vartheta \mapsto P_\vartheta$ in $\mathcal{P}$ mit Tangente g, für die (3.1) erfüllt ist.

Im Allgemeinen ist $\dot{\kappa} \in L_2^0(P_0)$ nicht eindeutig. So ist wegen $g \in T(P_0, \mathcal{P})$ auch $\dot{\kappa} + h$ für alle $h \in T(P_0, \mathcal{P})^\perp$ ein Gradient, wobei $T(P_0, \mathcal{P})^\perp$ den Orthogonalraum von $T(P_0, \mathcal{P})$ in $L_2^0(P_0)$ bezeichnet, denn es gilt

$$\int (\dot{\kappa} + h) g \, dP_0 = \int \dot{\kappa} g \, dP_0 + \int h g \, dP_0 = \int \dot{\kappa} g \, dP_0.$$

Es existiert jedoch bereits im Fall der dichten Differenzierbarkeit ein eindeutig bestimmter kanonischer Gradient $\tilde{\kappa} \in T(P_0, \mathcal{P})$.

Satz 3.3 (kanonischer Gradient)
In der Situation von Definition 3.2 existiert ein eindeutiger Gradient $\tilde{\kappa} \in T(P_0, \mathcal{P})$. Dieser wird als kanonischer Gradient bezeichnet und besitzt in der Menge aller Gradienten die kleinste Norm.

Beweis. Jeder Gradient $\dot{\kappa}$ liefert eine eindeutige stetige Linearform

$$\Psi : D \subset T(P_0, \mathcal{P}) \to \mathbb{R}, g \mapsto \int \dot{\kappa} g \, dP_0.$$

Nach Werner [40], S.48, Satz II.1.5 ist Ψ durch seine Werte auf der dichten Teilmenge $D \subset T(P_0, \mathcal{P})$ auch auf $T(P_0, \mathcal{P})$ bereits eindeutig bestimmt.
Nach dem Satz von Frechét-Riesz, siehe Satz A.11, gibt es daher genau ein Element

$\tilde{\kappa} \in T(P_0, \mathcal{P})$ mit

$$\Psi(g) = \int \tilde{\kappa} g \, dP_0 \quad \forall \, g \in T(P_0, \mathcal{P}).$$

Sei nun $\dot{\kappa} \in L_2^0(P_0)$ ein weiterer Gradient. Dann erhält man $\tilde{\kappa}$ durch die orthogonale Projektion $p : L_2^0(P_0) \to T(P_0, \mathcal{P})$ von $\dot{\kappa}$ auf $T(P_0, \mathcal{P})$, denn per Definition der orthogonalen Projektion gilt

$$\langle \dot{\kappa} - p(\dot{\kappa}), g \rangle = \int (\dot{\kappa} - p(\dot{\kappa})) g \, dP_0 = 0$$

$$\Leftrightarrow \int \dot{\kappa} g \, dP_0 = \int p(\dot{\kappa}) g \, dP_0$$

für alle $g \in T(P_0, \mathcal{P})$. Mit dem Satz des Pythagoras, vgl. Werner [40], S.223, folgt

$$\|\dot{\kappa}\|^2 = \|\dot{\kappa} - p(\dot{\kappa})\|^2 + \|p(\dot{\kappa})\|^2 \geq \|p(\dot{\kappa})\|^2 = \|\tilde{\kappa}\|^2$$

und $\tilde{\kappa}$ besitzt die kleinste Norm. ■

Bemerkung 3.4

a) Typischerweise ist $\tilde{\kappa} \neq 0$. Ansonsten sind im Folgenden $\mathcal{N}\left(0, \|\tilde{\kappa}\|^2\right)$-Normalverteilungen als Einpunktverteilungen in 0 zu verstehen.
b) Die Existenz der dichten Teilmenge $D \subset T(P_0, \mathcal{P})$ mit $D \subset K(P_0, \mathcal{P})$ impliziert $T(P_0, \mathcal{P}) = \overline{K(P_0, \mathcal{P})}$.
c) Im Beweis wurde gezeigt, dass man den kanonischen Gradienten $\tilde{\kappa}$ aus einem beliebigen Gradienten $\dot{\kappa} \in L_2^0(P_0)$ durch die orthogonale Projektion auf $T(P_0, \mathcal{P})$ erhält. Gilt $T(P_0, \mathcal{P}) = L_2^0(P_0)$ wie gemäß Lemma 1.8 für $\mathcal{P} = \mathcal{P}^*$, so ist $\dot{\kappa} \in L_2^0(P_0)$ bereits eindeutig bestimmt und die Projektion entfällt.

Beispiel 3.5

Sei $\mathcal{P} = \mathcal{P}^*$ das volle Modell, $P_0 \in \mathcal{P}$ und $h \in L_2(P_0)$. Ist für alle $L_2(P_0)$-differenzierbaren Kurven $\vartheta \mapsto P_\vartheta$ die Abbildung $\vartheta \mapsto \int h^2 \, dP_\vartheta$ stetig in 0, so ist das Funktional $\kappa(P) = \int h \, dP$ nach Janssen [19], Beispiel 18.2, S.168, differenzierbar im klassischen Sinn mit dem Gradienten $\dot{\kappa} = h$. Dies ist auch bereits der kanonische Gradient, $\tilde{\kappa} = h$, da nach Lemma 1.8 für das volle Modell $T(P_0, \mathcal{P}) = L_2^0(P_0)$ gilt.
Schon bei einem solch einfachen Beispiel wird nun deutlich werden, dass die dichte Differenzierbarkeit eine starke Vereinfachung darstellt.
Sei $D = D^* \subset T(P_0, \mathcal{P})$ die dichte Teilmenge der beschränkten Tangenten aus Lemma 1.9. Zu $g \in D$ wähle die $L_2(P_0)$-differenzierbare Kurve $\vartheta \mapsto P_\vartheta$ aus Lemma 1.13

mit $\frac{dP_\vartheta}{dP_0} = 1 + \vartheta g$. Dann gilt

$$\begin{aligned}\frac{1}{\vartheta}\left(\kappa(P_\vartheta) - \kappa(P_0)\right) &= \frac{1}{\vartheta}\left(\int h\, dP_\vartheta - \int h\, dP_0\right)\\ &= \frac{1}{\vartheta}\int h(1+\vartheta g - 1)\, dP_0 = \int hg\, dP_0\end{aligned}$$

und κ ist dicht differenzierbar bezüglich D mit dem kanonischen Gradienten h.

Korollar 3.6 (Kettenregel)
Sei $h : \mathbb{R} \to \mathbb{R}$ eine differenzierbare Funktion mit Ableitung h' und $\kappa : \mathcal{P} \to \mathbb{R}$ ein in $P_0 \in \mathcal{P}$ bezüglich D dicht differenzierbares Funktional mit dem kanonischen Gradienten $\tilde{\kappa}$. Dann ist das Funktional $\iota := h \circ \kappa$ ebenfalls in $P_0 \in \mathcal{P}$ bezüglich D dicht differenzierbar mit dem kanonischen Gradienten

$$\tilde{\iota} = h'(\kappa(P_0))\tilde{\kappa}.$$

Beweis. Der Beweis ergibt sich sofort aus der Kettenregel für Funktionen $\mathbb{R} \to \mathbb{R}$. Sei $g \in A$. Dann gibt es eine $L_2(P_0)$-differenzierbare Kurve $\vartheta \mapsto P_\vartheta$ in $\mathcal{P}$ mit Tangente g und

$$\frac{d}{d\vartheta}\kappa(P_\vartheta)_{|\vartheta=0} = \int \tilde{\kappa} g\, dP_0.$$

Mit der Kettenregel folgt

$$\frac{d}{d\vartheta}\iota(P_\vartheta)_{|\vartheta=0} = h'(\kappa(P_0))\int \tilde{\kappa} g\, dP_0.$$

■

Als nächstes erweitern wir analog den Begriff der Regularität auf die nichtparametrische Situation.

Definition 3.7
Sei $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen, sei κ ein in P_0 bezüglich $D \subset T(P_0, \mathcal{P})$ dicht differenzierbares Funktional und sei $T_n : \Omega^n \to \mathbb{R}$ eine Schätzfolge für κ. Dann heißt T_n D-regulär in P_0, falls es ein Wahrscheinlichkeitsmaß L gibt, für das gilt: Für alle $g \in D$ existiert eine $L_2(P_0)$-differenzierbare Kurve $\Theta \to \mathcal{P}$, $\vartheta \mapsto P_\vartheta$ in $\mathcal{P}$ mit Tangente g, so dass für alle $\vartheta \in \Theta$ gilt:

$$\mathcal{L}\left(\sqrt{n}\left(T_n - \kappa\left(P_{\frac{\vartheta}{\sqrt{n}}}\right)\right)|P^n_{\frac{\vartheta}{\sqrt{n}}}\right) \to L$$

schwach für $n \to \infty$.

Ist κ nun dicht differenzierbar mit kanonischem Gradienten $\tilde{\kappa}$, so gilt nach dem Faltungssatz 2.12 für alle Tangenten $g \in D$, $g \neq 0$, mit zugehörigen Kurven $\vartheta \mapsto P_\vartheta$:

$$L = \mathcal{N}\left(0, \frac{\langle \tilde{\kappa}, g\rangle_{P_0}^2}{\langle g, g\rangle_{P_0}}\right) * \nu_g.$$

Die Cauchy-Schwarz-Ungleichung liefert zudem

$$\frac{\langle \tilde{\kappa}, g\rangle_{P_0}^2}{\langle g, g\rangle_{P_0}} \leq \frac{\langle \tilde{\kappa}, \tilde{\kappa}\rangle_{P_0}\langle g, g\rangle_{P_0}}{\langle g, g\rangle_{P_0}} = \langle \tilde{\kappa}, \tilde{\kappa}\rangle_{P_0} = \|\tilde{\kappa}\|^2.$$

Gleichheit gilt hier, wenn $\tilde{\kappa}$ und g linear abhängig sind. Der Ausdruck wird somit maximiert für $g = \tilde{\kappa} \in T(P_0, \mathcal{P})$. Für den Spezialfall $\tilde{\kappa} \in D$ und dass T_n D-regulär ist, ist daher unmittelbar klar, dass $\|\tilde{\kappa}\|^2$ in einem solchen nichtparametrischen Modell eine untere Schranke für die Varianz der Grenzverteilung L eines solchen Schätzers T_n bildet. Die zugehörige Kurve an $g = \tilde{\kappa}$ ist in diesem Fall das am schwersten zu schätzende Teilmodell, auch least favorable submodel genannt. Es ist jedoch nicht notwendig, dass $\tilde{\kappa}$ bereits in D liegt und durch ein Teilmodell erreichbar ist.

Satz 3.8

Sei $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen, sei κ ein in P_0 bezüglich $D \subset T(P_0, \mathcal{P})$ dicht differenzierbares Funktional mit kanonischem Gradienten $\tilde{\kappa} \neq 0$ und sei $T_n : \Omega^n \to \mathbb{R}$ eine in P_0 D-reguläre Schätzfolge. Dann gibt es ein Wahrscheinlichkeitsmaß ν mit

$$\mathcal{L}(\sqrt{n}(T_n - \kappa(P_0))|P_0) \to \mathcal{N}\left(0, \|\tilde{\kappa}\|^2\right) * \nu \tag{3.2}$$

schwach für $n \to \infty$.

Beweis. Wähle $g_m \in D$ mit $\|\tilde{\kappa} - g_m\|_{L_2(P_0)} \xrightarrow[m\to\infty]{} 0$. Dann gibt es zugehörige L_2-differenzierbare Kurven $\vartheta \mapsto P_\vartheta^{(m)}$ mit Tangenten g_m und ein Wahrscheinlichkeitsmaß L mit

$$\mathcal{L}\left(\sqrt{n}\left(T_n - \kappa\left(P_{n,\vartheta}^{(m)}\right)\right) \mid \left(P_{n,\vartheta}^{(m)}\right)^n\right) \xrightarrow[n\to\infty]{} L$$

schwach für alle m. Mit dem Faltungssatz 2.12 folgt

$$L = \mathcal{N}\left(0, \frac{\langle \tilde{\kappa}, g_m\rangle^2}{\langle g_m, g_m\rangle}\right) * \nu_m$$

für alle m. Weiterhin gilt

$$\frac{\langle \tilde{\kappa}, g_m\rangle^2}{\langle g_m, g_m\rangle} \xrightarrow[m\to\infty]{} \frac{\langle \tilde{\kappa}, \tilde{\kappa}\rangle^2}{\langle \tilde{\kappa}, \tilde{\kappa}\rangle} = \|\tilde{\kappa}\|^2.$$

Nun gilt nach dem Eindeutigkeitssatz für die Fouriertransformierte

$$\hat{L}(y) = \left(\mathcal{N}\left(0, \frac{\langle \tilde{\kappa}, g_m \rangle^2}{\langle g_m, g_m \rangle}\right) * \nu_m \right)^{\widehat{}} (y) = \widehat{\mathcal{N}\left(0, \frac{\langle \tilde{\kappa}, g_m \rangle^2}{\langle g_m, g_m \rangle}\right)}(y) \cdot \widehat{\nu_m}(y)$$

und wegen

$$\begin{aligned} \widehat{\mathcal{N}\left(0, \frac{\langle \tilde{\kappa}, g_m \rangle^2}{\langle g_m, g_m \rangle}\right)}(y) &= \exp\left(-\frac{1}{2}\frac{\langle \tilde{\kappa}, g_m \rangle^2}{\langle g_m, g_m \rangle} y^2\right) \\ &\xrightarrow[m\to\infty]{} \exp\left(-\frac{1}{2}\|\tilde{\kappa}\|^2 y^2\right) = \widehat{\mathcal{N}(0, \|\tilde{\kappa}\|^2)}(y) \end{aligned}$$

gibt es eine Funktion $\widehat{\nu}$ mit

$$\widehat{\nu_m} \xrightarrow[m\to\infty]{} \widehat{\nu}.$$

Nach dem Stetigkeitssatz für Fouriertransformierte existiert daher ein Wahrscheinlichkeitsmaß ν mit

$$\nu_m \xrightarrow[m\to\infty]{} \nu$$

schwach. Insgesamt folgt damit

$$L = \mathcal{N}\left(0, \|\tilde{\kappa}\|^2\right) * \nu.$$

■

$\|\tilde{\kappa}\|^2$ bildet also auch in diesem Fall eine untere Schranke für die asymptotische Varianz $Var(L)$ eines regulären Schätzers, falls diese existiert. Diese Schranke nennen wir auch Informationsschranke, da sie die größtmögliche Schranke darstellt, die wir allein aus Informationen über das Modell erhalten können. Im einparametrischen Fall und für $\kappa(P_\vartheta) = \vartheta$ ist sie das Inverse der Fisher-Information.

Ein Schätzer T_n besitzt somit die optimale Grenzverteilung, wenn ν verschwindet, das heißt wenn gilt $\nu = \varepsilon_0$. In diesem Fall nennen wir T_n asymptotisch effizient.

Definition 3.9

a) Sei $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen, sei κ ein in $P_0 \in \mathcal{P}$ bezüglich $D \subset T(P_0, \mathcal{P})$ dicht differenzierbares Funktional mit kanonischem Gradienten $\tilde{\kappa}$ und sei $T_n : \Omega^n \to \mathbb{R}$ eine in P_0 D-reguläre Schätzfolge. Dann heißt T_n asymptotisch effizient im Fußpunkt P_0, wenn gilt

$$\sqrt{n}(T_n - \kappa(P_0)) \xrightarrow[n\to\infty]{\mathcal{D}} Z \sim \mathcal{N}\left(0, \|\tilde{\kappa}\|^2\right)$$

unter P_0.

b) Gilt (a) für alle $P_0 \in \mathcal{P}$, so nennen wir T_n asymptotisch effizient.

Es sei darauf hingewiesen, dass die asymptotische Effizienz in (a) ein lokales Konzept in einem festen Fußpunkt P_0 ist und sowohl die dichte Teilmenge D als auch der kanonische Gradient $\tilde{\kappa}$ von P_0 abhängen. Für die asymptotische Effizienz muss es also für jedes $P_0 \in \mathcal{P}$ eine solche dichte Teilmenge $D_{P_0} \subset T(P_0, \mathcal{P})$ geben, bezüglich derer κ dicht differenzierbar ist mit einem kanonischen Gradienten $\tilde{\kappa}_{P_0}$ und T_n muss unter allen P_0 obige Verteilungskonvergenz zeigen mit der von P_0 abhängigen Varianz $\|\tilde{\kappa}_{P_0}\|^2_{P_0}$ der Grenzverteilung L_{P_0}.

Zum Abschluss dieses Abschnitts betrachten wir noch spezielle Teilmodelle unseres nichtparametrischen Modells.

Definition 3.10
Sei $\mathcal{P}$ eine nichtparametrische Familie von Wahrscheinlichkeitsmaßen. Sei $A \subset T(P_0, \mathcal{P})$ ein linearer Teilraum. Ein Funktional $\kappa : \mathcal{P} \to \mathbb{R}$ heißt differenzierbar bezüglich A in P_0 mit einem Gradienten $\dot{\kappa} \in L_2^0(P_0)$, falls für alle $g \in A$ eine $L_2(P_0)$-differenzierbare Kurve $\vartheta \mapsto P_\vartheta$ in $\mathcal{P}$ mit Tangente g existiert, die (3.1) erfüllt.

Sei nun $A \subset T$ eine solche Teilmenge, bezüglich derer κ differenzierbar sei. Dann bildet $\overline{A} \subset T(P_0, \mathcal{P})$ einen Teilhilbertraum und nach dem Satz von Fischer-Riesz gibt es einen Gradienten $\tilde{\kappa}' \in \overline{A}$ mit $\langle \dot{\kappa}, g \rangle = \langle \tilde{\kappa}', g \rangle$ für alle $g \in A$, der unter allen Gradienten die kleinste Norm besitzt, vgl. Satz 3.3.
Beispielsweise kann mit $\overline{A} = T(P_0, \mathcal{P}')$ der Tangentialraum eines Teilmodells $\mathcal{P}' \subset \mathcal{P}$ mit dem kanonischen Gradienten $\tilde{\kappa}' \in T(P_0, \mathcal{P}')$ sein. In diesem Fall erhalten wir über $\tilde{\kappa}'$ mit Satz 3.8 eine untere Schranke für die asymptotische Varianz eines A-regulären Schätzers im Teilmodell $\mathcal{P}'$, wobei wir A-Regularität analog zu Definition 3.7 bezüglich der dichten Teilmenge $A \subset T(P_0, \mathcal{P}')$ verstehen, also Regularität für Kurven mit Tangente $g \in A$ fordern. Diese Schranke ist kleiner oder gleich der Schranke des gesamten Modells $\mathcal{P}$, da die Projektion normverkürzend wirkt. Heuristisch folgt dies bereits aus der Tatsache, dass in einem größeren Modell nicht besser geschätzt werden kann als in einem Teilmodell $\mathcal{P}'$. Nimmt ein A-regulärer Schätzer die Schranke $\|\tilde{\kappa}'\|^2$ an, so ist er asymptotisch effizient für das Teilmodell $\mathcal{P}'$.
Existiert ein kanonischer Gradient $\tilde{\kappa} \in T(P_0, \mathcal{P})$ für das gesamte Modell (ist κ also zumindest dicht differenzierbar), so ist $\tilde{\kappa}'$ die orthogonale Projektion von $\tilde{\kappa}$ auf $\overline{A}$. Ist der kanonische Gradient $\tilde{\kappa}$ für $\kappa_{|\mathcal{P}}$ bereits im Abschluss $\overline{A}$ von A enthalten, das

heißt $\tilde{\kappa} = \tilde{\kappa}'$, so ergibt sich die Informationsschranke in dem gesamten Modell $\mathcal{P}$ an der Stelle P_0 bereits durch die Betrachtung von Kurven mit Tangenten in A.
Es sei darauf hingewiesen, dass zur Berechnung von $\tilde{\kappa}$ aber die dichte Differenzierbarkeit von $\kappa : \mathcal{P} \to \mathbb{R}$ erforderlich ist.

3.2 Asymptotische Linearität

Dieser Abschnitt beruht auf der Arbeit von Pfanzagl, Wefelmeyer [30]. Es soll ein Zusammenhang zwischen asymptotischer Effizienz und asymptotischer Linearität hergestellt werden. Die asymptotische Normalität unserer asymptotisch effizienten Schätzer mit der entsprechenden Varianz impliziert eine Entwicklung um den kanonischen Gradienten und umgekehrt.

Definition 3.11 (asymptotische Linearität)
Sei $\mathcal{P}$ eine nichtparametrische Familie von Wahrscheinlichkeitsmaßen, seien $X_1, ..., X_n$ unabhängig identisch nach $P_0 \in \mathcal{P}$ verteilt und sei T_n eine Schätzerfolge für ein statistisches Funktional $\kappa : \mathcal{P} \to \mathbb{R}$. Dann ist T_n asymptotisch linear in P_0 mit der Einflussfunktion $\chi \in L_2(P_0)$ mit $E_{P_0}(\chi) = 0$, falls gilt

$$\sqrt{n}(T_n - \kappa(P_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \chi(X_i) + o_{P_0}(1).$$

Ein Schätzer ist nun genau dann asymptotisch effizient, wenn er im kanonischen Gradienten asymptotisch linear ist.

Satz 3.12 (vgl. Van der Vaart [37], Lemma 25.23, S.367)
Seien $X_1, ..., X_n$ unabhängig identisch nach $P_0 \in \mathcal{P}$ verteilt. Sei $\kappa : \mathcal{P} \to \mathbb{R}$ differenzierbar in P_0 mit dem kanonischen Gradienten $\tilde{\kappa}$ und T_n eine Schätzerfolge für κ. Ist T_n asymptotisch linear in P_0 mit Einflussfunktion $\tilde{\kappa}$, das heißt wenn gilt

$$\sqrt{n}(T_n - \kappa(P_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\kappa}(X_i) + o_{P_0}(1), \tag{3.3}$$

dann ist T_n asymptotisch effizient in P_0.

Beweis. Die asymptotische Normalität folgt sofort mit dem zentralen Grenzwertsatz, denn es gilt

$$E(\tilde{\kappa}(X_i)) = \int \tilde{\kappa}\, dP_0 = 0, \quad Var(\tilde{\kappa}(X_i)) = \int \tilde{\kappa}^2\, dP_0 < \infty$$

und zusammen mit dem Lemma von Slutsky folgt

$$\mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\kappa}(X_i)\right) \xrightarrow[n\to\infty]{} N\left(0, \int \tilde{\kappa}^2\, dP_0\right).$$

Die Informationsschranke wird also angenommen.

Zu zeigen bleibt die Regularität von T_n. Sei dazu $\vartheta \mapsto P_\vartheta$ eine $L_2(P_0)$-differenzierbare Kurve in $\mathcal{P}$ mit Tangente g. Nach Satz 2.2 gilt

$$\log \prod_{i=1}^{n} \frac{dP_{1/\sqrt{n}}}{dP_0}(X_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i) - \frac{1}{2}\int g^2\, dP_0 + o_{P_0}(1).$$

Definiere jetzt die zweidimensionale Zufallsvariable

$$S_n = (Y_{n1}, Y_{n2}) := \left(\sqrt{n}(T_n - \kappa(P_0)), \log \prod_{i=1}^{n} \frac{dP_{1/\sqrt{n}}}{dP_0}(X_i)\right)$$

und die führenden linearen Terme

$$S_n' = (Y_{n1}', Y_{n2}') := \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\kappa}(X_i), \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i) - \frac{1}{2}\int g^2\, dP_0\right).$$

Wir bestimmen nun mit Hilfe des mehrdimensionalen zentralen Grenzwertsatzes die Grenzverteilung von S_n'. Nach dem allgemeinen Lemma von Slutsky entspricht diese wegen der stochastischen Konvergenz $S_n - S_n' \xrightarrow[n\to\infty]{} 0$ der Grenzverteilung von S_n.
Da die X_i i.i.d. sind, gilt

$$\lim_{n\to\infty} Cov(Y_{n1}', Y_{n2}') = \lim_{n\to\infty} E\left(\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\kappa}(X_i)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i) - \frac{1}{2}\int g^2\, dP_0\right)\right)$$

$$= \lim_{n\to\infty} E\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\kappa}(X_i)g(X_i)\right) + \frac{1}{\sqrt{n}}\left(-\frac{1}{2}\int g^2\, dP_0\right)\sum_{i=1}^{n} E(\tilde{\kappa}(X_i))$$

$$= E(\tilde{\kappa}(X_1)g(X_1)) = \int \tilde{\kappa} g\, dP_0 = \kappa'(P_0).$$

Nach dem mehrdimensionalen zentralen Grenzwertsatz und dem mehrdimensionalen Lemma von Slutsky, vgl. Witting, Müller-Funk [42], S.76, Korollar 5.84, gilt nun wegen $\int g\, dP_0 = 0$ und $\int g^2\, dP_0 < \infty$

$$\mathcal{L}(S_n|P_0^n) \xrightarrow[n\to\infty]{} \mathcal{N}\left(\begin{pmatrix} 0 \\ -\frac{1}{2}\int g^2\, dP_0 \end{pmatrix}, \begin{pmatrix} \int \tilde{\kappa}^2\, dP_0 & \kappa'(P_0) \\ \kappa'(P_0) & \int g^2\, dP_0 \end{pmatrix}\right).$$

Mit dem dritten Lemma von Le Cam, bzw. genauer nach Korollar 2.7, folgt

$$\mathcal{L}(\sqrt{n}(T_n - \kappa(P_0))|P_{\vartheta/\sqrt{n}}^n) \xrightarrow[n\to\infty]{} N(\kappa'(P_0), \int \tilde{\kappa}^2\, dP_0)$$

schwach. Weiterhin gilt

$$\sqrt{n}(T_n - \kappa(P_{1/\sqrt{n}})) = \sqrt{n}(T_n - \kappa(P_0) - (\kappa(P_{1/\sqrt{n}}) - \kappa(P_0)))$$
$$= \sqrt{n}(T_n - \kappa(P_0)) - \sqrt{n}(\kappa(P_{1/\sqrt{n}}) - \kappa(P_0))$$

und wegen der Differenzierbarkeit von κ

$$\sqrt{n}(\kappa(P_{1/\sqrt{n}}) - \kappa(P_0)) \xrightarrow[n\to\infty]{} \kappa'(P_0).$$

Daher folgt mit dem Lemma von Slutsky

$$\sqrt{n}(T_n - \kappa(P_{1/\sqrt{n}})) \xrightarrow[n\to\infty]{P_{1/\sqrt{n}}} Z - \kappa'(P_0) \sim N(0, \int \tilde{\kappa}^2 \, dP_0)$$

und T_n ist regulär. ■

Bemerkung 3.13

a) Wegen (3.3) heißt der kanonische Gradient $\tilde{\kappa}$ auch effiziente Einflussfunktion. (3.3) ist sogar notwendig für die asymptotische Effizienz, vgl. Van der Vaart [37], Lemma 25.23.

b) Im Rahmen der asymptotischen Theorie erwartungstreuer Schätzer mit existierenden Varianzen wurde die asymptotische Linearität im Zusammenhang mit der Fisher-Effizienz von Schätzern in Janssen [15] diskutiert.

Satz 3.12 liefert keinen konstruktiven Vorschlag für einen asymptotisch effizienten Schätzer, da $\tilde{\kappa}$ im Allgemeinen vom Fußpunkt P_0 abhängt und der Restterm $o_{P_0}(1)$ für jeden Fußpunkt ein anderer sein kann. Vielmehr liefert der Satz eine Entwicklung eines asymptotisch effizienten Schätzers und eine weitere Möglichkeit die asymptotische Effizienz für einen gegebenen Schätzer nachzuweisen. Ist in speziellen Fällen der kanonische Gradient unabhängig von P_0, so ist die Gestalt des asymptotischen Schätzers trivial.

Beispiel 3.14 (Fortsetzung Beispiel 3.5)
Sei $\mathcal{P} = \mathcal{P}^*$ das volle Modell, $P_0 \in \mathcal{P}$ und $h \in L_2(P_0)$ und $\kappa(P) = \int h \, dP$. Sei $D = D^* \subset T(P_0, \mathcal{P})$ die dichte Teilmenge der beschränkten Tangenten aus Lemma 1.9.
Gemäß Beispiel 3.5 ist κ dicht differenzierbar bezüglich D mit dem kanonischen Gradienten h.

Seien $X_1, ..., X_n$ i.i.d. Zufallsvariablen mit Verteilung $P^{X_1} = P_0$. Der kanonische Schätzer für κ ist die empirische Version des Funktionals

$$T_n(X_1, ..., X_n) = \sum_{i=1}^{n} h(X_i).$$

T_n ist asymptotisch effizient nach Satz 3.12, da der Schätzer (3.3) erfüllt, also asymptotisch linear in $\tilde{\kappa} = h$ ist. Der Restterm $o_{P_0}(1)$ verschwindet hier.

Kapitel 4

Asymptotische Effizienz von L-Schätzern

In diesem Kapitel soll die Theorie aus Kapitel drei nun explizit auf die Klasse der L-Funktionale angewendet werden. Für diese große Klasse von Funktionalen, in die beispielsweise der Median, aber auch der Erwartungswert hineinfallen, diskutieren wir die dichte Differenzierbarkeit, ermitteln die kanonischen Gradienten und bestimmen hiermit eine Informationsschranke.
In Spezialfällen wurde die Differenzierbarkeit von L-Funktionalen bereits diskutiert. Verwiesen wird auf die Arbeiten zur Frechet-Differenzierbarkeit von Boos [8] und Pfanzagl, Wefelmeyer [31], S.147, Lemma 5.4.3, und von Beutner, Zähle [5] für ein modifiziertes Konzept der Hadamard-Differenzierbarkeit. Beide Konzepte ersetzen nicht die dichte Differenzierbarkeit, sondern können etwa verwendet werden, um beispielsweise mit der Delta-Methode die zugehörigen L-Statistiken auf asymptotische Normalität zu untersuchen.
Wir weisen diese und die Regularität der L-Statistiken als kanonische Schätzer der L-Funktionale hier mit Hilfe eines Satzes von Shorack, Wellner [34] nach und bestimmen in diesem Zusammenhang die asymptotische Varianz. Da diese mit der Informationsschranke übereinstimmt, erbringen wir so den Nachweis der asymptotischen Effizienz dieser L-Schätzer. Dabei betrachten wir hier hauptsächlich L-Funktionale mit stetigen Gewichten. Für diskrete Gewichtungen siehe beispielsweise ausführlich von Müller [38], S.46 ff.

4.1 L-Funktionale und ihre kanonischen Schätzer

Sei $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen. Da L-Funktionale über die inverse Verteilungsfunktion definiert sind, ergibt sich über die empirischen Quantile sofort ein kanonischer Schätzer.

Definition 4.1 (L-Funktional und L-Schätzer)
ν sei ein endliches Maß auf $((0,1), \mathcal{B}(0,1))$, h eine reellwertige messbare Funktion auf $\mathbb{R}$ und $\mathcal{P}$ eine Familie von Wahrscheinlichkeitsmaßen, so dass für alle $P \in \mathcal{P}$ und zugehörigen Verteilungsfunktionen F mit Inversen F^{-1} gilt:

$$\int_0^1 |h(F^{-1}(u))| d\nu(u) < \infty. \tag{4.1}$$

Dann heißt das durch

$$\kappa(P) = \int_0^1 h(F^{-1}(u)) d\nu(u)$$

definierte Funktional $\kappa : \mathcal{P} \to \mathbb{R}$ L-Funktional zur Bewertungsfunktion h und Gewichtung ν. Ist ν absolut stetig mit $\lambda\!\!\!\lambda$-Dichte $b = \frac{d\nu}{d\lambda\!\!\!\lambda}$, so gilt

$$\kappa(P) = \int_0^1 h(F^{-1}(u)) b(u)\, du. \tag{4.2}$$

Seien nun $X_1, ..., X_n$ reelle i.i.d. Zufallsvariablen mit der unbekannten Verteilung P. Ersetzt man im L-Funktional die inverse Verteilungsfunktion F^{-1} durch das empirische Quantil $\hat{F}_n^{-1}$, vgl. Definition A.3, so ergibt sich der L-Schätzer

$$\hat{\kappa}_n = \hat{\kappa}_n(X_1, ..., X_n) := \int_0^1 h(\hat{F}_n^{-1}(u)) d\nu(u) \tag{4.3}$$

als kanonischer Schätzer des L-Funktionals.
Für eine diskrete Gewichtung $\nu = \sum_{i=1}^{m} c_i \varepsilon_{u_i}$ mit $c_i > 0, u_i \in (0,1)$ erhält man ein L-Funktional mit diskreten Gewichten, kurz diskretes L-Funktional,

$$\kappa(P) = \sum_{i=1}^{m} c_i h(F^{-1}(u_i))$$

mit dem zugehörigen kanonischen L-Schätzer

$$\hat{\kappa}_n = \sum_{i=1}^{m} c_i h(\hat{F}_n^{-1}(u_i)). \tag{4.4}$$

Im Spezialfall $m = 1$ ist κ ein (bewertetes) Quantil und $\hat{\kappa}_n$ ein (bewertetes) empirisches Quantil.

Es bezeichne ab jetzt $X_{i:n}$ die i-te Orderstatistik der nach der unbekannten Verteilung P verteilten reellen Zufallsvariablen $X_1, ..., X_n$. Da das empirische Quantil $\hat{F}_n^{-1}$ per Definition auf $\left(\frac{i-1}{n}, \frac{i}{n}\right]$ konstant gleich $X_{i:n}$ ist, lässt sich der kanonische L-Schätzer auch als L-Statistik, also als Linearkombination von (bewerteten) Orderstatistiken, darstellen.

Bemerkung 4.2 (L-Schätzer als L-Statistiken)
Sei $u \in (0,1)$. Dann gilt für das empirische Quantil

$$\begin{aligned}\hat{F}_n^{-1}(u) &= \sum_{i=1}^{n} X_{i:n} 1_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(u) = \sum_{i=1}^{n} X_{i:n} 1_{(i-1,i]}(nu) \\ &= \sum_{i=1}^{n} X_{i:n} 1_{\{i\}}(\lceil nu \rceil) = X_{\lceil nu \rceil : n}.\end{aligned}$$

Dementsprechend lässt sich der L-Schätzer (4.4) auf folgende Weise als L-Statistik schreiben:

$$\hat{\kappa}_n = \sum_{i=1}^{m} c_i h(\hat{F}_n^{-1}(u_i)) = \sum_{i=1}^{m} c_i h(X_{\lceil nu_i \rceil : n}).$$

Ist die Gewichtung ν hingegen absolut stetig mit λ-Dichte b, so ist der kanonische L-Schätzer (4.3) darstellbar als L-Statistik der Form

$$\begin{aligned}\hat{\kappa}_n &= \int_0^1 h(\hat{F}_n^{-1}(u)) d\nu(u) = \sum_{i=1}^{n} \int_{\frac{i-1}{n}}^{\frac{i}{n}} h(X_{i:n}) b(u)\, du \\ &= \frac{1}{n} \sum_{i=1}^{n} h(X_{i:n}) b_{ni} \quad \text{mit } b_{ni} := n \int_{\frac{i-1}{n}}^{\frac{i}{n}} b(u)\, du. \qquad (4.5)\end{aligned}$$

Bemerkung 4.3 (gemittelte und approximative Gewichte)
Die hier gewählte Definition der kanonischen L-Schätzer $\hat{\kappa}_n^m := \hat{\kappa}_n$ wird aufgrund obiger Darstellung auch L-Statistik mit gemittelten Gewichten genannt. Da nach Satz A.5 ein L-Funktional mit stetiger Verteilungsfunktion F auch geschrieben werden kann als

$$\kappa(P) = \int_0^1 h(F^{-1}(u)) b(u)\, du = \int h(x) b(F(x))\, dP(x), \qquad (4.6)$$

bietet sich als alternative Definition des kanonischen L-Schätzers auch

$$\hat{\kappa}_n^a := \int h(x) b(\hat{F}_n(x))\, d\hat{P}(x)$$

an, wobei $\hat{P}$ die zur empirischen Verteilungsfunktion $\hat{F}_n$ zugehörige diskrete Verteilung bezeichne. Im Unterschied zu Bemerkung 4.2 ergibt sich in anderer Darstellung die L-Statistik mit approximativen Gewichten

$$\begin{aligned}\hat{\kappa}_n^a &= \int h(x) b(\hat{F}_n(x))\, d\left(\sum_{i=1}^n \frac{1}{n} \varepsilon_{X_{i:n}}(x)\right) \\ &= \frac{1}{n}\sum_{i=1}^n h(X_{i:n}) b(\hat{F}_n(X_{i:n})) = \frac{1}{n}\sum_{i=1}^n h(X_{i:n}) b\left(\frac{i}{n}\right).\end{aligned}$$

Man beachte, dass $\hat{\kappa}_n^m$ und $\hat{\kappa}_n^a$ im Normalfall nicht übereinstimmen, da $\hat{F}_n$ nicht stetig ist und (4.6) für $F = \hat{F}_n$ nicht erfüllt ist, weil in dem Fall b nicht mit $b \circ F \circ F^{-1}$ übereinstimmen muss. Für $h \in L_2(P)$ und $b' \in L_2(\lambda\!\!\!\lambda_{|(0,1)})$ gilt jedoch immerhin

$$|\hat{\kappa}_n^m - \hat{\kappa}_n^a| = o(n^{-1/2}) \quad P - f.s.,$$

vgl. Witting, Müller-Funk [42], S.610.

Beispiel 4.4

Wichtige Beispiele für L-Funktionale sind für diskretes ν

der Median	$\kappa(P) = F^{-1}\left(\frac{1}{2}\right),$
die Quartilsmitte	$\kappa(P) = \frac{1}{2}\left(F^{-1}\left(\frac{1}{4}\right) + F^{-1}\left(\frac{3}{4}\right)\right),$
das Gastwirthfunktional	$\kappa(P) = \frac{3}{10}F^{-1}\left(\frac{1}{4}\right) + \frac{4}{10}F^{-1}\left(\frac{1}{2}\right) + \frac{3}{10}F^{-1}\left(\frac{3}{4}\right)$

und für absolut stetiges ν mit $\lambda\!\!\!\lambda$-Dichte $b = \frac{d\nu}{d\lambda\!\!\!\lambda}$

der Erwartungswert	$\kappa(P) = \int_0^1 F^{-1}(u)\, du = \int id\, dP,$
das α-getrimmte Mittel	$\kappa(P) = \frac{1}{1-2\alpha}\int_\alpha^{1-\alpha} F^{-1}(u)\, du,$
und Funktionale der Form	$\kappa(P) = \int_0^1 (F^{-1}(u))^\beta\, du = \int_0^1 x^\beta\, dP(x).$

Es stellt sich nun die Frage, wie gut der kanonische L-Schätzer ist bzw. ob er in gewisser Hinsicht optimal ist oder ob es einen besseren gibt. Dazu ist zunächst zu klären, was in diesem Zusammenhang als optimal angesehen wird. In dieser Arbeit wird die Frage nach der asymptotischen Effizienz gestellt, also das Grenzverhalten der L-Statistiken auf Optimalität hin untersucht. Eine vollkommen andere Fragestellung ist die nach der Konsistenz, der in dieser Arbeit zwar nicht weiter nachgegangen werden soll, aber zumindest für das empirische Quantil wird sie später zum Nachweis der asymptotischen Linearität verwendet und soll daher hier kurz angeführt werden.

Satz 4.5 (starke Konsistenz des empirischen Quantils)
Seien $X_j, j \in \mathbb{N}$ reelle i.i.d. Zufallsvariablen mit Verteilungsfunktion F und sei $u \in S(F^{-1})$. Dann gilt für $\frac{j_n}{n} = u + o(1)$

$$X_{j_n:n} \xrightarrow[n\to\infty]{} F^{-1}(u) \quad P-f.s..$$

Insbesondere gilt für das empirische u-Quantil

$$\hat{F}_n^{-1}(u) \xrightarrow[n\to\infty]{} F^{-1}(u) \quad P-f.s..$$

Beweis. Siehe Witting, Müller-Funk [42], S. 576, Satz 7.108. ■

Ziel ist es nun, die asymptotische Effizienz dieser Schätzer nachzuweisen. Gemäß Kapitel eins muss dazu das Funktional κ dicht differenzierbar sein und der Schätzer $\hat{\kappa}_n$ eine gewisse Regularität aufweisen.

4.2 Die kanonischen Gradienten der L-Funktionale

Wir betrachten im Folgenden das volle Modell $\mathcal{P}^*$ (1.1) aus Kapitel 1.3 mit der Zusatzbedingung (4.1), das L-Funktional $\kappa(P) = \int\limits_0^1 h(F^{-1}(u))b(u)\,du$ mit $\lambda\!\!\lambda$-Dichte b (4.2) und die dichte Teilmenge $D := D^{**} \subset T(P_0, \mathcal{P}^*)$ der beschränkten Tangenten mit kompaktem Träger, der vollständig im Inneren des Trägers von $P_0 \in \mathcal{P}^*$ liegt, aus Lemma 1.11.

In einem ersten Schritt zeigen wir jetzt die Differenzierbarkeit des Quantils $F^{-1}(u)$ für $u \in (0,1)$.

Lemma 4.6
Sei $\vartheta \mapsto P_\vartheta$ eine in P_0 L_1-differenzierbare Kurve in $\mathcal{P}^*$ mit Tangente $g \in L_1(\mu)$.

Seien $x_\vartheta, x_0 \in \mathbb{R}$ mit $x_\vartheta \to x_0$ für $\vartheta \to 0$. Dann gilt

$$\lim_{\vartheta \to 0} \frac{F_\vartheta(x_\vartheta) - F_0(x_\vartheta)}{\vartheta} = \int_{-\infty}^{x_0} g \, dP_0.$$

Beweis. Es gilt

$$\begin{aligned}
&\left| \frac{F_\vartheta(x_\vartheta) - F_0(x_\vartheta)}{\vartheta} - \int_{-\infty}^{x_0} g \, dP_0 \right| \\
&= \left| \frac{1}{\vartheta} \left(\int_{-\infty}^{x_\vartheta} dP_\vartheta - \int_{-\infty}^{x_\vartheta} dP_0 \right) - \int_{-\infty}^{x_\vartheta} g \, dP_0 + \int_{-\infty}^{x_\vartheta} g \, dP_0 - \int_{-\infty}^{x_0} g \, dP_0 \right| \\
&= \left| \int_{-\infty}^{x_\vartheta} \frac{1}{\vartheta} \left(\frac{dP_\vartheta}{d\mu} - \frac{dP_0}{d\mu} \right) - g \frac{dP_0}{d\mu} \, d\mu + \int_{-\infty}^{x_\vartheta} g \, dP_0 - \int_{-\infty}^{x_0} g \, dP_0 \right| \\
&\leq \int \left| \frac{1}{\vartheta} \left(\frac{dP_\vartheta}{d\mu} - \frac{dP_0}{d\mu} \right) - g \frac{dP_0}{d\mu} \right| d\mu + \left| \int_{-\infty}^{x_\vartheta} g \, dP_0 - \int_{-\infty}^{x_0} g \, dP_0 \right| \xrightarrow[\vartheta \to 0]{} 0.
\end{aligned}$$

■

Satz 4.7 (Differenzierbarkeit des Quantils)
Sei $P_0 \in \mathcal{P}^*$ und $\kappa(P) = F^{-1}(u)$ ein Quantil-Funktional für $u \in (0,1)$. Dann ist κ differenzierbar in P_0 und besitzt den kanonischen Gradienten

$$\tilde{\kappa}(x) = \frac{1}{f_0(F_0^{-1}(u))} (u - 1_{(0,u]}(F_0(x))). \tag{4.7}$$

Beweis. Setze $x_\vartheta := F_\vartheta^{-1}(u)$.
Sei $\vartheta \mapsto P_\vartheta$ eine in P_0 L_2-differenzierbare Kurve in $\mathcal{P}^*$ mit Tangente $g \in L_2^0(P_0)$. Dann gilt

$$\begin{aligned}
\frac{\kappa(P_\vartheta) - \kappa(P_0)}{\vartheta} &= \frac{x_\vartheta - x_0}{\vartheta} \\
&= \frac{F_0^{-1}(F_0(x_\vartheta)) - x_0}{F_0(x_\vartheta) - F_\vartheta(x_\vartheta)} \cdot \frac{F_0(x_\vartheta) - F_\vartheta(x_\vartheta)}{\vartheta} \\
&= \frac{F_0^{-1}(F_0(x_\vartheta)) - F_0^{-1}(u)}{F_0(x_\vartheta) - u} \cdot \frac{F_0(x_\vartheta) - F_\vartheta(x_\vartheta)}{\vartheta}.
\end{aligned}$$

Da F_0^{-1} stetig ist und $F_\vartheta \xrightarrow[n \to \infty]{} F_0$ auf $S(F_0)$, gilt $x_\vartheta \to x_0$ für $\vartheta \to 0$. Mit Lemma 4.6 folgt

$$\lim_{\vartheta \to 0} \frac{F_\vartheta(x_\vartheta) - F_0(x_\vartheta)}{\vartheta} = \int_{-\infty}^{x_0} g \, dP_0.$$

Da $x_\vartheta \to x_0$ für $\vartheta \to 0$ und F_0 stetig ist, gilt auch $F_0(x_\vartheta) \to F_0(x_0) = u$. Außerdem ist F_0^{-1} stetig differenzierbar in u und zusammen gilt

$$\frac{F_0^{-1}(F_0(x_\vartheta)) - F_0^{-1}(u)}{F_0(x_\vartheta) - u} \xrightarrow[\vartheta \to 0]{} (F_0^{-1})'(u) = \frac{1}{f_0(F_0^{-1}(u))}.$$

Insgesamt folgt

$$\begin{aligned} \lim_{\vartheta \to 0} \frac{\kappa(P_\vartheta) - \kappa(P_0)}{\vartheta} &= \frac{1}{f_0(F_0^{-1}(u))} \left(- \int_{-\infty}^{x_0} g \, dP_0 \right) \\ &= \int - \frac{1}{f_0(F_0^{-1}(u))} 1_{(-\infty, x_0]}(x) g(x) \, dP_0(x). \end{aligned}$$

Die Funktion

$$l(x) := - \frac{1}{f_0(F_0^{-1}(u))} 1_{(-\infty, x_0]}(x) = - \frac{1}{f_0(F_0^{-1}(u))} 1_{(0,u]}(F_0(x))$$

ist allerdings noch nicht der kanonische Gradient, da sie nicht in $L_2^0(P_0)$ liegt. Dieser ergibt sich aus der Zentrierung. Es gilt

$$\begin{aligned} E_{P_0}(l) &= \int - \frac{1}{f_0(F_0^{-1}(u))} 1_{(0,u]}(F_0(x)) \, dP_0(x) \\ &= - \frac{1}{f_0(F_0^{-1}(u))} \int_0^u d\lambda\!\!\!\lambda_{|(0,1)} = - \frac{u}{f_0(F_0^{-1}(u))} \end{aligned}$$

und folglich ist der kanonische Gradient durch

$$\tilde{\kappa}(x) = l(x) + \frac{u}{f_0(F_0^{-1}(u))} = \frac{1}{f_0(F_0^{-1}(u))} (u - 1_{(0,u]}(F_0(x)))$$

gegeben. ∎

Korollar 4.8

Sei h differenzierbar in $F_0^{-1}(u)$ mit Ableitung h' und $\kappa(P) = h(F^{-1}(u))$. Dann gilt nach der Kettenregel

$$\frac{d}{d\vartheta} h(F_\vartheta^{-1}(u))_{|\vartheta=0} = h'(F_0^{-1}(u)) \int \tilde{\kappa}(x) g(x) \, dP_0(x)$$

für $\tilde{\kappa}$ aus Satz 4.7 und das Funktional κ besitzt den kanonischen Gradienten

$$\tilde{\kappa}_h(x) = h'\left(F_0^{-1}(u)\right) \frac{1}{f_0(F_0^{-1}(u))} (u - 1_{(0,u]}(F_0(x))). \tag{4.8}$$

Bemerkung 4.9
Statt in P_0 kann analog auch die Ableitung des Funktionals in einem beliebigen Punkt P_ϑ entlang einer $L_2(P_\vartheta)$-differenzierbaren Kurve $\vartheta \mapsto P_\vartheta$ mit Tangente g_ϑ bestimmt werden und man erhält

$$\frac{d}{d\vartheta} h(F_\vartheta^{-1}(u)) = h'(F_\vartheta^{-1}(u)) \frac{1}{f_\vartheta(F_\vartheta^{-1}(u))} \int (u - 1_{(0,u]}(F_\vartheta(x))) g_\vartheta(x)\, dP_\vartheta(x).$$

Betrachten wir nun wieder unser L-Funktional mit stetigen Gewichten. Wir zeigen zunächst die Differenzierbarkeit des L-Funktionals in $\lambda\!\!\!\lambda_{|(0,1)}$.

Satz 4.10 (dichte Differenzierbarkeit in $\lambda\!\!\!\lambda_{|(0,1)}$)
Sei $P_0 = \lambda\!\!\!\lambda_{|(0,1)}$ die Gleichverteilung auf $(0,1)$. Sei h differenzierbar mit Ableitung h'. Sind außerdem
a) h beschränkt und $b = \frac{d\nu}{d\lambda\!\!\!\lambda}$ stetig differenzierbar mit beschränkter Ableitung b' oder
b) h' und b beschränkt,
dann ist κ dicht differenzierbar in P_0 bezüglich der Menge D und besitzt dort den kanonischen Gradienten

$$\tilde{\kappa}_1(x) = \int_0^1 h'(u)(u - 1_{(0,u]}(x))\, b(u)\, du.$$

Beweis. Sei $g \in D$. Zu zeigen ist nun die Differenzierbarkeit des Funktionals entlang den entsprechenden $L_2(P_0)$-differenzierbaren Kurven $\vartheta \mapsto P_\vartheta$ aus Lemma 1.13 für $|\vartheta| < \delta$. Sei also g beschränkt, etwas $|g| < c \in \mathbb{R}$ und

$$\frac{dP_\vartheta}{dP_0} = \frac{dP_\vartheta}{d\lambda\!\!\!\lambda_{|(0,1)}} = (1 + \vartheta g) 1_{(0,1)} =: f_\vartheta.$$

a) Es gilt

$$F_\vartheta(x) = \int_{-\infty}^{x} dP_\vartheta = F_0(x) + \vartheta \int_0^x g\, dP_0.$$

Sei $A \subset (0,1)$ der kompakte Träger von g, also $g(x) = 0\ \forall x \in A^c$.
Weiter gilt

$$\kappa(P_\vartheta) = \int_0^1 h(F_\vartheta^{-1}(u)) b(u)\, du = \int h(x) b(F_\vartheta(x))\, dP_\vartheta(x)$$

$$= \int_0^1 h(x) b(F_\vartheta(x))(1 + \vartheta g(x))\, dx$$

$$= \int_A h(x)b(F_\vartheta(x))(1+\vartheta g(x))\,dx + \int_{A^c} h(x)b(F_\vartheta(x))\,dx.$$

Wir betrachten zunächst den ersten Summanden. Es gilt

$$\frac{d}{d\vartheta}h(x)b\left(F_0(x)+\vartheta\int_0^x g\,dP_0\right)(1+\vartheta g(x))$$

$$= h(x)\left(b'(F_\vartheta(x))\left(\int_0^x g\,dP_0\right)(1+\vartheta g(x)) + b(F_\vartheta(x))g(x)\right).$$

Da h, b' und F_ϑ und somit auch $b \circ F_\vartheta$ stetig sind und stetige Funktionen auf kompakten Mengen beschränkt sind, sind h und $b \circ F_\vartheta$ beschränkt auf A.
Ebenso sind die Faktoren $\left|\int_0^x g\,dP_0\right| \leq \int |g|\,dP_0 \leq c$ und $|1+\vartheta g| \leq 1+\delta c$ beschränkt. Also darf nach Lemma A.12 beim ersten Integral Differentiation und Integration vertauscht werden.
Wir betrachten nun den zweiten Summanden. Es gilt

$$\frac{d}{d\vartheta}h(x)b\left(F_0(x)+\vartheta\int_0^x g\,dP_0\right) = h(x)b'(F_\vartheta(x))\int_0^x g\,dP_0.$$

Nun sind neben $\int_0^x g\,dP_0$ nach Voraussetzung auch h und b' beschränkt und es darf auch unter dem zweiten Integral differenziert werden.
Insgesamt darf beim gesamten Funktional also Differentiation und Integration vertauscht werden. Das heißt κ ist differenzierbar entlang dieser Kurve und mit partieller Integration und dem Satz von Fubini erhalten wir

$$\frac{d}{d\vartheta}\left(\int_0^1 h(F_\vartheta^{-1}(u))b(u)\,du\right)_{|\vartheta=0}$$

$$= \int_0^1 \frac{d}{d\vartheta}h(x)b\left(F_0(x)+\vartheta\int_0^x g\,dP_0\right)(1+\vartheta g(x))_{|\vartheta=0}\,dx$$

$$= \int_0^1 h(x)\left(b'(x)\int_0^x g\,dP_0 + b(x)g(x)\right)\,dx$$

$$= \int_0^1 h(x)\left(\frac{d}{dx}b(x)\int_0^x g(y)\,dy\right)\,dx$$

$$= \left[h(x)b(x)\int_0^x g(y)\,dy\right]_0^1 - \int_0^1 h'(x)b(x)\int_0^x g(y)\,dy\,dx$$

$$= \int \underbrace{\int_0^1 -h'(x)1_{(0,x]}(y)b(x)\,dx}_{=:l(y)}\; g(y)\,dP_0(y)$$

Analog zu Satz 4.7 ist l noch nicht der kanonische Gradient, sondern muss noch zentriert werden. Nach Fubini gilt

$$E_{P_0}(l) = \int_0^1 \int_0^1 -h'(u)1_{(0,u]}(y)b(u)\,du\,dy$$

$$= \int_0^1 -h'(u)b(u)\int_0^u dy\,du = \int_0^1 -h'(u)b(u)u\,du$$

und der kanonische Gradient ergibt sich zu

$$\tilde{\kappa}_1(x) = \int_0^1 -h'(u)1_{(0,u]}(x)b(u)\,du - \int_0^1 -h'(u)b(u)u\,du$$

$$= \int_0^1 h'(u)(u - 1_{(0,u]}(x))\,b(u)\,du.$$

b) Da h' und b als beschränkt gegeben sind, sei $|h'| \leq c_h$, $b \leq c_b$ und $c_1 := c_h c_b$. Da g beschränkt ist und $|\vartheta| < \delta$, gilt für δ ausreichend klein außerdem $1 + \vartheta g \geq c_2$ für ein $c_2 > 0$. Sei F_ϑ die Verteilungsfunktion von P_ϑ. Nach Bemerkung 4.9 und mit der Cauchy-Schwarz-Ungleichung gilt dann

$$\left|\frac{d}{d\vartheta}h(F_\vartheta^{-1}(u))b(u)\right|$$

$$= \left|h'(F_\vartheta^{-1}(u))\frac{1}{f_\vartheta(F_\vartheta^{-1}(u))}\int (u - 1_{(0,u]}(F_\vartheta(x)))g_\vartheta(x)\,dP_\vartheta(x)\,b(u)\right|$$

$$\leq \frac{c_1}{f_\vartheta(F_\vartheta^{-1}(u))}\left|\int (u - 1_{(0,u]}(F_\vartheta(x)))g_\vartheta(x)\,dP_\vartheta(x)\right|$$

$$\overset{CSU}{\leq} \frac{c_1}{f_\vartheta(F_\vartheta^{-1}(u))}\left(\int (u - 1_{(0,u]}(F_\vartheta(x)))^2\,dP_\vartheta(x)\right)^{\frac{1}{2}}\left(\int g_\vartheta^2\,dP_\vartheta\right)^{\frac{1}{2}}$$

$$
\begin{aligned}
&= \frac{c_1}{f_\vartheta(F_\vartheta^{-1}(u))} \left(\int (u - 1_{(0,u]}(y))^2 \, d\lambda\!\!\!\lambda_{|(0,1)}(y) \right)^{\frac{1}{2}} \left(\int \frac{g^2}{(1+\vartheta g)^2} (1+\vartheta g) \, dP_0 \right)^{\frac{1}{2}} \\
&= \frac{c_1}{f_\vartheta(F_\vartheta^{-1}(u))} \left(\int_0^1 u^2 - 2u 1_{(0,u]}(y) + 1_{(0,u]}(y) \, dy \right)^{\frac{1}{2}} \left(\int \frac{g^2}{1+\vartheta g} \, dP_0 \right)^{\frac{1}{2}} \\
&\le \frac{c_1}{f_\vartheta(F_\vartheta^{-1}(u))} \sqrt{u^2 - 2u^2 + u} \underbrace{\left(\frac{1}{c_2} \int g^2 \, dP_0 \right)^{\frac{1}{2}}}_{<\infty \text{ da } g \in L_2(P_0)} \\
&= \frac{c_1'}{\sqrt{c_2} f_\vartheta(F_\vartheta^{-1}(u))} \sqrt{u(1-u)}
\end{aligned}
$$

für

$$
c_1' := c_1 \left(\int g^2 \, dP_0 \right)^{\frac{1}{2}} \in \mathbb{R}.
$$

Auf $(0,1)$ ist $\sqrt{u(1-u)} \le 1$, also gilt

$$
\left| \frac{d}{d\vartheta} h(F_\vartheta^{-1}(u)) b(u) \right| \le \frac{c_1'}{\sqrt{c_2}} \frac{1}{f_\vartheta(F_\vartheta^{-1}(u))}. \tag{4.9}
$$

Weiterhin gilt wie oben für δ ausreichend klein

$$
f_\vartheta = 1 + \vartheta g \ge c_2.
$$

Somit gilt

$$
\left| \frac{d}{d\vartheta} h(F_\vartheta^{-1}(u)) b(u) \right| \le c_1' c_2^{-\frac{3}{2}} \in L_1(\lambda\!\!\!\lambda_{|(0,1)}).
$$

Also ist κ ist differenzierbar entlang dieser Kurve, es kann jetzt Integration und Differentiation vertauscht werden und nach Bemerkung 4.9 und dem Satz von Fubini gilt

$$
\begin{aligned}
&\frac{d}{d\vartheta} \kappa(P_\vartheta)_{|\vartheta=0} = \frac{d}{d\vartheta} \left(\int_0^1 h(F_\vartheta^{-1}(u)) b(u) \, du \right)_{|\vartheta=0} \\
&= \int_0^1 \frac{d}{d\vartheta} (h(F_\vartheta^{-1}(u)) b(u))_{|\vartheta=0} \, du \\
&= \int_0^1 \int \frac{h'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))} (u - 1_{(0,u]}(F_0(x))) g(x) \, dP_0(x) \, b(u) \, du
\end{aligned}
$$

$$= \int \underbrace{\int_0^1 h'(u)(u - 1_{(0,u]}(x))b(u)\,du}_{\tilde{\kappa}_1(x)}\; g(x)\,dP_0(x).$$

■

Liegt nun dichte Differenzierbarkeit des Funktionals in $\lambda\!\!\!\lambda_{|(0,1)}$ bezüglich D vor, so überträgt sie sich in folgender Weise bereits auf einen beliebigen Fußpunkt $P_0 \in \mathcal{P}^*$:

Satz 4.11

Ist κ in $Q_0 = \lambda\!\!\!\lambda_{|(0,1)}$ bezüglich der Menge D und den entsprechenden Kurven aus Lemma 1.13 dicht differenzierbar, dann ist κ auch in jedem beliebigen $P_0 \in \mathcal{P}^*$ mit Verteilungsfunktion F_0 dicht differenzierbar mit dem kanonischen Gradienten

$$\tilde{\kappa}(x) = \int_0^1 \frac{h'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}(u - 1_{(0,u]}(F_0(x)))\,b(u)\,du.$$

Beweis. Sei $D = D(Q_0, \mathcal{P}^*) \subset T(Q_0, \mathcal{P}^*)$ wie in Lemma 1.13 a), $g \in D$, also g beschränkt, und $\vartheta \mapsto Q_\vartheta$ die entsprechende $L_2(Q_0)$-differenzierbare Kurve auf $(0,1)$ mit Tangente $g : (0,1) \to \mathbb{R}$ aus Lemma 1.13 b), die gegeben ist durch

$$\frac{dQ_\vartheta}{dQ_0} = \frac{dQ_\vartheta}{d\lambda\!\!\!\lambda} = (1 + \vartheta g)1_{(0,1)}.$$

Sei nun $P_0 \in \mathcal{P}^*$ beliebig mit Verteilungsfunktion F_0. Betrachte jetzt die Kurve

$$\vartheta \mapsto P_\vartheta := Q_\vartheta^{F_0^{-1}}.$$

Diese besitzt gerade den Fußpunkt

$$Q_0^{F_0^{-1}} = \left(\lambda\!\!\!\lambda_{|(0,1)}\right)^{F_0^{-1}} = P_0$$

und ist L_2-differenzierbar in P_0 mit der Tangente $g \circ F_0$, denn es gilt

$$\frac{dQ_\vartheta^{F_0^{-1}}}{dQ_0^{F_0^{-1}}} = \frac{dQ_\vartheta}{dQ_0} \circ F_0$$

und somit

$$\left\| \frac{2}{\vartheta}\left(\left(\frac{dQ_\vartheta^{F_0^{-1}}}{dQ_0^{F_0^{-1}}}\right)^{\frac{1}{2}} - 1\right) - g \circ F_0 \right\|_{L_2\left(Q_0^{F_0^{-1}}\right)}$$

$$= \left\| \frac{2}{\vartheta}\left(\left(\frac{dQ_\vartheta}{dQ_0}\right)^{\frac{1}{2}} - 1\right) - g \right\|_{L_2(Q_0)} \xrightarrow[\vartheta\to 0]{} 0.$$

Da D dicht in $T(Q_0,\mathcal{P}^*)$ (bzgl. $L_2(Q_0)$) liegt, liegt die Menge $\{g\circ F_0 : g\in D\}$ dicht in $T(P_0,\mathcal{P}^*)$ (bzgl. $L_2(P_0)$), denn ist $\tilde g \in T(P_0,\mathcal{P}^*) = L_2^0(P_0)$, dann gilt

$$\int \tilde g \circ F_0^{-1}\, d\lambda\!\!\lambda_{|(0,1)} = \int \tilde g\, dP_0 = 0,$$

also $\tilde g \circ F_0^{-1} \in T(Q_0,\mathcal{P}^*) = L_2^0(\lambda\!\!\lambda_{|(0,1)})$ und es gibt eine Folge g_n in D mit $g_n \xrightarrow[n\to\infty]{} \tilde g\circ F_0^{-1}$ in $L_2(Q_0)$. Nun gilt

$$\|g_n\circ F_0 - \tilde g\|_{L_2(P_0)} = \left(\int (g_n\circ F_0 - \tilde g)^2\, dP_0\right)^{1/2}$$

$$= \left(\int (g_n - \tilde g\circ F_0^{-1})^2\, d\lambda\!\!\lambda_{|(0,1)}\right)^{1/2} \xrightarrow[n\to\infty]{} 0.$$

Zu zeigen bleibt nun die Differenzierbarkeit des Funktionals entlang der Kurven $\vartheta \mapsto P_\vartheta$. Zunächst gilt

$$\frac{dP_\vartheta}{dP_0} = \frac{dQ_\vartheta^{F_0^{-1}}}{dQ_0^{F_0^{-1}}} = \frac{dQ_\vartheta}{dQ_0}\circ F_0 = (1+\vartheta g)1_{(0,1)}\circ F_0.$$

Sei zudem F_ϑ die zu P_ϑ zugehörige Verteilungsfunktion. Dann gilt

$$F_\vartheta(x) := P_\vartheta((-\infty,x]) = Q_\vartheta^{F_0^{-1}}((-\infty,x]) = Q_\vartheta([0,F_0(x)])$$

und somit

$$\kappa(P_\vartheta) = \kappa(Q_\vartheta^{F_0^{-1}}) = \int_0^1 h(F_\vartheta^{-1}(u))b(u)\,du = \int h(x)b(F_\vartheta(x))\,dP_\vartheta(x)$$

$$= \int h(x)b(Q_\vartheta([0,F_0(x)]))(1+\vartheta g(F_0(x)))1_{(0,1)}(F_0(x))\,dP_0(x)$$

$$= \int_0^1 h(F_0^{-1}(u))b(Q_\vartheta([0,u]))(1+\vartheta g(u))\,du = \int_0^1 h(F_0^{-1}(u))b(Q_\vartheta([0,u]))\,dQ_\vartheta(u).$$

Sei G_ϑ die zugehörige Verteilungsfunktion von Q_ϑ. Dann gilt nach Voraussetzung

$$\frac{d}{d\vartheta}\int h(x)b(G_\vartheta(x))\,dQ_\vartheta(x)_{|\vartheta=0} = \frac{d}{d\vartheta}\int_0^1 h(G_\vartheta^{-1}(u))b(u)\,du_{|\vartheta=0}$$

$$= \int \tilde{\kappa}_1(x) g(x)\, dQ_0(x) = \int_0^1 \int_0^1 h'(u)(u - 1_{(0,u]}(x))\, b(u)\, du\; g(x)\, dx.$$

Mit dem Satz von Fubini folgt jetzt

$$\frac{d}{d\vartheta}\kappa(P_\vartheta)_{|\vartheta=0} = \frac{d}{d\vartheta}\left(\int_0^1 h(F_0^{-1}(u)) b(G_\vartheta(u))\, dQ_\vartheta(u)\right)_{|\vartheta=0}$$

$$= \int_0^1 \int_0^1 \left(\frac{d}{du} h(F_0^{-1}(u))\right)(u - 1_{(0,u]}(x)) b(u)\, du\; g(x)\, dx$$

$$= \int_0^1 \int_0^1 h'(F_0^{-1}(u)) \frac{1}{f_0(F_0^{-1}(u))}(u - 1_{(0,u]}(x)) b(u)\, du\; g(x)\, dx$$

$$= \int_0^1 \frac{h'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))} \int_0^1 (u - 1_{(0,u]}(x)) g(x)\, dx\; b(u)\, du$$

$$= \int_0^1 \frac{h'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))} \int (u - 1_{(0,u]}(F_0(x))) g(F_0(x))\, dP_0(x)\; b(u)\, du$$

$$= \int \underbrace{\int_0^1 \frac{h'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}(u - 1_{(0,u]}(F_0(x))) b(u)\, du}_{\tilde{\kappa}(x)}\; g(F_0(x))\, dP_0(x)$$

und die Behauptung ist gezeigt. ■

Aus der dichten Differenzierbarkeit der L-Funktionale und ihren kanonischen Gradienten $\tilde{\kappa}$ erhalten wir gemäß Satz 3.8 eine untere Schranke für die Varianz der Grenzverteilung L eines regulären Schätzers T_n für das L-Funktional κ mit

$$\mathcal{L}(\sqrt{n}(T_n - \kappa(P_0))|P_0) \xrightarrow[n\to\infty]{} L$$

schwach. Ein zentraler Grenzwertsatzes für Dreiecksschemata von L-Statistiken von Shorack, Wellner [34] liefert diese Konvergenz auch unter Alternativen, also insbesondere die Regularität dieser kanonischen Schätzer, und die entsprechende asymptotische Varianz, siehe Satz A.14. Durch einen Vergleich mit unserer Informationsschranke können wir so die asymptotische Effizienz der L-Schätzer nachweisen.

Satz 4.12 (Asymptotische Effizienz von L-Schätzern)
Sei $\mathcal{P}^*$ das volle Modell (1.1), κ ein L-Funktional mit absolut stetiger Gewichtung und $\lambda\!\!\!\lambda$-Dichte b (4.2) und zu beliebigem $P_0 \in \mathcal{P}^*$ sei $D := D^{**} \subset T(P_0, \mathcal{P}^*)$ die dichte Teilmenge der beschränkten Tangenten mit kompaktem Träger, der vollständig im Inneren des Trägers von $P_0 \in \mathcal{P}^*$ liegt, aus Lemma 1.11. Sei

$$T_n := \hat{\kappa}_n = \frac{1}{n}\sum_{i=1}^{n} h(X_{i:n})b_{ni} \quad \text{mit } b_{ni} := n\int_{\frac{i-1}{n}}^{\frac{i}{n}} b(u)\,du$$

die zu κ zugehörige L-Statistik (4.5). Sind dann die Voraussetzungen von Satz 4.10 und Satz A.14 erfüllt, insbesondere seien h' und b beschränkt, dann ist T_n ein asymptotisch effizienter Schätzer für κ.

Beweis. Sei $P_0 \in \mathcal{P}^*$ beliebig. Nach Satz 4.10 ist κ dicht differenzierbar in $\lambda\!\!\!\lambda_{(0,1)}$ bezüglich D und nach Satz 4.11 somit auch in P_0 mit dem kanonischen Gradienten

$$\tilde{\kappa}(x) = \int_0^1 \frac{h'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}(u - 1_{(0,u]}(F_0(x)))\,b(u)\,du.$$

Wir berechnen zunächst die Informationsschranke. Es gilt

$$\int \tilde{\kappa}^2\,dP_0 = \int\left(\int_0^1 \frac{h'(F_0^{-1}(u))}{f_0(F_0^{-1}(u))}(u - 1_{(0,u]}(F_0(x)))\,b(u)\,du\right)^2 dP_0(x)$$

$$= \int\int_0^1\int_0^1 \frac{h'(F_0^{-1}(u_1))h'(F_0^{-1}(u_2))}{f_0(F_0^{-1}(u_1))f_0(F_0^{-1}(u_2))}(u_1u_2 - u_1 1_{(0,u_2]}(F_0(x)) - u_2 1_{(0,u_1]}(F_0(x))$$

$$+1_{(0,u_1]}(F_0(x))1_{(0,u_2]}(F_0(x)))\,b(u_1)b(u_2)\,du_1\,du_2\,dP_0(x)$$

$$= \int_0^1\int_0^1 \frac{h'(F_0^{-1}(u_1))h'(F_0^{-1}(u_2))}{f_0(F_0^{-1}(u_1))f_0(F_0^{-1}(u_2))}\,b(u_1)b(u_2)$$

$$\int u_1u_2 - u_1 1_{(0,u_2]}(F_0(x)) - u_2 1_{(0,u_1]}(F_0(x)) + 1_{(0,u_1\wedge u_2]}(F_0(x))\,dP_0(x)\,du_1\,du_2$$

$$= \int_0^1\int_0^1 \frac{h'(F_0^{-1}(u_1))h'(F_0^{-1}(u_2))}{f_0(F_0^{-1}(u_1))f_0(F_0^{-1}(u_2))}\,b(u_1)b(u_2)$$

$$\int u_1u_2 - u_1 1_{(0,u_2]}(x) - u_2 1_{(0,u_1]}(x) + 1_{(0,u_1\wedge u_2]}(x)\,d\lambda\!\!\!\lambda_{|(0,1)}(x)\,du_1\,du_2$$

$$= \int_0^1\int_0^1 \frac{h'(F_0^{-1}(u_1))h'(F_0^{-1}(u_2))}{f_0(F_0^{-1}(u_1))f_0(F_0^{-1}(u_2))}\,b(u_1)b(u_2)(u_1u_2 - u_1u_2 - u_2u_1 + u_1\wedge u_2)\,du_1\,du_2$$

$$= \int\int \frac{h'(y)h'(z)}{f_0(y)f_0(z)}\, b(F_0(y))b(F_0(z))(F_0(y) \wedge F_0(z) - F_0(y)F_0(z))\, dP_0(y)\, dP_0(z)$$
$$= \int\int h'(y)h'(z)\, b(F_0(y))b(F_0(z))(F_0(y) \wedge F_0(z) - F_0(y)F_0(z))\, dy\, dz.$$

Sei nun $g \in D$ und $\vartheta \mapsto P_\vartheta$ eine beliebige $L_2(P_0)$-differenzierbare Kurve lokal in $\mathcal{P}^*$ mit Tangente g.
Sei zudem $X_{ni}, 1 \leq i \leq n$ ein Dreiecksschema von zeilenweise i.i.d. Zufallsvariablen mit Verteilung $X_{n1} \sim P_{\frac{\vartheta}{\sqrt{n}}}$ und zugehöriger Verteilungsfunktion $F_{\frac{\vartheta}{\sqrt{n}}}$.
Setze jetzt $J_n(t) := \sum_{i=1}^{n} b_{ni} 1_{(\frac{i-1}{n}, \frac{i}{n}]}(t)$ und F_{ni}, $\overline{F}_n$, $\overline{F}_n^{-1}$, G_{ni}, $K_n(s,t)$, g_n, μ_n und σ_n^2 wie in Satz A.14.
Dann gilt $F_{ni} \equiv F_{\frac{\vartheta}{\sqrt{n}}}$ und folglich auch $\overline{F}_n = F_{\frac{\vartheta}{\sqrt{n}}}$ und $\overline{F}_n^{-1} = F^{-1}_{\frac{\vartheta}{\sqrt{n}}}$ und somit $G_{ni} = F_{\frac{\vartheta}{\sqrt{n}}} \circ F^{-1}_{\frac{\vartheta}{\sqrt{n}}} = id$, $K_n(s,t) = s \wedge t - \frac{1}{n}\sum_{i=1}^{n} G_{ni}(s)G_{ni}(t) = s \wedge t - st = \Gamma(s,t) =: K(s,t)$ und $g_n = h \circ F^{-1}_{\frac{\vartheta}{\sqrt{n}}}$.
Weiterhin gilt

$$J_n(t) \xrightarrow[n\to\infty]{} b(t) \text{ auf } (0,1),$$

$$\mu_n = \int_0^1 g_n(t)b(t)\, dt = \int_0^1 h(F^{-1}_{\frac{\vartheta}{\sqrt{n}}}(t)b(t)\, dt = \kappa(P_{\frac{\vartheta}{\sqrt{n}}})$$

und

$$\begin{aligned}\sigma_n^2 &= \int_0^1\int_0^1 K(s,t)b(s)b(t)\, dg_n(s)\, dg_n(t)\\ &= \int\int h'(F^{-1}_{\frac{\vartheta}{\sqrt{n}}}(s))h'(F^{-1}_{\frac{\vartheta}{\sqrt{n}}}(t))\Gamma(s,t)b(s)b(t)\, dF^{-1}_{\frac{\vartheta}{\sqrt{n}}}(s)\, dF^{-1}_{\frac{\vartheta}{\sqrt{n}}}(t)\\ &= \int\int h'(y)h'(z)\Gamma(F_{\frac{\vartheta}{\sqrt{n}}}(y), F_{\frac{\vartheta}{\sqrt{n}}}(z))b(F_{\frac{\vartheta}{\sqrt{n}}}(y))b(F_{\frac{\vartheta}{\sqrt{n}}}(z))\, dy\, dz\end{aligned}$$

mit

$$\Gamma(s,t) := s \wedge t - st.$$

Aus der schwachen Konvergenz $P_{\frac{\vartheta}{\sqrt{n}}} \xrightarrow[n\to\infty]{} P_0$, vgl. Lemma 1.2 b), folgt mit dem Satz von Helly-Bray, da F_0 stetig ist, die Konvergenz von $F_{\frac{\vartheta}{\sqrt{n}}} \xrightarrow[n\to\infty]{} F_0$. Da b und h' beschränkt sind, folgt mit dem Satz von der dominierten Konvergenz:

$$\lim_{n\to\infty} \sigma_n^2 = \int\int \lim_{n\to\infty} h'(y)h'(z)\Gamma(F_{\frac{\vartheta}{\sqrt{n}}}(y), F_{\frac{\vartheta}{\sqrt{n}}}(z))b(F_{\frac{\vartheta}{\sqrt{n}}}(y))b(F_{\frac{\vartheta}{\sqrt{n}}}(z))\, dy\, dz$$

$$= \int \int h'(y)h'(z)\Gamma(F_0(y), F_0(z))b(F_0(y))b(F_0(z))\, dy\, dz = \sigma^2 > 0.$$

Mit Satz A.14 folgt nun

$$\frac{\sqrt{n}(T_n - \kappa(P_{\frac{\vartheta}{\sqrt{n}}}))}{\sigma_n} \xrightarrow[n\to\infty]{\mathcal{D}} Z \sim N(0,1)$$

und mit dem Satz von Slutsky folgt

$$\sqrt{n}(T_n - \kappa(P_{\frac{\vartheta}{\sqrt{n}}})) = \frac{\sigma_n}{\sigma_n}\sqrt{n}(T_n - \kappa(P_{\frac{\vartheta}{\sqrt{n}}})) \xrightarrow[n\to\infty]{\mathcal{D}} \sigma Z \sim N(0, \sigma^2).$$

Also ist der L-Schätzer T_n regulär bezüglich der dichten Teilmenge $D \subset T(P_0, \mathcal{P}^*)$, insbesondere asymptotisch normal und wir stellen fest, dass $\|\tilde{\kappa}\|^2 = \sigma^2$ gilt. Somit nimmt T_n die Informationsschranke an und ist gemäß Satz 3.8 asymptotisch effizient.

■

Bemerkung 4.13

Die L-Statistiken mit gemittelten Gewichten b_{ni} können bei der Betrachtung des Grenzverhaltens durch die L-Statistiken mit approximativen Gewichten

$$\hat{\kappa}_n^a = \frac{1}{n}\sum_{i=1}^{n} b_{ni}^a h(X_{i:n}) \quad \text{mit } b_{ni}^a := b\left(\frac{i}{n}\right)$$

ersetzt werden. Denn nach Bemerkung 4.3 gilt dann

$$|\hat{\kappa}_n - \hat{\kappa}_n^a| = o(n^{-1/2}) \quad P - f.s.$$

und damit auch

$$\sqrt{n}(\hat{\kappa}_n - \hat{\kappa}_n^a) \xrightarrow[n\to\infty]{} 0 \quad \text{stochastisch}$$

und mit dem Satz von Slutsky folgt die Äquivalenz

$$\sqrt{n}(\hat{\kappa}_n - \kappa(P)) \xrightarrow[n\to\infty]{\mathcal{D}} Z \Leftrightarrow \sqrt{n}(\hat{\kappa}_n^a - \kappa(P)) \xrightarrow[n\to\infty]{\mathcal{D}} Z.$$

Kapitel 5

Asymptotische Effizienz in der Survival Analysis

In diesem Kapitel betrachten wir als weitere Anwendung der Theorie aus Kapitel eins bis drei Funktionale in der Survival Analysis und ihre Schätzer. Wir wiederholen dabei zunächst einige Grundbegriffe und Konzepte der Survival Analysis und leiten auch ein bekanntes Modell für rechtszensierte Daten her, das auf Gill [11] zurückgeht, siehe auch Andersen, Borgan, Gill, Keiding [2], von nun an durch ABGK abgekürzt, sowie Klein, Moeschberger [3]. Anschließend befassen wir uns mit der Differenzierbarkeit bestimmter Funktionale in diesem Zusammenhang sowohl im unzensierten als auch im rechtszensierten Fall. Mit Hilfe unserer Konzepte können wir auch für solche Funktionale Informationsschranken ableiten. Wichtige Beispiele stellen die Survivalfunktion und die kumulative Hazardfunktion dar, die gemeinhin durch den Kaplan-Meier-Schätzer respektive den Nelson-Aalen-Schätzer geschätzt werden. Auf diese Weise können wir hier erneut die asymptotische Effizienz dieser beiden Schätzer explizit nachweisen.

5.1 Grundbegriffe und bekannte Resultate der Survival Analysis

Im Folgenden sei X die Zeit bis zu einem bestimmten Ereignis (Überlebenszeit) mit stetiger Verteilung P. Sei F die stetige Verteilungsfunktion von X. Dann bezeichnet

$$S(x) := 1 - F(x)$$

die Survivalfunktion.
Zur Motivation besitze P die Lebesguedichte f. Die Hazardrate λ wird definiert duch

$$\lambda(t) := \lim_{\Delta\to 0} \frac{P(X \in [t, t+\Delta)|X \geq t)}{\Delta},$$

das heißt λ ist die Rate, mit der das relevante Ereignis zum Zeitpunkt t stattfindet unter der Bedingung, dass es bis zum Zeitpunkt t noch nicht stattgefunden hat. Hazardraten brauchen jedoch nicht zu existieren.
Allgemein, also unabhängig von der Existenz von Dichten und Hazardraten lässt sich aber durch

$$\frac{d\Lambda}{dF} := \frac{1}{S}$$

das unbeschränkte Hazardmaß Λ definieren. Im Falle der Existenz von Dichten und Hazardraten gilt dann

$$\lambda = \frac{d\Lambda}{d\lambda\!\!\lambda} = \frac{f}{S}.$$

Die kumulative Hazardfunktion ist gegeben durch

$$\Lambda(x) := \Lambda((-\infty, x]) = \int_{-\infty}^{x} \lambda(t)\, dt = -\log(S(x))$$

bzw. umgekehrt

$$S(x) = \exp(-\Lambda(x)).$$

Betrachten wir nun erneut (einparametrige) Familien von Wahrscheinlichkeitsmaßen $\{P_\vartheta : \vartheta \in \Theta\}$, so heißt

$$\frac{d\Lambda_\vartheta}{d\Lambda_0} = \frac{dP_\vartheta}{dP_0}\frac{1-F_0}{1-F_\vartheta}$$

der Hazardratenquotient von P_ϑ bezüglich P_0.
Auf Seiten der Wahrscheinlichkeitsmaße bzw. Dichten haben wir für L_1- bzw. L_2-differenzierbare Kurven $\vartheta \mapsto P_\vartheta$ Ableitungen der Dichtequotienten (Tangenten) vorliegen. Auf Seiten der Hazards bezeichnen wir die Ableitungen der Hazardratenquotienten

$$\gamma := \frac{d}{d\vartheta}\frac{d\Lambda_\vartheta}{d\Lambda_0}_{|\vartheta=0} = \lim_{\vartheta\to 0}\frac{1}{\vartheta}\left(\frac{\lambda_\vartheta}{\lambda_0} - 1\right)$$

entsprechend als Hazardratenableitungen.
Für diese gilt nach Janssen [17], Lemma 2.1 für eine L_2-differenzierbare Kurve mit Tangente g in P_0 und zugehörigen Hazardmaßen Λ_ϑ

$$\frac{1}{\vartheta}\left(\frac{d\Lambda_\vartheta}{d\Lambda_0} - 1\right) \to \gamma = R(g)$$

P_0-stochastisch für $\vartheta \to 0$ und auf $(-\infty, a]$ für $F_0(a) < 1$ zumindest auch in $L_1(P_0)$ mit

$$R(g)(x) := g(x) - \int_x^\infty g\, dP_0 \frac{1}{1 - F_0(x)}. \tag{5.1}$$

Der Operator $R : L_2^0(P_0) \to L_2(P_0)$ bildet also eine Tangente auf die zugehörige Hazardratenableitung ab. Janssen [17],[19], Efron, Johnstone [10] bzw. Ritov, Wellner [33] haben gezeigt, dass R eine Hilbertraumisometrie ist mit der Umkehrabbildung

$$L(\gamma)(x) := \gamma(x) - \int_{-\infty}^x \frac{\gamma}{1 - F_0}\, dP_0.$$

Das heißt R ist linear, bijektiv und erhält das Skalarprodukt auf $L_2^0(P_0)$.
Darüber hinaus gilt für $g, g', \gamma \in L_2(P_0)$:

$$\int L(\gamma)\, dP_0 = 0,$$

$$R(L(\gamma)) = \gamma \ P_0\text{-f.s.},$$

$$L(R(g)) = g - E(g),$$

$$\langle \gamma, R(g) \rangle = \langle L(\gamma), g \rangle, \text{ falls die Integrale für } |\gamma| \text{ und } |g| \text{ endlich sind und}$$

$$\langle R(g), R(g') \rangle = \langle g - E(g), g' - E(g') \rangle, \tag{5.2}$$

vgl. Janssen [17], Lemma 2.2, S.357.
Der Operator R stellt somit ein nützliches Instrument dar, um von Tangenten zu Hazardratenableitungen überzugehen und umgekehrt, worauf wir in Abschnitt 5.4 zurückgreifen werden.

Beispiel 5.1 (proportionales Hazardratenmodell)
Sei $Y \sim exp(1)$ standard-exponentialverteilt und $X := \exp(-\vartheta)Y$, $\vartheta \in \mathbb{R}$, $P_\vartheta = P^X$. Dann gilt

$$F_\vartheta(x) := P_\vartheta(X \le x) = P(Y \le \exp(\vartheta)x) = 1 - \exp(-x\exp(\vartheta))1_{(0,\infty)}(x) \text{ und}$$

$$f_\vartheta(x) = \exp(-x\exp(\vartheta))\exp(\vartheta)1_{(0,\infty)}(x).$$

Die Kurve ist eine Exponentialfamilie in der Statistik $T(x) = -x$ und somit $L_2(P_0)$-differenzierbar mit der Tangente

$$g(x) = T(x) - E_0(T) = 1 - x.$$

Die Hazardrate

$$\lambda_\vartheta(x) = \frac{f_\vartheta(x)}{1 - F_\vartheta(x)} = \exp(\vartheta)$$

ist hier konstant, man spricht von einem proportionalen Hazardratenmodell. Es ergibt sich die Hazardratenableitung

$$\frac{d}{d\vartheta} \log(\lambda_\vartheta)_{|\vartheta=0} = \frac{d}{d\vartheta} \vartheta_{|\vartheta=0} = 1.$$

Nach (5.1) ergibt sich ebenso

$$\begin{aligned} R(g)(x) &= 1 - x - \int_x^\infty (1-y)\, dP_0(y) \frac{1}{1 - F_0(x)} \\ &= 1 - x - \int_x^\infty (1-y) \exp(-y)\, dy \frac{1}{\exp(-x)} = 1 - x + x \exp(-x) \exp(x) = 1 \end{aligned}$$

Im Folgenden bezeichne bei gegebenen Wahrscheinlichkeitsmaßen P, P_0, P_ϑ stets $\Lambda, \Lambda_0, \Lambda_\vartheta$ bzw. $\lambda, \lambda_0, \lambda_\vartheta$ die entsprechend zugehörigen Hazardmaße bzw. Hazardraten.

5.2 Rechtszensierte Daten

Sei T eine Überlebenszeit (bzw. Zeit bis zu einem bestimmten Ereignis). In der Praxis können wir Realisierungen der Überlebenszeit oft nicht direkt beobachten, sondern T ist zufällig zensiert. Sei C eine von T stochastisch unabhängige Zufallsvariable (Zensierung). Dann ist oft nur die Zufallsvariable (X, Δ) mit

$$X := \min(T, C), \quad \Delta = 1_{\{T \leq C\}}$$

beobachtbar. Das heißt wir können die Zeit beobachten, zu der das Objekt aus der Studie ausfällt und gleichzeitig, ob dies durch Zensierung oder Tod (bzw. Eintritt des Ereignisses) stattgefunden hat, wissen aber im Falle einer Zensierung nicht, wie lange das Obejekt ohne Zensierung überlebt hätte (bzw. die Zeit bis zum Ereignis ohne Zensierung).
In der Praxis tritt dieser Fall etwa auf, wenn Patienten eine Studie vorzeitig verlassen oder die Studie vorbei ist, bevor alle Ereignisse eingetreten sind.

Sei $\mathcal{P} = \mathcal{P}^*$ (1.1) die Menge der für die Überlebenszeit T in Frage kommenden Verteilungen P^T und $P^C \in \mathcal{P}$ eine unbekannte, aber feste Verteilung der Zensierungsvariablen. Sei

$$\mathcal{Q} = \{\mathcal{L}((X, \Delta) | P^T \otimes P^C) : P^T \in \mathcal{P}\}$$

das Modell der Verteilungen der beobachtbaren Zufallsvariablen (X, Δ).
$Q \in \mathcal{Q}$ ist eindeutig bestimmt durch die Subverteilungen

$$Q((-\infty, x] \times \{1\}) = P(X \leq x, \Delta = 1) = P(T \leq x, T \leq C)$$

$$= \int_{-\infty}^{x} \int_{a}^{\infty} dP^C(b)\, dP^T(a) = \int_{-\infty}^{x} 1 - F_C\, dP^T,$$

$$Q((-\infty, x] \times \{0\}) = P(X \leq x, \Delta = 0) = P(C \leq x, T > C)$$

$$= \int_{-\infty}^{x} \int_{b}^{\infty} dP^T(a)\, dP^C(b) = \int_{-\infty}^{x} 1 - F_T\, dP^C,$$

wobei F_T, F_C die Verteilungsfunktionen von P^T bzw. P^C bezeichnen.
Seien $f_T := \frac{dP^T}{d\mu}, f_C := \frac{dP^C}{d\mu}$ die zugehörigen Dichten. Für $x \in \mathbb{R}, \delta \in \{0, 1\}$ ergibt sich

$$Q((-\infty, x] \times \{\delta\}) = \delta \int_{-\infty}^{x} 1 - F_C\, dP^T + (1 - \delta) \int_{-\infty}^{x} 1 - F_T\, dP^C$$

$$= \int_{(-\infty,x]\times\{0,1\}} \delta(1 - F_C(y))f_T(y) + (1 - \delta)(1 - F_T(y))f_C(y)\, d\mu \otimes \#(y, \delta),$$

wobei $\#$ das Zählmaß auf $\{0, 1\}$ bezeichne. Also besitzt Q die Dichte

$$\frac{dQ}{d\mu \otimes \#}(x, \delta) = \delta(1 - F_C(y))f_T(y) + (1 - \delta)(1 - F_T(y))f_C(y),$$

vgl. z.B. Van der Vaart [36], S.104, oder Janssen, Werft [20], S.12.

Für $h \in L_1(Q)$ und $A \in \mathcal{B}$ gilt somit

$$\int_{A\times\{0,1\}} h(x, \delta)\, dQ(x, \delta)$$

$$= \int \int_A h(x, \delta)\delta(1 - F_C(x))f_T(x) + h(x, \delta)(1 - \delta)(1 - F_T(x))f_C(x)\, d\mu \otimes \#(x, \delta)$$

$$= \int_A h(x, 1)(1 - F_C(x))\, dP^T(x) + \int_A h(x, 0)(1 - F_T(x))\, dP^C(x). \tag{5.3}$$

Sei nun $P_0 \in \mathcal{P}, Q_0 = \mathcal{L}((X, \Delta)|P_0 \otimes P^C)$.
Für eine $L_2(P_0)$-differenzierbare Kurve $\vartheta \mapsto P_\vartheta$ lokal in $\mathcal{P}$ mit der Tangente $g \in K(P_0, \mathcal{P})$ ist die Kurve

$$\vartheta \mapsto Q_\vartheta = \mathcal{L}((X, \Delta)|P_\vartheta \otimes P^C)$$

lokal in $\mathcal{Q}$ ebenfalls L_2-differenzierbar mit der Tangente

$$\tilde{g}(x,\delta) := \delta g(x) + (1-\delta)\overline{R}(g)(x)$$

für

$$\overline{R}(g) := g - R(g) = \int_x^\infty g\, dP_0 \frac{1}{1-F_0(x)}.$$

Dies folgt beispielsweise aus Van der Vaart [36], S.107, oder Janssen [17], Lemma 2, [18], S.358.

So wie durch $R : L_2^0(P_0) \to L_2(P_0)$ eine Isometrie mit Umkehrfunktion L auf dem Tangentialraum von $P_0 \in \mathcal{P}$ gegeben ist, so existiert auch auf $L_2^0(Q_0)$ eine Isometrie gegeben durch

$$\tilde{R}(\tilde{g})(x,\delta) = \delta R(g)(x)$$

mit der Umkehrfunktion

$$\tilde{L}(\tilde{\gamma})(x,\delta) = \delta L(\gamma)(x) - (1-\delta)\overline{L}(\gamma)(x) = \delta\gamma(x) - \overline{L}(\gamma)(x)$$

für

$$\overline{L}(\gamma) := \gamma - L(\gamma) = \int_{-\infty}^x \frac{\gamma}{1-F_0}\, dP_0.$$

Analog zu (5.2) gilt für $\tilde{\gamma}, \tilde{g} \in L_2(Q_0)$:

$$\langle \tilde{\gamma}, \tilde{R}(\tilde{g}) \rangle = \langle \tilde{L}(\tilde{\gamma}), \tilde{g} \rangle, \tag{5.4}$$

falls eine der beiden Seiten für $|\tilde{\gamma}|, |\tilde{g}|$ existiert, vgl. Janssen [18], Theorem 3.1.

5.3 Kaplan-Meier- und Nelson-Aalen-Schätzer

Wir führen nun noch zwei bekannte Schätzer der wichtigsten Größen in der Survival Analysis ein, wobei wir uns zunächst allgemein halten und Bindungen der X_i zulassen. Werden die X_i über einen stetigen Zeitraum beobachtet und sind die Verteilungsfunktionen F_X stetig, so treten (P^X-fast-sicher) keine Bindungen auf und die Schätzer lassen sich leicht vereinfacht darstellen.
Seien $X_1, ..., X_n$ unabhängig identisch nach P^X verteilt und s_i Realisierungen von X_i, $1 \leq i \leq n$. Seien $t_1 < t_2 < ... < t_m$ die verschiedenen Zeitpunkte, zu denen Ereignisse eintreten, also angeordnete Realisierungen von X_i mit $\Delta_i = 1$, und sei δ_i

die Anzahl der Ereignisse zum Zeitpunkt t_i. Sei außerdem Y_i die Anzahl der Objekte unter Risiko zum Zeitpunkt t_i, das heißt

$$Y_i := \sum_{j=1}^{n} 1_{[t_i,\infty)}(X_j)$$

also die Anzahl der Objekte, die den Zeitpunkt t_i erleben.

Der Standard-Schätzer für die Survivalfunktion $S(t) = 1 - F_T(t)$ unter (möglicherweise) rechtszensierten Daten ist der nach seinen Schöpfern benannte und bereits 1958 eingeführte Kaplan-Meier-Schätzer [23] oder auch Produkt-Limit-Schätzer

$$\hat{S}_{n,KM}(t) := \begin{cases} 1 & , \quad t < t_1 \\ \prod\limits_{t_i \le t} (1 - \frac{\delta_i}{Y_i}) & , \quad t_1 \le t \end{cases} , \tag{5.5}$$

vgl. Klein, Moeschberger [3], S.84. Er ist der Maximum-Likelihood-Schätzer für $S(t)$, vgl. Johansen [22].
Wegen

$$\Lambda(t) = -\log S(t) \tag{5.6}$$

kann der Produkt-Limit-Schätzer auch zur Schätzung der kumulativen Hazardfunktion verwendet werden. Es ergibt sich der Schätzer

$$\hat{\Lambda}_n(t) = -\log \hat{S}_{n,KM}(t).$$

Da (5.6) jedoch nur im stetigen Fall, also für stetige Verteilungsfunktionen F bzw. Survivalfunktionen S erfüllt ist und die empirische Version $\hat{S}_{n,KM}$ offensichtlich nicht stetig ist, basiert der Standard-Schätzer für die kumulative Hazardfunktion jedoch nicht auf dem Kaplan-Meier-Schätzer. Dieser Standard-Schätzer, der Nelson-Aalen-Schätzer, wurde zum ersten Mal 1972 von Nelson [28] vorgeschlagen und 1978 von Aalen [1] wiederentdeckt und über Zählprozesse bzw. Produktintegration neu hergeleitet, vgl. auch Gill, Johansen [12]. Er zeigt ein besseres Verhalten bei kleinen Stichprobenumfängen n und hat folgende Gestalt:

$$\hat{\Lambda}_{n,NA}(t) =:= \begin{cases} 0 & , \quad t < t_1 \\ \sum\limits_{t_i \le t} \frac{\delta_i}{Y_i} & , \quad t_1 \le t \end{cases} , \tag{5.7}$$

vgl. Klein, Moeschberger [3], S.86.
Via

$$S(t) = \exp(-\Lambda(t))$$

kann natürlich umgekehrt auch die Survivalfunktion alternativ zum Produkt-Limit-Schätzer geschätzt werden durch

$$\hat{S}_n = \exp(-\hat{\Lambda}_{n,NA}(t)),$$

wobei man den selben Fehler begeht wie oben beschrieben.
Da das in diesem Kapitel betrachtete volle Modell nur stetige Verteilungsfunktionen F_X umfasst, hier also keine Bindungen auftreten, gibt es P^X-fast sicher n verschiedene Zeitpunkte $t_1 < t_2 < ... < t_n$ und $\delta_i \in \{0, 1\}$ gibt lediglich an, ob die jeweiligen Daten zensiert sind oder nicht. Die Anzahl der Objekte unter Risiko vereinfacht sich zu

$$Y_i = \sum_{j=1}^{n} 1_{[t_i,\infty)}(X_j) = n - i + 1.$$

Sind die Daten unszensiert, so gilt $\delta_i = 1$ für alle i und die Schätzer nehmen eine noch einfachere Gestalt an.

5.4 Hazard-Funktionale

Wir betrachten im Folgenden erneut das volle Modell $\mathcal{P} := \mathcal{P}^*$ (1.1), einen beliebigen Fußpunkt $P_0 \in \mathcal{P}$ und die dichte Teilmenge $D := D^{**} \subset T(P_0, \mathcal{P}^*)$ der beschränkten Tangenten mit kompaktem Träger, der vollständig im Inneren des Trägers von P_0 liegt, aus Lemma 1.11.

Sei $h : \Omega \to \mathbb{R}$ eine reellwertige messbare Funktion mit

$$\int |h| \, d\Lambda < \infty$$

für alle $P \in \mathcal{P}$ und zugehörigen Hazardmaßen Λ und Verteilungsfunktionen F.
Dann definiert

$$\kappa(P) := \int h \, d\Lambda = \int \frac{h}{1-F} \, dP \tag{5.8}$$

ein einfaches Hazard-Funktional.

Einen Schätzer für κ erhalten wir, indem wir die Verteilungsfunktion durch die empirische Verteilungsfunktion und das Integral durch das arithmetische Mittel ersetzen. Verwenden wir anstelle der Verteilungsfunktion $F(x)$ ihre linksstetige Version $F(x-)$, so ergibt sich mit $\hat{F}_n(X_{i:n}-) = \frac{i-1}{n}$ der Schätzer

$$\hat{\kappa}_n = \frac{1}{n}\sum_{i=1}^{n} \frac{h(X_i)}{1-\hat{F}_n(X_i-)} = \frac{1}{n}\sum_{i=1}^{n} \frac{h(X_{i:n})}{1-\hat{F}_n(X_{i:n}-)} = \sum_{i=1}^{n} \frac{h(X_{i:n})}{n-i+1}. \tag{5.9}$$

als empirische Version des Funktionals (5.8).

Bemerkung 5.2
κ ist durch

$$\kappa(P) = \int \frac{h}{1-F}\, dP = \int_0^1 h(F^{-1}(u))\frac{1}{1-u}\, du$$

als L-Funktional darstellbar. Allerdings ist hier weder $b(u) = \frac{1}{1-u}$ noch $b'(u) = \frac{1}{(1-u)^2}$ beschränkt, so dass Satz 4.11 nur unter zusätzlichen Voraussetzungen anwendbar ist.

Wir wollen die dichte Differenzierbarkeit daher speziell für Hazard-Funktionale in größter Allgemeinheit auf andere Weise zeigen. Hierzu konstruieren wir für Tangenten $g \in D$ zunächst neue Kurven $\vartheta \mapsto P_\vartheta$ mit Tangente g, die implizit durch ihren Hazardratenquotienten $\frac{d\Lambda_\vartheta}{d\Lambda_0}$ gegeben sind und dadurch im Rahmen der Hazard-Funktionale einfacher zu handhaben sein werden. Gemäß Abschnitt 5.1 erhalten wir aus der L_1-Differenzierbarkeit einer Kurve $\vartheta \mapsto P_\vartheta$ eine Differenzierbarkeit des Hazardratenquotienten. Wir wollen nun keinesfalls auf Seiten der Hazards die Differenzierbarkeit von Modellen $\{\Lambda_\vartheta : \vartheta \in \Theta\}$ studieren, da dies aufgrund der Unbeschränktheit der Hazardmaße einige Probleme mit sich bringt. Stattdessen wollen wir den Operator R nutzen, um zu den Tangenten überzugehen. Daher stellt sich uns nun die Frage, welche Bedingungen wir an den Hazardratenquotienten stellen müssen, um die L_2-Differenzierbarkeit der Kurve $\vartheta \mapsto P_\vartheta$ zu erhalten. Die Antwort auf diese Frage wird hier allerdings nicht in dieser Allgemeinheit benötigt, sondern es soll lediglich für spezielle gegebene Hazardratenquotienten die L_2-Differenzierbarkeit der zugehörigen Kurve gezeigt und die Tangente angegeben werden.

Satz 5.3
Sei $P_0 \in \mathcal{P}$ und Λ_0 das zugehörige Hazardmaß. Sei $g \in D$. Dann gibt es ein $\delta > 0$ so dass für $|\vartheta| \leq \delta$ durch

$$\frac{d\Lambda_\vartheta}{d\Lambda_0} = 1 + \vartheta R(g)$$

eine $L_2(P_0)$-differenzierbare Kurve $\vartheta \mapsto P_\vartheta$ in $\mathcal{P}$ mit Tangente g definiert wird. Diese ist explizit gegeben durch

$$\frac{dP_\vartheta}{dP_0}(x) = (1 + \vartheta R(g)(x)) \exp(\vartheta(g(x) - R(g)(x))).$$

Beweis. Sei $|g| < c_1$, $c_1 \in \mathbb{R}$.
Trivialerweise gilt $1 - F_0 > 0$ P_0-f.s. und somit P_0-f.s.

$$|g(x) - R(g)(x)| = \left| \int_x^\infty g dP_0 \frac{1}{1 - F_0(x)} \right|$$

$$\leq \int_x^\infty |g| dP_0 \frac{1}{1 - F_0(x)} \leq c_1 \frac{1 - F_0(x)}{1 - F_0(x)} = c_1$$

und

$$|R(g)| \leq |g| + |g - R(g)| \leq c_1 + c_1 = 2c_1.$$

Wir zeigen nun zunächst, dass durch den Hazardquotienten $\frac{d\Lambda_\vartheta}{d\Lambda_0} = 1 + \vartheta R(g)$ für betraglich ausreichend kleine ϑ überhaupt eine Kurve lokal in $\mathcal{P}$ gegeben ist, also Wahrscheinlichkeitsmaße $P_\vartheta \in \mathcal{P}$ definiert werden. Es gilt

$$\Lambda_\vartheta(t) = \int_{-\infty}^t \frac{d\Lambda_\vartheta}{d\Lambda_0}\, d\Lambda_0 = \int_{-\infty}^t 1 + \vartheta R(g)\, d\Lambda_0 = \Lambda_0(t) + \vartheta \int_{-\infty}^t R(g)\, d\Lambda_0$$

und

$$\begin{aligned} F_\vartheta(t) &= 1 - \exp(-\Lambda_\vartheta(t)) = 1 - \exp\left(-\Lambda_0(t) - \vartheta \int_{-\infty}^t R(g)\, d\Lambda_0 \right) \\ &= 1 - \exp\left(-\Lambda_0(t) + \vartheta(L(R(g))(t) - R(g)(t))\right) \\ &= 1 - \exp\left(\log(1 - F_0(t)) + \vartheta(g(t) - R(g)(t))\right) \\ &= 1 - (1 - F_0(t)) \exp\left(\vartheta(g(t) - R(g)(t))\right) \end{aligned}$$

und die Kurve ist gegeben durch

$$\frac{dP_\vartheta}{dP_0}(x) = \frac{d\Lambda_\vartheta}{d\Lambda_0}(x) \frac{1 - F_\vartheta(x)}{1 - F_0(x)} = (1 + \vartheta R(g)(x)) \exp(\vartheta(g(x) - R(g)(x))).$$

Nun gilt

$$\lim_{t \to -\infty} g(t) - R(g)(t) = \lim_{t \to -\infty} \int_t^\infty g\, dP_0 \frac{1}{1 - F_0(t)} = \int g\, dP_0 = 0$$

und somit

$$\lim_{t \to -\infty} F_\vartheta(t) = 1 - exp(0) = 0$$

und da $g - R(g)$ beschränkt ist

$$\lim_{t \to \infty} F_\vartheta(t) = 1.$$

Für $|\vartheta|$ ausreichend klein gilt außerdem $\frac{dP_\vartheta}{dP_0} \geq 0$ und folglich ist P_ϑ ein Wahrscheinlichkeitsmaß mit Verteilungsfunktion F_ϑ.
Zu zeigen bleibt die $L_2(P_0)$-Differenzierbarkeit dieser Kurve mit Tangente g.
Wir bestimmen zuerst die Tangente, das heißt wir bilden die Ableitung des Dichtequotienten und erhalten somit die stochastische Konvergenz des entsprechenden Differenzenquotienten. Es gilt

$$\begin{aligned}
&\frac{d}{d\vartheta}\frac{dP_\vartheta}{dP_0}_{|\vartheta=0} \\
=&R(g)\exp(\vartheta(g - R(g)))_{|\vartheta=0} + (1 + \vartheta R(g))\exp(\vartheta(g - R(g)))(g - R(g))_{|\vartheta=0} \\
=&R(g) + g - R(g) = g.
\end{aligned}$$

Es folgt der Nachweis der $L_2(P_0)$-Differenzierbarkeit. Setze

$$C(\vartheta) := \left(\int \exp(\vartheta(g - R(g)))\, dP_0 \right)^{-1} < \infty$$

und definiere eine Kurve $\{Q_\vartheta : |\vartheta| < \delta\}$ durch

$$\frac{dQ_\vartheta}{dP_0} := C(\vartheta)\exp(\vartheta(g - R(g))).$$

Dann ist $\vartheta \mapsto Q_\vartheta$ eine Exponentialfamilie und daher nach BKRW [7], S.14, Beispiel 1, $L_2(P_0)$-differenzierbar, d.h. es gilt

$$\frac{2}{\vartheta}\left(\sqrt{C(\vartheta)}\exp\left(\frac{\vartheta}{2}(g - R(g))\right) - 1 \right) \to g - R(g) + E_{P_0}(R(g)) \tag{5.10}$$

in $L_2(P_0)$.
Weiter gilt

$$\begin{aligned}
&\frac{2}{\vartheta}\left(\left(\frac{dP_\vartheta}{dP_0}^{1/2} \right) - 1 \right) = \frac{2}{\vartheta}\left(C(\vartheta)^{-1/2}\sqrt{1 + \vartheta R(g)}\sqrt{C(\vartheta)}\exp\left(\frac{\vartheta}{2}(g - R(g))\right) - 1 \right) \\
&= 2\sqrt{C(\vartheta)}\exp\left(\frac{\vartheta}{2}(g - R(g))\right)\left(C(\vartheta)^{-1/2}\frac{\sqrt{1 + \vartheta R(g)} - 1}{\vartheta} + \frac{C(\vartheta)^{-1/2} - 1}{\vartheta} \right)
\end{aligned}$$

$$+ 2\frac{\sqrt{C(\vartheta)}\exp\left(\frac{\vartheta}{2}(g - R(g))\right) - 1}{\vartheta}. \tag{5.11}$$

Aufgrund der Beschränktheit von $g - R(g)$ sind auch $\sqrt{C(\vartheta)}$ sowie $C(\vartheta)^{-1/2}$ und ebenso $\exp\left(\frac{\vartheta}{2}(g - R(g))\right)$ beschränkt für $|\vartheta| < \delta$.
Außerdem gilt

$$\frac{\sqrt{1 + \vartheta R(g)} - 1}{\vartheta} = \frac{1 + \vartheta R(g) - 1}{\vartheta}\frac{1}{\sqrt{1 + \vartheta R(g)} + 1} = \frac{R(g)}{\sqrt{1 + \vartheta R(g)} + 1}$$

und auch dieser Bruch ist beschränkt für $|\vartheta| < \delta$, da $R(g)$ beschränkt ist.
Zuletzt gilt

$$2\left|\frac{C(\vartheta)^{-1/2} - 1}{\vartheta}\right| \to 2\frac{d}{d\vartheta}C(\vartheta)^{-1/2}{}_{|\vartheta=0} = E_{P_0}(R(g))$$

für $\vartheta \to 0$ unabhängig von $x \in \Omega_0$.
Zusammen mit (5.10) folgt die $L_2(P_0)$-Konvergenz des gesamten Ausdrucks (5.11) und somit die $L_2(P_0)$-Differenzierbarkeit von $\vartheta \mapsto P_\vartheta$. ■

Beispiel 5.4 (Fortsetzung von Beispiel 5.1)
Sei $P_0 = \exp(1)$ die Standard-Exponentialverteilung mit der Dichte

$$f_0(x) = \exp(-x)1_{[0,\infty)}(x)$$

und $g(x) := 1 - x$.
Dann ist $g \in L_2^0(P_0)$ eine Tangente mit $R(g) = 1$.
Definiere nun eine Kurve $\vartheta \mapsto P_\vartheta$ durch $\frac{d\Lambda_\vartheta}{d\Lambda_0} = 1 + \vartheta R(g) = 1 + \vartheta$. Dann gilt

$$\begin{aligned}\frac{dP_\vartheta}{dP_0}(x) &= (1 + \vartheta)\exp(\vartheta(g(x) - R(g)(x))) \\ &= (1 + \vartheta)\exp(\vartheta(1 - x - 1)) = (1 + \vartheta)\exp(-\vartheta x)\end{aligned}$$

und

$$\begin{aligned}f_\vartheta(x) &= \frac{dP_\vartheta}{dP_0}(x)f_0(x) = (1 + \vartheta)\exp(-\vartheta x)\exp(-x)1_{[0,\infty)}(x) \\ &= (1 + \vartheta)\exp(-(1 + \vartheta)x)1_{[0,\infty)}(x).\end{aligned}$$

Es ergibt sich also die Familie der Exponentialverteilungen $P_\vartheta = \exp(1 + \vartheta)$ zum Parameter $1 + \vartheta$. Diese ist als Exponentialfamilie in der Statistik $T(x) = -x$ wie in Beispiel 5.1 L_2-differenzierbar in $\vartheta = 0$ mit der Tangente $T(x) - E_0(T) = x - 1 = g(x)$.

Mit Hilfe dieser Kurven können wir jetzt die dichte Differenzierbarkeit des Hazardfunktionals zeigen und den kanonischen Gradienten berechnen.

Satz 5.5
Das Hazardfunktional κ (5.8) ist dicht differenzierbar in P_0 bezüglich D mit dem kanonischen Gradienten

$$\tilde{\kappa} = L\left(\frac{h}{1-F_0}\right) \in L_2(P_0).$$

Beweis. Sei $g \in D$. Nach Lemma 5.3 ist durch $\frac{d\Lambda_\vartheta}{d\Lambda_0} = 1 + \vartheta R(g)$, $|\vartheta| < \delta$ eine $L_2(P_0)$-differenzierbare Kurve mit der Tangente g gegeben. Es folgt

$$\begin{aligned}
\frac{\kappa(P_\vartheta) - \kappa(P_0)}{\vartheta} &= \frac{\int h \, d\Lambda_\vartheta - \int h \, d\Lambda_0}{\vartheta} \\
= &\frac{\int h(1 + \vartheta R(g)) - h \, d\Lambda_0}{\vartheta} = \int hR(g) \, d\Lambda_0 \\
= &\int \frac{h}{1-F_0} R(g) \, dP_0 \\
\overset{(4.5)}{=} &\int L\left(\frac{h}{1-F_0}\right) g \, dP_0
\end{aligned}$$

und die Behauptung ist gezeigt. ■

Beispiel 5.6 (Nelson-Aalen-Schätzer für unzensierte Daten)
Für $h = 1_{(-\infty,t]}$, $t \in \mathbb{R}$ mit $F_0(t) < 1$, ergibt sich das Funktional

$$\kappa(P) = \int_{-\infty}^{t} d\Lambda = \Lambda(t)$$

als kumulative Hazardfunktion an der Stelle t.
Nach Satz 5.5 ist κ dicht differenzierbar und der kanonische Gradient dieses Funktionals ergibt sich zu

$$\begin{aligned}
&\tilde{\kappa}(x) = L\left(\frac{h}{1-F_0}\right)(x) = \frac{h(x)}{1-F_0(x)} - \int_{-\infty}^{x} \frac{h(y)}{1-F_0(y)} \frac{1}{1-F_0(y)} \, dP(y) \\
&= \frac{1_{(-\infty,t]}(x)}{1-F_0(x)} - \int_{-\infty}^{x} \frac{1_{(-\infty,t]}(y)}{(1-F_0(y))^2} \, dP(y) \\
&= 1_{(-\infty,t]}(x) \left(\frac{1}{1-F_0(x)} - \int_{-\infty}^{x} \frac{1}{(1-F_0(y))^2} \, dP(y) \right) - 1_{(t,\infty)}(x) \int_{-\infty}^{t} \frac{1}{(1-F_0(y))^2} \, dP(y)
\end{aligned}$$

$$=1_{(-\infty,t]}(x)\left(\frac{1}{1-F_0(x)}-\left(\frac{1}{1-F_0(x)}-1\right)\right)-1_{(t,\infty)}(x)\left(\frac{1}{1-F_0(t)}-1\right)$$
$$=1_{(-\infty,t]}(x)-1_{(t,\infty)}(x)\frac{F_0(t)}{1-F_0(t)}$$

Für diese spezielle Wahl von h ist $\kappa(P)=\Lambda(t)=-\log(1-F(t))$ sogar differenzierbar im klassischen Sinn, das heißt (3.1) gilt für für jede $L_2(P_0)$-differenzierbare Kurve $\vartheta\mapsto P_\vartheta$ mit Tangente $g\in L_2^0(P_0)$. Denn für jede solche Kurve gilt

$$\frac{d}{d\vartheta}\kappa(P_\vartheta)_{|\vartheta=0}=\frac{d}{d\vartheta}-\log(1-F_\vartheta(t))=\frac{1}{1-F_0(t)}\frac{d}{d\vartheta}\left(\int\limits_{-\infty}^{t}dP_\vartheta\right)_{|\vartheta=0}$$
$$=\frac{1}{1-F_0(t)}\int\limits_{-\infty}^{t}gdP_0=\int\limits_{-\infty}^{t}gdP_0\left(1+\frac{F_0(t)}{1-F_0(t)}\right)$$
$$=\int\limits_{-\infty}^{t}gdP_0-\frac{F_0(t)}{1-F_0(t)}\int\limits_{t}^{\infty}gdP_0=\int\left(1_{(-\infty,t]}-\frac{F_0(t)}{1-F_0(t)}1_{(t,\infty)}\right)gdP_0$$

und es ergibt sich auf diese Weise derselbe kanonische Gradient $\tilde{\kappa}$.

Eine untere Schranke für die Varianz der Grenzverteilung eines regulären Schätzers für die kumulative Hazardfunktion ist nun gemäß Kapitel eins gegeben durch

$$\int\tilde{\kappa}^2\,dP_0=\int\limits_{-\infty}^{t}dP_0+\int\limits_{t}^{\infty}\frac{F_0(t)^2}{(1-F_0(t))^2}dP_0=F_0(t)+\frac{F_0(t)^2}{1-F_0(t)}=\frac{F_0(t)}{1-F_0(t)}.$$

Seien $t_1<t_2<...<t_n$ geordnete Realisierungen der sortierten Überlebenszeiten $X_{1:n},...,X_{n:n}$ und

$$Y_i=\sum_{j=1}^{n}1_{[t_i,\infty)}(X_j)=n-i+1$$

die Anzahl der Objekte unter Risiko zum Zeitpunkt t_i.
Der Schätzer (5.9) ergibt sich für unser h dann zu

$$\hat{\Lambda}_n(t)=\sum_{i=1}^{n}\frac{1_{(-\infty,t]}(X_{i:n})}{n-i+1}=\sum_{t_i\le t}\frac{1}{Y_i}$$

und entspricht dem Nelson-Aalen-Schätzer (5.7) für unzensierte Daten. Dieser ist regulär und asymptotisch normal mit der asymptotischen Varianz

$$\int_{-\infty}^{t}\frac{\lambda_0(x)}{S_0(x)}\,dx=\int_{-\infty}^{t}\frac{1}{S_0^2}\,dP_0=\frac{1}{1-F_0(t)}-1=\frac{F_0(t)}{1-F_0(t)},$$

vgl. ABGK [2], S.626, bzw. Theorem IV.1.2, S.191.
Der Nelson-Aalen-Schätzer nimmt also die Schranke für die asymptotische Varianz an und ist somit asymptotisch effizient.

5.5 Hazard-Funktionale für rechtszensierte Daten

Auch unter Rechtzensierung sei

$$\kappa\left(P^T\right) = \int h \, d\Lambda_T$$

nach wie vor die zu schätzende Größe. Sie bildet zunächst aber kein Funktional unseres Modells $\mathcal{Q}$ sondern ein Funktional von $P^T \in \mathcal{P}$, dessen Realisierungen nicht mehr unmittelbar beobachtbar sind. Wir können daher nur Funktionale $\kappa : \mathcal{Q} \to \mathbb{R}$ behandeln und müssen das Hazardfunktional umformulieren.
Da T und C unabhängig sind, gilt

$$1 - F_X(x) = P(\min(T, C) > x) = P(T > x, C > x) = (1 - F_T(x))(1 - F_C(x)).$$

Mit (5.3) ergibt sich für $h(x, \delta) = \delta h(x)$

$$\begin{aligned} \kappa(Q) := \int \frac{\delta h(x)}{1 - F_X(x)} \, dQ(x, \delta) &= \int \frac{\delta h(x)}{(1 - F_T(x))(1 - F_C(x))} \, dQ(x, \delta) \qquad (5.12) \\ = \int \frac{h(x)}{1 - F_T(x)} \, dP^T(x) &= \int h \, d\Lambda_T. \end{aligned}$$

als das gesuchte Funktional.

Für Beobachtungen $X = (X_1, ..., X_n)$ sei $d_i(X)$ der Antirang von i, also der Index der i-ten kleinsten Beobachtung, vgl. etwa Janssen [19], S.2. Dann ist

$$\Delta_{d_i(X)} = \begin{cases} 1 & \text{die Daten } X_{i:n} \text{ sind nicht zensiert} \\ 0 & \text{sonst} \end{cases} .$$

Analog zu (5.9) ergibt sich ein Schätzer für (5.12) durch

$$\hat{\kappa}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{h(X_i)\Delta_i}{1 - \hat{F}_n(X_i-)} = \frac{1}{n} \sum_{i=1}^{n} \frac{h(X_{i:n})\Delta_{d_i(X)}}{1 - \hat{F}_n(X_{i:n}-)} = \sum_{i=1}^{n} \frac{h(X_{i:n})\Delta_{d_i(X)}}{n - i + 1}. \qquad (5.13)$$

Sei $P_0 \in \mathcal{P}, Q_0 = \mathcal{L}((X, \Delta)|P_0 \otimes P^C)$. Sei $D \subset T(P_0, \mathcal{P})$ wie in Abschnitt 5.4 die dichte Teilmenge der beschränkten Tangenten mit kompaktem Träger, der vollständig

im Inneren des Trägers von P_0 liegt.
Da für eine $L_2(P_0)$-differenzierbare Kurve $\vartheta \mapsto P_\vartheta$ lokal in $\mathcal{P}$ mit der Tangente $g \in K(P_0, \mathcal{P})$ die Kurve

$$\vartheta \mapsto Q_\vartheta = \mathcal{L}((X, \Delta)|P_\vartheta \otimes P^C)$$

lokal in $\mathcal{Q}$ ebenfalls L_2-differenzierbar mit der Tangente

$$\tilde{g}(x, \delta) := \delta g(x) + (1 - \delta)\overline{R}(g)(x)$$

ist, definiere

$$\tilde{D} := \{\tilde{g} : g \in D\} \subset T(Q_0, \mathcal{Q}).$$

Satz 5.7
Das Hazardfunktional

$$\kappa(Q) = \int \frac{\delta h(x)}{1 - F_X(x)}\, dQ(x, \delta) = \int h\, d\Lambda_0$$

aus (5.12) ist differenzierbar in Q_0 bezüglich $\tilde{D}$ mit dem kanonischen Gradienten

$$\tilde{\kappa}(x, \delta) = \tilde{L}\left(\tilde{\gamma}\right)(x, \delta) \in \overline{\tilde{D}} \text{ mit}$$

$$\tilde{\gamma}(x, \delta) := \frac{\delta h(x)}{1 - F_X(x)}.$$

Beweis. Sei $\tilde{g} \in \tilde{D}, \tilde{g} = \delta g + (1 - \delta)\overline{R}(g), g \in D$.
Dann ist die Kurve $\vartheta \mapsto P_\vartheta$, gegeben durch $\frac{d\Lambda_\vartheta}{d\Lambda_0} = 1 + \vartheta R(g)$ nach Lemma 5.3 L_2-differenzierbar mit Tangente g und $\vartheta \mapsto Q_\vartheta$ somit L_2-differenzierbar mit Tangente $\tilde{g}$, wobei $\Lambda_\vartheta = \Lambda_{\vartheta,T}$ die kumulative Hazardfunktion von $P_\vartheta = (P^T)_\vartheta$ bezeichne.
Für diese gilt dann mit (5.4)

$$\begin{aligned}
&\frac{1}{\vartheta}\left(\kappa(Q_\vartheta) - \kappa(Q_0)\right) = \frac{1}{\vartheta}\left(\int h\, d\Lambda_\vartheta - \int h\, d\Lambda_0\right) \\
&= \frac{1}{\vartheta}\int h(1 + \vartheta R(g) - 1)\, d\Lambda_0 = \int hR(g)\, d\Lambda_0 = \int \frac{h}{1 - F_0} R(g)\, dP_0 \\
&= \int \frac{\delta h(x)}{(1 - F_0(x))(1 - F_C(x))} R(g)(x)\, dQ_0(x, \delta) \\
&= \int \frac{\delta h(x)}{1 - F_X(x)} \delta R(g)(x)\, dQ_0(x, \delta) \\
&= \int \frac{\delta h(x)}{1 - F_X(x)} \tilde{R}(\tilde{g})(x, \delta)\, dQ_0(x, \delta)
\end{aligned}$$

$$= \int \tilde{L}\left(\tilde{\gamma}\right)(x,\delta)\tilde{g}(x)\, dQ_0(x,\delta).$$

■

Nach Satz 3.8 bildet $\|\tilde{\kappa}\|^2$ eine Informationsschranke von κ auf $\mathcal{Q}$ und nimmt die asymptotische Varianz eines regulären Schätzers diese an, so ist er asymptotisch effizient.

Korollar 5.8

Sei $Q_0 \in \mathcal{Q}$ beliebig und κ, $\tilde{\kappa}$ und $\tilde{D}$ wie in Satz 5.7. Sei T_n ein in Q_0 regulärer Schätzer bezüglich $\tilde{\tilde{D}} \subset \tilde{D}$ mit $\tilde{\kappa} \in \overline{\tilde{\tilde{D}}}$ und

$$\mathcal{L}(\sqrt{n}(T_n - \kappa(P_0)|P_0)) \xrightarrow[n\to\infty]{} \mathcal{N}(0,\sigma^2).$$

Dann gilt

$$\sigma^2 \geq \int \frac{h(x)}{(1-F_0(x))^2(1-F_C(x))}\, dP_0(x).$$

Bei Gleichheit ist T_n asymptotisch effizient für κ.

Beweis. Mit (5.3) und der Isometrieeigenschaft von $\tilde{L}$ gilt

$$\int \tilde{\kappa}^2\, dQ_0 = \int \left(\tilde{L}\left(\frac{\delta h(x)}{1-F_X(x)}\right)(x,\delta)\right)^2 dQ_0(x,\delta)$$

$$= \int \left(\frac{\delta h(x)}{1-F_X(x)}\right)^2 dQ_0(x,\delta)$$

$$= \int \frac{h(x)(1-F_C(x))}{(1-F_X(x))^2}\, dP_0(x) = \int \frac{h(x)}{(1-F_0(x))^2(1-F_C(x))}\, dP_0(x).$$

Die Behauptung folgt damit sofort aus Satz 3.8. ■

Beispiel 5.9 (Nelson-Aalen-Schätzer für rechtszensierte Daten)

Wie in Beispiel 5.6 ergibt sich für $h = 1_{(-\infty,t]}, t \in \mathbb{R}$ mit $F_0(t) < 1$ das Hazardfunktional

$$\kappa(P^{(X,\Delta)}) = \int_{-\infty}^{t} d\Lambda_T = \Lambda_T(t),$$

also die kumulative Hazardfunktion. Nach Satz 5.7 ist dieses Funktional für

$$\tilde{\gamma}(x,\delta) = \frac{\delta 1_{(-\infty,t](x)}}{1-F_X(x)}$$

zumindest bezüglich $\tilde{D}$ differenzierbar mit dem kanonischen Gradienten

$$\tilde{\kappa}(x,\delta) = \tilde{L}\left(\tilde{\gamma}\right)(x,\delta) = \delta\frac{1_{(-\infty,t](x)}}{1-F_X(x)} - \int_{-\infty}^{x} \frac{1_{(-\infty,t](y)}}{(1-F_X(y))(1-F_0(y))}\, dP_0(y).$$

In Beispiel 5.6 haben wir den kanonischen Gradienten weiter ausgerechnet und die Norm direkt bestimmt. Hier führt dieser Weg zu einer langen Rechnung, daher nutzen wir diesmal die Isometrie-Eigenschaft von $\tilde{L}$ aus bzw. wenden Korollar 5.8 an. Demnach gilt

$$\int \tilde{\kappa}^2 \, dQ_0 = \int\limits_{-\infty}^{t} \frac{1}{(1-F_0(x))^2(1-F_C(x))} \, dP_0(x).$$

Seien $t_1 < t_2 < ... < t_n$ geordnete Realisierungen der sortierten zensierten Überlebenszeiten $X_{1:n}, ..., X_{n:n}$ und $\delta_i \in \{0,1\}$ die zugehörigen Realisierungen von $\Delta_{d_i(X)}$. $\delta_i = 0$ bedeutet also, dass das i-te früheste Ereignis aus einer Zensierung resultiert. Sei zudem

$$Y_i = \sum_{j=1}^{n} 1_{[t_i,\infty)}(X_j) = n - i + 1$$

die Anzahl der zum Zeitpunkt t_i unter Risiko stehenden Objekte.
Dann lässt sich der Schätzer (5.13) schreiben als

$$\hat{\kappa}_n(t) = \sum_{i=1}^{n} \frac{h(X_{i:n})\Delta_{d_i(X)}}{n-i+1} = \sum_{t_i \le t} \frac{\delta_i}{Y_i}$$

und entspricht dem Nelson-Aalen-Schätzer $\hat{\Lambda}_{n,NA}(t)$ (5.7) für rechtszensierte Daten. Aus ABGK [2], Theorem IV.1.2 bzw. Beispiel IV.1.6, folgt die asymptotische Normalität des Nelson-Aalen-Schätzers

$$\mathcal{L}(\sqrt{n}(\hat{\Lambda}_{NA}(t) - \Lambda(t))|Q_0) \xrightarrow[n\to\infty]{} \mathcal{N}(0, \sigma^2(t))$$

(schwache Konvergenz der Maße) mit

$$\sigma^2 = \int\limits_{-\infty}^{t} \frac{\lambda(u)}{(1-F_0(u))(1-F_C(u))} \, du = \int\limits_{-\infty}^{t} \frac{1}{(1-F_0(x))^2(1-F_C(x))} \, dP_0(x) = \|\tilde{\kappa}\|^2.$$

$\hat{\Lambda}_{NA}$ nimmt die Schranke also asymptotisch an und ist nach ABGK [2], S.626, zudem regulär. Somit ist der Nelson-Aalen-Schätzer für rechtszensierte Daten asymptotisch effizient.

Bemerkung 5.10
Die asymptotische Effizienz des Nelson-Aalen-Schätzers ist wohlbekannt, vgl. ABGK [2], S.624 ff.

Wie in Abschnitt 5.3 schon erwähnt stand vor der Entwicklung des Nelson-Aalen-Schätzers historisch der Kaplan-Meier-Schätzer als Maximum-Likelihood-Schätzer für die Survivalfunktion. Eine asymptotische Optimalität dieses Schätzers wurde zum ersten Mal von Wellner betrachtet und in diesem Zuge die asymptotische Effizienz desselben bewiesen, vgl. Wellner [39], S.597, Theorem 1. Wir können dieses Ergebnis nun als weiteres Anwendungsbeispiel auch mit unseren Methoden verifizieren.

Beispiel 5.11 (Kaplan-Meier-Schätzer)
Auch wenn der Kaplan-Meier-Schätzer wie in Abschnitt 5.3 erläutert nicht auf dem Nelson-Aalen-Schätzer basiert oder umgekehrt, lässt sich der zu schätzende Wert $S(t)$ aber dennoch wegen

$$S(t) = \exp(-\Lambda_0(t))$$

als Funktional

$$\kappa_s(Q) = \exp(-\kappa(Q)) = S(t) \tag{5.14}$$

schreiben mit κ aus (5.12) für $h = 1_{(-\infty,t]}$.
Nach Satz 5.7 ist κ differenzierbar in Q_0 bezüglich $\tilde{D}$ mit dem kanonischen Gradienten

$$\tilde{\kappa}(x,\delta) = \tilde{L}\left(\frac{\delta h(x)}{1 - F_X(x)}\right)(x,\delta) \in \overline{\tilde{D}}.$$

Weiterhin gilt

$$\frac{d}{dx}\exp(-x)_{|\kappa(Q_0)} = -\exp(-\Lambda(t)) = -S(t)$$

für $h = 1_{(-\infty,t]}$.
Nach der Kettenregel, Korollar 3.6, ist $\kappa_S(Q)$ (5.14) nun differenzierbar in Q_0 bezüglich $\tilde{D}$ mit dem kanonischen Gradienten

$$\tilde{\kappa}_s(x,\delta) = -S(t)\tilde{L}\left(\frac{\delta 1_{(-\infty,t]}(x)}{1 - F_X(x)}\right)(x,\delta) \in \overline{\tilde{D}}.$$

Eine untere Schranke für die asymptotische Varianz eines regulären Schätzers ist somit gegeben durch

$$\int \tilde{\kappa}_s^2 \, dQ_0 = S(t)^2 \int\limits_{-\infty}^{t} \frac{1}{(1 - F_0(x))^2(1 - F_C(x))} \, dP_0(x) =: \sigma^2.$$

Der Kaplan-Meier-Schätzer $\hat{S}_{n,KM}(t)$ (5.5) ist gemäß Breslow, Crowley [9], Theorem 5, vgl. auch Wellner [39], S.597, regulär und

$$\sqrt{n}(\hat{S}_{n,KM}(t) - S(t))$$

konvergiert in Verteilung gegen eine normalverteilte Zufallsvariable mit Erwartungswert 0 und Varianz σ^2. Somit ist der Kaplan-Meier-Schätzer ein asymptotisch effizienter Schätzer für die Survivalfunktion.

Anhang

A.1 Verteilungsfunktionen

Definition A.1 (inverse Verteilungsfunktion)
Sei $F(y) := P(X \leq y)$ eine Verteilungsfunktion. Dann heißt

$$F^{-1} : (0,1) \to \mathbb{R}, F^{-1}(u) := \inf\{y \in \mathbb{R} : F(y) \geq u\} \text{ für } u \in (0,1)$$

die inverse Verteilungsfunktion von F.

Definition A.2 (empirische Verteilungsfunktion)
Seien $X_1, ..., X_n$ i.i.d. Zufallsvariablen und $x_1, ..., x_n$ Realisierungen. Dann heißt

$$x = (x_1, ..., x_n) \mapsto \hat{F}_n(y) := \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty,y]}(x_i)$$

die empirische Verteilungsfunktion zur Realisierung $x_1, ..., x_n$.

Die empirische Verteilungsfunktion ordnet also jeder Realisierung $x = (x_1, ..., x_n)$ die diskrete Verteilung zu, die in jedem Punkt $x_1, ..., x_n$ die Wahrscheinlichkeit $\frac{1}{n}$ besitzt. Hieraus ergibt sich die folgende Definition der empirischen Quantilfunktion.

Definition A.3 (empirisches Quantil)
Seien $X_1, ..., X_n$ i.i.d. Zufallsvariablen und $x_1, ..., x_n$ Realisierungen. Dann heißt

$$x = (x_1, ..., x_n) \mapsto \hat{F}_n^{-1}(u) := \begin{cases} x_{1:n} & \text{für } u = 0 \\ \sum\limits_{i=1}^{n} x_{i:n} 1_{\left(\frac{i-1}{n}, \frac{i}{n}\right]}(u) & \text{für } u > 0 \end{cases}$$

die empirische Quantilfunktion zur Realisierung $x_1, ..., x_n$.

Lemma A.4 (Ableitung der Inversen)
Sei F^{-1} die Umkehrfunktion einer streng monotonen Funktion $F : I \to \mathbb{R}$. Ist F

in $y_0 \in I$ *differenzierbar mit* $f(y_0) := F'(y_0) \neq 0$*, so ist* F^{-1} *in* $x_0 = f(y_0)$ *differenzierbar mit*

$$(F^{-1})'(x_0) = \frac{1}{f(y_0)} = \frac{1}{f(F^{-1}(x_0))}.$$

Beweis. Siehe Königsberger [25], S.143. ■

Satz A.5

Sei F *eine Verteilungsfunktion mit zugehöriger Verteilung* P*.*

a) Sei F^{-1} *die inverse Verteilungsfunktion zu* F*. Dann gilt*

$$P = (\lambda\!\!\!\lambda_{|(0,1)})^{F^{-1}}.$$

b) Ist F *stetig, so folgt*

$$P^F = (\lambda\!\!\!\lambda_{|(0,1)}).$$

Beweis. Siehe Witting [41], S.215, Hilfssatz 2.29. ■

Lemma A.6

Seien F_n, F *Verteilungsfunktionen mit Inversen* F_n^{-1}, F^{-1}*. Dann sind äquivalent*

$$F_n(x) \xrightarrow[n\to\infty]{} F(x)\ \forall\, x \in S(F)$$

und

$$F_n^{-1}(u) \to F^{-1}(u)\ \forall\, u \in S(F^{-1}) \cap (0,1).$$

Beweis. Siehe Witting, Müller-Funk [42], S.71, Satz 5.76. ■

Lemma A.7

Seien X_n, X *reelle Zufallsvariablen mit Verteilungsfunktionen* F_n *bzw.* F *und* $X_n \xrightarrow[n\to\infty]{} X$ *in Verteilung.*

a) Seien $c_n \in \mathbb{R}$ *mit* $c_n \to c \in \mathbb{R}$ *und* F *stetig in* c*. Dann gilt*

$$F_n(c_n) \xrightarrow[n\to\infty]{} F(c).$$

b) Seien $c_n \in \mathbb{R}$ mit $c_n \to \infty$. Dann gilt

$$F_n(c_n) \xrightarrow[n\to\infty]{} 1 \tag{A.15}$$

und

$$F_n(-c_n) \xrightarrow[n\to\infty]{} 0. \tag{A.16}$$

Beweis. a) Nach dem Lemma von Slutsky gilt

$$X_n - c_n \xrightarrow[n\to\infty]{\mathcal{D}} X - c.$$

Da F stetig in c ist, ist $t \mapsto P(X-c \leq t)$ stetig in 0 und mit dem Satz von Helly-Bray folgt

$$F_n(c_n) = P(X_n \leq c_n) = P(X_n - c_n \leq 0) \xrightarrow[n\to\infty]{} P(X - c \leq 0) = P(X \leq c) = F(c).$$

b) Nach Witting, Müller-Funk [42], S.56, Satz 5.51, folgt aus $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$, dass F_n straff ist, das heißt $\forall\, \varepsilon > 0\ \exists\, k > 0$ mit

$$\sup_n (1 - F_n(x)) \leq \varepsilon \quad \forall x \geq k \tag{A.17}$$

und

$$\sup_n F_n(x) \leq \varepsilon \quad \forall x \leq -k. \tag{A.18}$$

Sei nun $\varepsilon > 0$ beliebig und $k > 0$ so, dass (A.17) gilt. Da $c_n \to \infty$ gibt es ein $n_0 \in \mathbb{N}$ mit $c_n > k$ für alle $n > n_0$. Weil F_n für alle n monoton wachsend ist, folgt $1 - F_n(c_n) \leq 1 - F_n(k)$ für alle $n > n_0$ und somit auch

$$\sup_{n>n_0} (1 - F_n(c_n)) \leq \sup_{n>n_0} (1 - F_n(k)) \leq \varepsilon.$$

Es gibt also für alle $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ mit $1 - F_n(c_n) \leq \varepsilon$ für alle $n > n_0$, also gilt (A.15). Analog zeige (A.16) mit (A.18). ■

A.2 Hilbertraumtheorie

Definition A.8 (Hilbertraum)
Ein vollständiger, normierter Vektorraum $(H, \|\cdot\|)$ über einem Körper $\mathbb{K}$ mit Skalarprodukt $\langle\cdot,\cdot\rangle$ und Norm $\|x\| := \sqrt{\langle x, x\rangle}$ für alle $x \in H$ heißt Hilbertraum.

Lemma A.9
Jeder abgeschlossene Unterraum eines Hilbertraums ist selbst ein Hilbertraum.

Definition A.10 (Dualraum)
Sei H ein Hilbertraum. Dann heißt der Raum $H' := L(H, \mathbb{K})$ der linearen Abbildungen von H auf den zugehörigen Grundkörper $\mathbb{K}$ der Dualraum von H.

Satz A.11 (Darstellungssatz von Frechét-Riesz)
Sei H ein Hilbertraum und H' der zugehörige Dualraum. Zu $\Psi \in H'$ existiert genau ein $y \in H$ mit $\Psi(x) = \langle x, y \rangle$ für alle $x \in H$ und es gilt $\|\Psi\| = \|y\|$.

Beweis. Siehe Werner [40], S.226, Theorem V.3.6. ■

A.3 Analytische Hilfsmittel

Lemma A.12 (Differentiations-Lemma)
Sei I ein nicht ausgeartetes (das heißt weder leeres noch einpunktiges) Intervall in $\mathbb{R}$ und $f : I \times \Omega \to \mathbb{R}$ eine Funktion mit folgenden Eigenschaften:

a) $\omega \mapsto f(x, \omega)$ ist μ-integrierbar für alle $x \in I$;

b) $x \mapsto f(x, \omega)$ ist auf I differenzierbar für alle $\omega \in \Omega$; die Ableitung werde mit $f'(x, \omega)$ bezeichnet;

c) es existiert eine μ-integrierbare Funktion $h \geq 0$ auf Ω mit

$$|f'(x, \omega)| \leq h(\omega)$$

für alle $\omega \in \Omega$ und alle $x \in I$.

Dann ist die auf I definierte Funktion

$$\varphi(x) := \int f(x, \omega)\, d\mu(\omega)$$

differenzierbar und für alle $x \in I$ ist $\omega \mapsto f'(x, \omega)$ μ-integrierbar und es gilt

$$\varphi'(x) = \int f'(x, \omega)\, d\mu(\omega)$$

für alle $x \in I$.

Beweis. Siehe Bauer [4], S.102, Lemma 16.2. ■

Lemma A.13
Seien $y, z \in \mathbb{R}$ mit $y \geq \frac{1}{4}$ und $y + z \geq \frac{1}{4}$. Dann gilt

$$\left|\sqrt{y+z} - \sqrt{y}\right| \leq |z|.$$

Beweis. Bemerke zuerst, dass $f(x) := \sqrt{x}, x \geq 0$, konkav ist, also die Steigung $\frac{d}{dx}\sqrt{x}$ monoton fällt, und dass gilt

$$\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}} \leq 1 \Leftrightarrow 2\sqrt{x} \geq 1 \Leftrightarrow x \geq \frac{1}{4}.$$

1. Fall: $z \geq 0$
Wegen $y \geq \frac{1}{4}$ wächst $\sqrt{x}$ für $y \leq x \leq y + z$ langsamer als linear und es folgt

$$\sqrt{y+z} \leq \sqrt{y} + z$$

und somit

$$\left|\sqrt{y+z} - \sqrt{y}\right| = \sqrt{y+z} - \sqrt{y} \leq \sqrt{y} + z - \sqrt{y} = z.$$

2. Fall: $z < 0$
Wegen $y + z \geq \frac{1}{4}$ wächst $\sqrt{x}$ für $y + z \leq x \leq y$ langsamer als linear und es folgt

$$\sqrt{y+z} \geq \sqrt{y} + z$$

und somit

$$\left|\sqrt{y+z} - \sqrt{y}\right| = \sqrt{y} - \sqrt{y+z} \leq \sqrt{y} - (\sqrt{y} + z) = -z = |z|.$$

■

A.4 Grenzwertsätze

Satz A.14 (Shorack/Wellner)
Sei $X_{n1}, ..., X_{nn}$, $n \in \mathbb{N}$ ein Dreiecksschema zeilenweise unabhängiger Zufallsvariablen mit stetigen Verteilungsfunktionen F_{ni}. Sei $\overline{F}_n := \frac{1}{n}\sum_{i=1}^{n} F_{in}$, $\overline{F}_n^{-1}$ die zugehörige inverse Verteilungsfunktion, $G_{ni} := F_{ni} \circ \overline{F}_n^{-1}$ und $K_n(s,t) := s \wedge t - \frac{1}{n}\sum_{i=1}^{n} G_{ni}(s) G_{ni}(t)$. Seien $c_{ni} \in \mathbb{R}$ feste scores und $J_n(t) := \sum_{i=1}^{n} c_{ni} 1_{(\frac{i-1}{n}, \frac{i}{n}]}(t)$ mit $J_n(t) \xrightarrow[n\to\infty]{} J(t)$ für

eine Funktion $J : (0,1) \to \mathbb{R}$*. Sei* h *eine feste Bewertungsfunktion und* $g_n := h \circ \overline{F}_n^{-1}$*. Die zu betrachtende L-Statistik sei nun definiert durch*

$$T_n := \frac{1}{n} \sum_{i=1}^{n} c_{ni} h(X_{i:n}),$$

eine natürliche Zentrierungskonstante durch

$$\mu_n(J_n) := \int_0^1 g_n(t) J_n(t)\, dt,$$

$$\mu_n := \mu_n(J) = \int_0^1 g_n(t) J(t)\, dt$$

und

$$\sigma_n^2 := \int_0^1 \int_0^1 K(s,t) J(s) J(t)\, dg_n(s)\, dg_n(t).$$

Weiterhin seien folgende Voraussetzungen erfüllt:

1. *(a)* $|J| \le B$ *und* $|J_n| \le B$ *auf* $(0,1)$
   mit $B(t) := Mt^{-b_1}(1-t)^{-b_2}$ *für feste* $M, b_1, b_2 \in \mathbb{R}, b_1 \vee b_2 < 1$.

   (b) $h = h_1 - h_2$ *mit* h_1, h_2 *monoton steigend, linksstetig und* $|h_i \circ \overline{F}_n^{-1}| \le D$ *mit* $D(t) := Mt^{-d_1}(1-t)^{-d_2}$ *für feste* $0 < d_1, d_2 < 1$.

2. $K_n(s,t) \xrightarrow[n\to\infty]{} K(s,t)$ *für alle* $0 \le s, t \le 1$.

3. $\liminf_{n\to\infty} \sigma_n^2 > 0$.

Dann gilt

$$\frac{\sqrt{n}(T_n - \mu_n)}{\sigma_n} \xrightarrow[n\to\infty]{\mathcal{D}} Z \sim N(0,1).$$

Beweis. Siehe Shorack, Wellner [34], S.823, Korollar 25.5.1. ■

A.5 Landau-Symbole

Definition A.15 (o und O)

a) *Folgen:*

Seien a_n, b_n, $n \in \mathbb{N}$ zwei Folgen in $\mathbb{R}$. Dann ist $a_n \in o(b_n)$, falls gilt

$$\lim_{n\to\infty} \frac{a_n}{b_n} = 0,$$

das heißt in Worten „a_n konvergiert schneller gegen 0 als b_n".

$a_n \in O(b_n)$ bedeutet, dass gilt

$$\lim_{n\to\infty} \frac{a_n}{b_n} = t$$

für ein $t \in (0, \infty)$. Die Kurzschreibweise $c_n = \tilde{c}_n + o(b_n)$ bedeutet, dass c_n darstellbar ist als $c_n = \tilde{c}_n + a_n$ mit $a_n \in o(b_n)$.

b) *Funktionen:*

Seien $f, g : \mathbb{R} \to \mathbb{R}$ zwei reellwertige Funktionen auf $\mathbb{R}$. Dann ist $f \in o(g)$ für $x \to a \in \mathbb{R}$, falls gilt

$$\lim_{x\to a} \frac{f(x)}{g(x)} = 0,$$

das heißt in Worten „f konvergiert schneller gegen 0 als g". Die Kurzschreibweise $h(x) = \tilde{h}(x) + o(g)$ für $x \to a$ bedeutet dementsprechend, dass h darstellbar ist als $h(x) = \tilde{h}(x) + f(x)$ mit $f \in o(g)$ für $x \to a$. Der Wert $a \in \mathbb{R}$ ist oft aus dem Zusammenhang klar und wird nicht explizit angegeben.

Symbolverzeichnis

$\#$	Zählmaß
$\langle \cdot, \cdot \rangle$	Skalarprodukt
$\sim$	verteilt nach
$x \vee y$, $x \wedge y$	Maximum von x und y, Minimum von x und y
$\lceil x \rceil$	obere Gaußklammer von x
$\xrightarrow{\mathcal{D}}$	Konvergenz in Verteilung
$\xrightarrow{P}$	stochastische Konvergenz
1_A	Indikatorfunktion der Menge A
$\mathcal{B}$	Borelsche σ-Algebra
C^1	Menge der differenzierbaren Funktionen
id	Identität $x \mapsto x$
$\lambda\!\!\lambda$	Lebesgue-Maß
$\mathcal{N}(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert μ und Varianz σ^2
$S(F)$	Menge der Stetigkeitsstellen von F
$K(P_0, \mathcal{P})$	Tangentenkegel (siehe S.12)
$T(P_0, \mathcal{P})$	Tangentialraum (siehe S.12)
$\mathcal{P}^*$	volles Modell (siehe S.12)
D^*	Teilmenge der beschränkten Tangenten (siehe S.13)
D^{**}	Teilmenge der beschränkten Tangenten mit kompaktem Träger (siehe S.15)

Abkürzungsverzeichnis

bzgl.	bezüglich
bzw.	beziehungsweise
CSU	Cauchy-Schwarz-Ungleichung
f.s.	fast sicher
f.ü.	fast überall
i.i.d.	unabhängig identisch verteilt
vgl.	vergleiche

Literaturverzeichnis

[1] Aalen, O.O. (1978). *Nonparametric Inference for a Family of Counting Processes.* Annals of Statistics 6, Number 4, 701-726.

[2] Andersen, P.K., Borgan, O., Gill, R.D., Keiding, N. (1993). *Statistical models based on counting processes.* Springer, New York.

[3] Klein,J.P., Moeschberger, M.L. (1997). *Survival Analysis. Techniques for censored and truncated data.* Springer, New York.

[4] Bauer, H. (1992). *Maß- und Integrationstheorie.* John Wiley and Sons, New York.

[5] Beutner, E., Zähle, H. (2010). *A modified functional delta method and its application to the estimation of risk functionals.* Journal of Multivariate Analysis, 101, 2452-2463.

[6] Billingsley, P. (1968). *Convergence of probability measures.* de Gruyter, Berlin.

[7] Bickel, P.J., Klaassen, C., Ritov, Y. und Wellner, J.A. (1993). *Efficient and adaptive estimation for semiparametric models.* The Johns Hopkins Univ. Pr., Baltimore.

[8] Boos, D.D. (1979). *A differential for L-statistics.* Annals of Statistics 7, Number 5, 955-959.

[9] Breslow, N., Crowley, J. (1974). *A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship.* Annals of Statistics 2, Number 3, 437-453.

[10] Efron, B., Johnstone, I. (1990). *Fisher's information in terms of the hazard rate.* Annals of Statistics 18, Number 1, 38-62.

[11] Gill, R.D. (1980). *Censoring and Stochastic Integrals.* Math. Centre Tracts 124, Mathematisch Centrum, Amsterdam.

[12] Gill, R.D., Johansen, S. (1990). *A Survey of Product-Integration with a View Toward Application in Survival Analysis.* Annals of Statistics 18, Number 4, 1501-1555.

[13] Hájek, J. (1970). *A characterization of limiting distributions of regular estimates.* Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 14, 323-330.

[14] Hájek, J., Sidák, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[15] Janssen, A. (2003). *A nonparametric Cramér-Rao inequality for estimators of statistical functionals.* Statistics & Probability Letters 64, 347-358. Elsevier.

[16] Janssen, A. (2003). *Asymptotic relative efficiency of tests at the boundary of regular statistical models.* Journal of Statistical Planning and Inference 126, 461-477. Elsevier.

[17] Janssen, A. (1989). *Local asymptotic normality for randomly censored data with applications to rank tests.* Statist. Neerlandica 43, 109-125.

[18] Janssen, A. (1994) *On local odds and hazard rate models in survival analysis.* Statistics & Probability Letters 20, 355-365. Elsevier.

[19] Janssen, A. (1998). *Zur Asymptotik nichtparametrischer Tests.* Lecture Notes. Skripten zur Stochastik Nr. 29, Gesellschaft zur Förderung der Mathematischen Statistik, Münster.

[20] Janssen, A., Werft, W. (2004). *A survey about the efficiency of two-sample survival tests for randomly censored data.* Mitteilungen aus dem Mathematischen Seminar Giessen 254, 1-47. Selbstverlag des Mathematischen Instituts, Giessen.

[21] Janssen, A., Ostrovski, V. (2005). *The convolution theorem of Hájek and Le Cam - revisited.* Preprint.

[22] Johansen, S. (1978). *The Product Limit Estimator as Maximum Likelihood Estimator.* Scandinavian Journal of Statistics 5, 195-199.

[23] Kaplan, E.L., Meier, P. (1958). *Nonparametric Estimation from Incomplete Observations.* Journal of the American Statistical Association 53, 457-481.

[24] Klenke, A. (2006). *Wahrscheinlichkeitstheorie.* Springer, Berlin.

[25] Königsberger, K. (2004). *Analysis 1.* Springer, Berlin.

[26] Le Cam, L. (1960). *Locally asymptotically normal families of distributions.* University of California Publications in Statistics 3, 37-98.

[27] Le Cam, L. (1986). *Asymptotic methods in statistical decision theory.* Springer Verlag, New York.

[28] Nelson, W. (1972). *Theory and Applications of Hazard Plotting for Censored Failure Data.* Technometrics 14, 945-965.

[29] Ostrovski, V. (2005). *Testen statistischer Funktionale für Zweistichprobenprobleme.* Dissertation, Heinrich-Heine-Universität Düsseldorf.

[30] Pfanzagl, J., Wefelmeyer, W. (1982). *Contributions to a General Asymptotic Statistical Theory.* Lecture Notes in Statistics 13, Springer, Berlin.

[31] Pfanzagl, J., Wefelmeyer, W. (1985). *Asymptotic Expansions for General Statistical Models.* Springer, Berlin.

[32] Pfanzagl, J. (2001). *A nonparametric asymptotic version of the Cramér-Rao bound.* de Gunst, M., Klaassen, C. and van der Vaart, A. eds. State of the art in probability and statistics: Festschrift for Willem R. van Zwet, IMS Lecture Notes Monograph Series 36, 499-517.

[33] Ritov, Y., Wellner, J.A. (1988). *Censoring, martingales and the Cox model.* N.U. Pabhu, ed., Statistical Inference from Stochastic Processes, Vol.80 (Contemporary Math. Amer. Math. Soc, Providence, R.I), 191-219.

[34] Shorack, G.R., Wellner, J.A. (1986). *Empirical processes with applications to statistics.* John Wiley & Sons, New York.

[35] Strasser, H. (1985). *Mathematical theory of statistics.* de Gruyter, Berlin.

[36] van der Vaart, A.W. (1988). *Statistical estimation in large parameter spaces.* CWI tract 44, Centrum voor Wiskunde en Informatica, Amsterdam.

[37] van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

[38] von Müller, F. (2008). *Zur asymptotischen Effizienz von L-Schätzern im nichtparametrischen Fall.* Diplomarbeit, Heinrich-Heine-Universität Düsseldorf.

[39] Wellner, J.A. (1982). *Asymptotic Optimality of the Product Limit Estimator.* Annals of Statistics 10, Number 2, 595-602.

[40] Werner, D. (2007). *Funktionalanalysis.* Springer, Berlin.

[41] Witting, H. (1985). *Mathematische Statistik I. Parametrische Verfahren bei festem Stichprobenumfang.* B.G. Teubner, Stuttgart.

[42] Witting, H., Müller-Funk, U. (1995). *Mathematische Statistik II.* B.G. Teubner, Stuttgart.

Erklärung

Die hier vorgelegte Dissertation habe ich eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen

Andreas Knoch Düsseldorf, den 23.10.2013

Danksagung

An dieser Stelle möchte ich die Gelegenheit nutzen, Professor A. Janssen für die umfangreiche Betreuung, Hilfestellungen und ergiebigen Diskussionen zu danken. Auch allen Mitarbeitern des Lehrstuhls für Mathematische Statistik und Wahrscheinlichkeitstheorie der HHU Düsseldorf, insbesondere PD Doktor M. Pauly, gilt mein Dank für ihre Unterstützung.