Statistik Stochastscher Prozesse

Prof. Dr. Uwe Küchler
Institut für Mathematik
Humboldt-Universität zu Berlin

Sommersemester 2005

21. Juli 2005

e-mail: kuechler@mathematik.hu-berlin.de

Inhaltsverzeichnis

Kapitel 1

Einleitung

Zum Begriff des Wortes „Statistik“

Umgangssprachlich versteht man unter einer Statistik eine Zusammenstellung von Zahlen über eine Bevölkerungsgruppe, ökonomische Tätigkeiten, Naturvorgänge, Krankheiten, Umwelteinflüsse und vieles andere mehr, vergleiche z. B. Statistisches Jahrbuch oder www.statistik-berlin.de. Beispiele sind Umsatzentwicklung eines Konzerns, Sterbetafeln, Wetterabläufe, Ausbreitung von AIDS, Wasserstandshöhe der Elbe usw.

Viele Statistiken beschreiben in gewisser Weise den Zustand des Staates. Für eine solche Beschreibung wurden sie im Mittelalter, seit etwa Mitte des 18 Jahrhunderts wohl auch zuerst benutzt. Daher kommt auch ihr Name: das Wort „Statistique“ entstammt dem Französischen und bedeutet Staatswissenschaft, dabei handelt es sich um ein Kunstwort, abgeleitet aus dem lateinischen „Status“, Zustand.

Statistiken zu erstellen kostet Arbeitszeit, ihre Aufbewahrung, Auswertung und Aktualisierung ebenfalls. Heute ist dank der Mikroelektronik die Erstellung, Speicherung und Auswertung extrem erleichtert und somit werden massenhaft Statistiken zu allen möglichen Prozessen erstellt und verarbeitet (Finanzdaten, Scannerdaten an Kassen usw).

Immer wieder stellen sich dabei die Fragen:

a) Wie soll man Daten gewinnen?

b) Wie soll man Daten beschreiben, d. h. darstellen?

c) Welche Schlüsse kann man aus Daten ziehen?

Einen Beitrag zur Frage c) leistet die Mathematik mit ihrem Teilgebiet „Mathematische Statistik“. Auch zu a) lassen sich mathematische Methoden einsetzen (statistische Versuchsplanung). b) ist das Gebiet der sogenannten „empirischen Statistik“, die durch die Möglichkeit der Darstellung auf Computern einen enormen Aufschwung erhalten hat.

Grundprinzip der mathematischen Statistik:

Die Daten x, $\quad x = (x_1, x_2, \ldots, x_n)$, $x = (x_t,\ t \in T)$ mit $T = \{1, 2, \ldots, N\}$ oder $T = [0, T_0]$ werden als Realisierung x einer Zufallsgröße X, eines zufälligen Vektors oder eines stochastischen Prozesses aufgefasst, die im Rahmen eines zufälligen Experimentes (stochastisches Modell) $(\Omega, \mathscr{A}, \mathbb{P})$ entstanden ist.

Meist ist $\mathbb{P}$ nicht bekannt, man weiß aber, zum Beispiel aus prinzipiellen Überlegungen oder aus dem Charakter des zufälligen Experiments, daß $\mathbb{P}$ zu einer Familie $\mathscr{P}$ von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathscr{A})$ gehört.
Die Frage c) kann man dann so formulieren: Aus den Daten x schließe man auf $\mathbb{P}$ bzw. auf Funktionale oder Eigenschaften von $\mathbb{P}$.

In dieser Vorlesung werden wir einen Einblick in mathematische Methoden der Statistik stochastischer Prozesse vermitteln. Wir gehen vom Fall der klassischen Statistik aus, bei dem im allgemeinen unabhängige und identisch verteilte zufallsbehaftete Beobachtungen vorliegen und zeigen anhand einiger Klassen stochastischer Prozesse, welche statistischen Methoden auf Grund ihrer speziellen Struktur möglich sind, bzw. welche Eigenschaften sie besitzen.

Kapitel 2

Statistische Experimente, statistische Modelle

2.1 Definitionen

In diesem Kapitel führen wir einige Begriffe ein, und zwar in einer solchen Allgemeinheit, daß sie auch für stochastische Prozesse einsetzbar sind.

Definition 2.1. Es seinen $(\Omega, \mathscr{A})$ ein meßbarer Raum und $\mathscr{P}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathscr{A})$. Dann nennen wir $(\Omega, \mathscr{A}, \mathscr{P})$ ein *statistisches Modell.*
Weiterhin sei X eine Zufallsgröße auf $(\Omega, \mathscr{A})$ mit Werten in einem meßbaren Raum $(E, \mathscr{E})$. Dann heißt $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein *statistisches Experiment* und X wird als eine *mathematische Stichprobe* bezeichnet.

Interpretation: Ein zufälliges Experiment wird gemäß $(\Omega, \mathscr{A}, \mathbb{P})$ mit einem bestimmten $\mathbb{P} \in \mathscr{P}$, das aber unbekannt ist, ausgeführt („wahres $\mathbb{P}$"). Dabei wird die Zufallsgröße X beobachtet, ihre Realisierungen x gehören zu E. x heißt *konkrete Stichprobe*, $(E, \mathscr{E})$ nennt man *Stichprobenraum.*

Durch $\mathbb{P}^X(B) = \mathbb{P}(X \in B),\ B \in \mathscr{E}$, ist ein Wahrscheinlichkeitsmaß $\mathbb{P}^X$ auf $\mathscr{E}$ definiert, die Wahrscheinlichkeitsverteilung von X unter $\mathbb{P}$ (*Stichprobenverteilung*).
Wir setzen $\mathscr{P}^X := \{\mathbb{P}^X : \mathbb{P} \in \mathscr{P}\}$.
$(E, \mathscr{E}, \mathscr{P}^{\mathscr{X}})$ ist ebenfalls ein statistisches Modell.

Definition 2.2. $\mathscr{P}^X$ heißt die *zum statistischen Experiment* $(\Omega, \mathscr{A}, \mathscr{P}, X)$ *gehörende Familie von Stichprobenverteilungen*

In dem genannten Experiment wird nur x beobachtet, $(\Omega, \mathscr{A}, \mathscr{P})$ und X sind Hilfskonstruktionen. $\mathscr{P}^X$ ist eine bekannte Familie, unter welcher wahren Verteilung $\mathbb{P}^X \in \mathscr{P}$ die Stichprobe realisiert wurde, ist unbekannt.
Ziel ist es, aus der Kenntnis von x Schlüsse auf das „wahre" $\mathbb{P}^X$ zu ziehen und die Güte dieser Rückschlüsse zu bewerten. Häufig indiziert man $\mathscr{P}$ der besseren Handhabung wegen, das heißt, man setzt $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und entsprechend $\mathscr{P}^X = (\mathbb{P}_\vartheta^X,\ \vartheta \in \Theta)$. Θ heißt *Parametermenge* oder *Parameterraum*.

Gilt $\Theta \subseteq \mathbb{R}^k$ für ein $k \geq 1$, so nennt man $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein *parametrisches statistisches Modell* (*k-parametrisches Modell*). Läßt sich Θ dagegen nicht durch endlich viele Parameter beschreiben, so spricht man von einem *nichtparametrischen Modell*.
Im Fall $E = \mathbb{R}^n, \mathscr{E} = \mathscr{B}^n$ ist X ein zufälliger Vektor, x eine seiner Realisierungen. Wie bereits erwähnt, bezeichnet man in diesem Fall X als eine (mathematische) Stichprobe und x im Unterschied dazu als eine konkrete Stichprobe.

Im Fall, daß E ein Funktionenraum ist, bildet $X = (X_t,\ t \in T)$ einen stochastischen Prozess und x eine Realisierung desselben. In diesem Fall ist der Begriff „Stichprobe" weniger gebräuchlich. Man spricht von *Trajektorien* oder *Pfaden*. Wir werden in dieser Vorlesung in jedem der genannten Fälle X bzw. x als mathematische bzw. konkrete Stichprobe bezeichnen.

2.2 Klassische Statistische Experimente

Die Wahrscheinlichkeitsverteilungen $\mathbb{P}_\vartheta^X,\ \vartheta \in \Theta$, der Stichprobe X bilden eine Ausgangsbasis der mathematischen Statistik.
Auf ihrer Grundlage werden Schätzer für das unbekannte ϑ oder Tests für Hypothesen über ϑ konstruiert und untersucht. Für ihre Beschreibung bedient man sich der sogenannten Likelihoodfunktion, die wir vorerst in zwei Beispielen definieren.

Es sei $(C, \mathscr{C})$ ein meßbarer Raum. Wir setzen $E = C^n, \mathscr{E} = \mathscr{C}^{\otimes n}$ und $X = (X_1, X_2, \ldots, X_n)$, wobei $X_k : (\Omega, \mathscr{A}) \longrightarrow (C, \mathscr{C})$ für jedes $k = 1, \ldots, n$ eine Zufallsgröße ist.
Es sei weiterhin $\mathscr{P}$ eine Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathscr{A})$, $\mathscr{P} = \{\mathbb{P}_\vartheta, \vartheta \in \Theta\}$. Unter folgenden Voraussetzungen

Voraussetzung 1 Für jedes $\mathbb{P}_\vartheta \in \mathscr{P}$ sind die Zufallsgrößen $X_1, \ldots, X_n$ identisch verteilt, das heißt,

$$\mathbb{P}_\vartheta(X_k \in B) = \mathbb{P}_\vartheta(X_1 \in B), \quad B \in \mathscr{C}, \quad k = 1, \ldots, n.$$

Voraussetzung 2 Für jedes $\mathbb{P}_\vartheta \in \mathscr{P}$ sind die $X_1, \ldots, X_n$ unter $\mathbb{P}_\vartheta$ in ihrer Gesamtheit unabhängig, d.h.

$$\mathbb{P}_\vartheta(X \in B_1 \times \ldots \times B_n) = \prod_{i=1}^{n} \mathbb{P}_\vartheta(X_i \in B_i).$$

gilt dann

$$\mathbb{P}_\vartheta^X(B_1 \times \ldots \times B_n) = \prod_{i=1}^{n} \mathbb{P}_\vartheta(X_1 \in B_i) = \prod_{i=1}^{n} \mathbb{P}_\vartheta^{X_1}(B_i) \tag{2.1}$$

Durch $\mathbb{P}_\vartheta^X$ ist ein Wahrscheinlichkeitsmaß auf $\mathscr{C}^{\otimes n}$ definiert, das wir ebenfalls mit $\mathbb{P}_\vartheta^X$ bezeichnen.

Beispiel 2.1 (diskreter Fall). Es seien $C = \{c_1, c_2, \ldots\}$, $\mathscr{C} = \mathscr{P}(C)$ die Potenzmenge von C, $E = C^n$ und $\mathscr{E} = \mathscr{C}^{\otimes n}$. X_1 nehme nur Werte aus C an, d.h. X_1 habe eine diskrete Verteilung mit

$$\mathbb{P}_\vartheta(X_1 = c_k) = p_k(\vartheta), \quad p_k(\vartheta) \geq 0, \quad k = 1, 2, \ldots, \quad \sum_{k=1}^{\infty} p_k(\vartheta) = 1.$$

Dann hat auch $X = (X_1, X_2, \ldots, X_n)$ eine diskrete Verteilung und es gilt:

$$\mathbb{P}_\vartheta^X(\,(c_{i_1}, \ldots, c_{i_n})\,) = \prod_{k=1}^{n} p_\vartheta(c_{i_k}) = \mathbb{P}_\vartheta(X_1 = c_{i_1}, \ldots, X_n = c_{i_n}) =: L_n(x, \vartheta), \quad x = (c_{i_1}, \ldots, c_{i_n})$$

Diese Funktion L_n heißt *Likelihoodfunktion* des statistischen Experiments $(\Omega, \mathscr{A}, \mathscr{P}, X)$.

Beispiel 2.2 (stetiger Fall). Es seien $C = \mathbb{R}$, $E = \mathbb{R}^n$ und $\mathscr{E} = \mathscr{B}(\mathbb{R}^n)$. X_1 besitze eine Dichte $f_\vartheta(x)$, d. h. es gelte

$$\mathbb{P}_\vartheta^X(D) = \int_D f_\vartheta(s)\, ds, \quad D \in \mathscr{B}.$$

Dann besitzt auch $X = (X_1, \ldots, X_n)$ eine Dichte $f_\vartheta^X(x_1, \ldots, x_n) = \prod_{i=1}^n f_\vartheta(x_i)$ und es gilt

$$\mathbb{P}_\vartheta^X(B_1 \times \ldots \times B_n) = \int_{B_1} \cdots \int_{B_n} f_\vartheta(x_1) \cdot \ldots \cdot f_\vartheta(x_n)\, dx_1 \ldots dx_n.$$

In diesem Fall bezeichnet man $L_n(x; \vartheta) := \prod_{k=1}^n f_\vartheta(x_k)$ als *Likelihoodfunktion* des statistischen Experiments $(\Omega, \mathscr{A}, \mathscr{P}, X)$.

Interpretation: Die Stichprobe $X = (X_1, \ldots, X_n)$ modelliert n voneinander unabhängige, unter gleichartigen Bedingungen ausgeführte zufällige Experimente, bei denen jeweils $X_1, X_2, \ldots, X_n$ beobachtet wird. Wir definieren $Q_\vartheta := \mathbb{P}_\vartheta^{X_1}$.
Man sagt, X sei eine *mathematische Stichprobe aus einer nach Q_ϑ verteilten Grundgesamtheit.*
Jede ihrer Realisierungen x nennt man eine *konkrete Stichprobe aus einer nach Q_ϑ verteilten Grundgesamtheit.*
Bezeichnung: Klassisches statistisches Experiment.

Wir kehren zurück zu allgemeinen statistischen Experimenten.

Definition 2.3. Es seien $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Experiment mit dem Stichprobenraum $(E, \mathscr{E})$ und H eine meßbare Abbildung von $(E, \mathscr{E})$ in einen meßbaren Raum $(F, \mathscr{F})$. H heißt eine *Stichprobenfunktion*. Insbesondere ist $(\Omega, \mathscr{A}, \mathscr{P}, H \circ X)$ ein statistisches Experiment mit dem Stichprobenraum $(F, \mathscr{F})$.

Im allgemeinen geht bei dieser Abbildung H Information verloren (Datenreduktion), andererseits kann $H(x)$ einfacher und übersichtlicher sein als x.

Setzt man in H die Zufallsgröße X ein, d. h. $H(X) = H\big((X_1, \ldots, X_n)\big)$, so erhält man eine neue Zufallsgröße, sie hat die Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta^H(B) = \mathbb{P}_\vartheta(H(X) \in B), \quad B \in \mathscr{F}$. Diese Wahrscheinlichkeitsverteilung wird unter anderem zum Studium der Eigenschaften der Stichprobe x in ihrem Verhältnis zum wahren Parameter ϑ herangezogen. Die Berechnung der Verteilungen von Stichprobenfunktionen $H(X)$ gehört zu den wesentlichen Aufgaben der mathematischen Statistik.

Anstelle von Stichprobenfunktion verwendet man auch einfach die Bezeichnung *Statistik*.

Wir nehmen an, daß $\mathscr{P}$ die Form $\mathscr{P} = \{\mathbb{P}_\vartheta,\ \vartheta \in \Theta\}$ mit irgendeiner Parametermenge Θ hat. Eine Grundaufgabe der Statistik ist es, von der Beobachtung x auf den Parameter ϑ bzw. eine Funktion $\gamma(\vartheta)$ zu schließen. Häufig möchte man ϑ bzw. $\gamma(\vartheta)$ mit möglichst großer Genauigkeit bestimmen, man sagt „schätzen".

Definition 2.4. Es seien γ eine Abbildung von Θ in eine Menge Γ, $\mathscr{A}_\Gamma$ eine σ-Algebra von Teilmengen von Γ und g eine meßbare Abbildung von $(E, \mathscr{E})$ in $(\Gamma, \mathscr{A}_\Gamma)$.
Dann heißt g eine *Schätzfunktion*, $g(X)$ ein *Schätzer* und $g(x)$ ein *Schätzwert* für $\gamma(\vartheta)$.

Jeder Schätzer ist also auch eine Stichprobenfunktion.

2.3 Ein Beispiel aus der klassischen Mathematischen Statistik

Es sei X_0 eine reellwertige Zufallsgröße mit unbekannter Verteilungsfunktion F. Zu schätzen sei der Erwartungswert

$$E(X_0) = \int_{\mathbb{R}} x\, dF(x) =: m_F.$$

Da man über F keine Vorinformation hat, setzt man

$$\Theta = \{F \,:\, F \text{ Verteilungsfunktion auf } (\mathbb{R}, \mathscr{B}) \text{ mit } |m_F| < \infty\}$$

Die Problemstellung legt nahe $\gamma(F) = m_F$ zu setzen.

Vorausgesetzt werde ferner, daß eine n-elementige Stichprobe $x = (x_1, x_2, \ldots, x_n)$ vorliegt, die aus n voneinander unabhängigen unter gleichartigen Bedingungen durchgeführten Versuchen gewonnen wurden. Dabei wird beim k-ten Versuch, $k = 1, \ldots, n$ registriert, welchen Wert die Zufallsgröße X_0 annimmt. Intuitiv verwenden wir als Schätzwert $g(x)$ für $\gamma(F) = m_F$ den Wert

$$g(x) = \overline{x}_n = \frac{1}{n} \sum_{k=1}^{n} x_k.$$

Die eben getroffenen Voraussetzungen legen es nahe, eine mathematische Stichprobe $X = (X_1, X_2, \ldots, X_n)$ zu betrachten, die aus n voneinander unabhängigen und identisch wie X_0 verteilten Zufallsgrößen besteht.
Als Schätzer für m_F ergibt sich

$$g(X) = \overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

Zur Illustration typischer Aussagen der Mathematischen Statistik stellen wir eine Reihe von Eigenschaften dieses Schätzers zusammen.

Aussage: Es gelten folgende Eigenschaften

a) $E_F(\overline{X}_n) = m_F$, man sagt, $\overline{X}_n$ ist ein *erwartungstreuer Schätzer*

b) $D_F^2\, \overline{X}_n = E_F\big((\overline{X}_n - m_F)^2\big) = \frac{D_F^2 X_0}{n}$, falls $\sigma^2 := D_F^2 X_0 < \infty$.
Insbesondere gilt

$$\mathbb{P}_F(|\overline{X}_n - m_F| \geq a) \leq \frac{\sigma_F^2}{na^2} \qquad a > 0, \quad n \geq 1$$

(Schwaches Gesetz der großen Zahlen). Man sagt, daß der Schätzer $\overline{X}_n$ *konsistent* ist.

c) $\lim_{n\to\infty} \overline{X}_n = m_F \quad \mathbb{P}_F$-fast sicher (Starkes Gesetz der großen Zahlen)

d) Angenommen,

$$I_F := \{t \in \mathbb{R} \, : \, E_F(e^{tX}) < \infty\}$$

ist eine Umgebung von 0. h_F sei die Cramertransformierte von F (sieh Übungen), ϵ sei irgendeine positive Zahl. Dann ist

$$P_F(\overline{X}_n \geq m_F + \epsilon) \leq \exp\{-nh_F(m_F+\epsilon)\} \text{ und } P_F(\overline{X}_n \leq m_F - \epsilon) \leq \exp\{-nh_F(m_F-\epsilon)\},$$

das bedeutet insbesondere, daß $P_F(\, |\overline{X}_n - m_F| \geq \epsilon)$ exponentiell schnell gegen 0 konvergiert.

e) Der Zentrale Grenzwertsatz besagt, daß

$$P_F(\, |\overline{X}_n - m_F| > \epsilon) = P_F\big(\, |\overline{X}_n - m_F| \frac{\sqrt{n}}{\sigma} > \epsilon \frac{\sqrt{n}}{\sigma}\big) \approx 2\Big(1 - \Phi(\epsilon\, \sigma_F^2 \sqrt{n})\Big).$$

Beide Methoden d) und e) führen zur Abschätzung der Genauigkeit der Approximation von m_F durch $\overline{X}_n$.

Die erste Methode liefert:

> Die Wahrscheinlichkeit, daß man sich irrt, wenn man sagt, m_F befinde sich in $(-\infty, \overline{X}_n + \epsilon)$ und in $(\overline{X}_n - \epsilon, \infty)$ konvergiert mit wachsendem n exponentiell schnell gegen Null.

Die zweite Methode liefert:

> Es wird ein $\alpha \in (0,1)$ fixiert und man erhält für große n Vertrauens-Intervalle
>
> $$\Big(-\infty, \overline{X}_n + \frac{\sigma_F}{\sqrt{n}} q_{1-\alpha}\Big],\ \Big[\overline{X}_n - \frac{\sigma_F}{\sqrt{n}} q_{1-\alpha}, \infty\Big),\ \Big[\overline{X}_n - \frac{\sigma_F}{\sqrt{n}} q_{1-\frac{\alpha}{2}}, \overline{X}_n + \frac{\sigma_F}{\sqrt{n}} q_{1-\frac{\alpha}{2}}\Big)$$
>
> des Niveaus $1 - \alpha$, von denen man sagen kann, daß m_F mit einer Irrtumswahrscheinlichkeit nahe bei α in denen liegt.

2.4 Empirische Schätzer

Klassischer Fall:
Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Experiment mit dem Stichprobenraum $E = \mathbb{R}^n$, $\mathscr{E} = \mathscr{B}^n$ und $X = (X_1, X_2, \ldots, X_n)$ bestehe aus reellwertigen, unabhängigen und identisch verteilten Zufallsgrößen $X_k,\ k = 1, \ldots, n$ mit Verteilungsfunktion F. Die Familie $\mathscr{P}$ sei parametrisiert: $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta),\ \Theta \subseteq \mathbb{R}^k$.

Empirische Verteilungsfunktion

Es seien

$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{X_n \leq x\}}, \qquad m_l(F) = \int_{\mathbb{R}} x^l \, dF(x), \quad l \in \mathbb{N}$$

Diese Verteilungsfunktion $\hat{F}_n(\,\cdot\,)$ gehört zur gleichmäßigen Verteilung auf $(X_1, X_2, \ldots, X_n)$. Sie hat $m_l(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} X_k^l$ als l-tes Moment.

Aussage 2.1 (Hauptsatz der mathematischen Statistik). *Es sei $(X_n,\, n \geq 1)$ eine Folge unabhängiger, identisch nach F verteilter Zufallsgrößen mit Werten in $\mathbb{R}^k$.*
Dann konvergiert die Folge $\big(\hat{F}_n(\,\cdot\,, \omega)\big)$ schwach gegen $F(\,\cdot\,)$ (d.h. $\hat{F}_n(f) \longrightarrow F(f), \forall f \in \mathbb{C}_b$). Ist $k = 1$, so erfolgt die Konvergenz $\mathbb{P}$-f.s. gleichmäßig, d.h.

$$\sup_{x \in \mathbb{R}} \mid \hat{F}_n(x, \omega) - F(x) \mid \longrightarrow 0 \qquad \textit{für } \mathbb{P}\textit{-f.a. } \omega \in \Omega.$$

Beweis:

s. Dacunha-Castelle, Duflo I, 4.4 □

Eine Konstruktionsmethode für Schätzer: Empirische Schätzer

Ist $\gamma(F)$ die zu schätzende Größe, so verwendet man $\gamma(\hat{F}_n)$ als Schätzer (sofern γ auf den Treppenfunktionen definiert ist, bzw. Sinn macht).

a) Momentenmethode
Zu schätzen sind der Erwartungswert $\mu = m_1(F) = \int_{\mathbb{R}} x \, dF(x)$ und die Streuung $\sigma_F^2 = m_2(F) - \big(m_1(F)\big)^2$.

Wir wenden die Abbildungen $F \longrightarrow m_1(F)$ und $F \longrightarrow \sigma_F^2$ auf $\hat{F}_n$ an und erhalten die „Momentenschätzer“

$$\hat{\mu} = m_1(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} X_k = \bar{X}_n, \qquad \text{(erwartungstreu)}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \big(\frac{1}{n}\sum_{k=1}^{n} X_k\big)^2 = \frac{1}{n}\sum_{k=1}^{n}\big(X_k - \bar{X}_k\big)^2.$$

Allgemeiner: Man berechne die Momente $m_l(F_\vartheta) = \mathbb{E}_\vartheta(X_1^l)$, ersetze $m_l(F_\vartheta)$ durch $m_l(\hat{F}_n)$ und löse die Gleichungen nach ϑ (bzw. nach $\gamma(\vartheta)$) auf. Im Ergebnis erhält man einen Schätzer $\hat{\vartheta}_n$ bzw. $\hat{\gamma}_n$ für ϑ bzw. $\gamma(\vartheta)$, einen sogenannten Momentenschätzer.

Beispiel 2.3. Es sei F_λ die Verteilungsfunktion der Exponentialverteilung mit Parameter $\lambda > 0$.
In diesem Fall gilt:

$$m_1(F) = \lambda \int_0^\infty x e^{-\lambda x}\, dx = \frac{1}{\lambda}, \quad \bar{X}_n = \widehat{\left(\frac{1}{\lambda}\right)}, \quad \text{wobei } \gamma(F_\lambda) = \frac{1}{\lambda}\,.$$

$\bar{X}_n$ ist ein erwartungstreuer Schätzer für $\gamma(\lambda) = \frac{1}{\lambda}$. Ein möglicher Schätzer für λ wäre zum Beispiel $\dfrac{1}{\bar{X}_n}$. Es wird sich aber herausstellen, daß dieser Schätzer nicht erwartungstreu ist.

Wir kehren zur Schätzung von σ_F^2 zurück:

$$\hat{\sigma}_F^2 := \frac{1}{n}\sum_{k=1}^{n}\big(X_k - \bar{X}_k\big)^2 \qquad \text{ist die Momentenschätzung für } \sigma_F^2.$$

Es gilt:

$$\begin{aligned}
E_F(\hat{\sigma}_F^2) &= \frac{1}{n} E_F\left(\left(\sum_{k=1}^{n} X_k^2\right) - n\bar{X}_n^2\right) \\
&= \frac{1}{n}\sum_{k=1}^{n} E_F X_k^2 - E_F \bar{X}_n^2 \\
&= \frac{1}{n}\left(\left(\sigma_F^2 + \mu_F^2\right) n\right) - \frac{\sigma_F^2}{n} - \mu^2 \\
&= \sigma_F^2 + \mu_F^2 - \frac{\sigma_F^2}{n} - \mu^2 \\
&= \sigma_F^2 \left(\frac{n-1}{n}\right)
\end{aligned}$$

Also ist $\hat{\sigma}_F^2$ nicht erwartungstreu, man unterschätzt σ_F^2 regelmäßig. Aber

$$S_n^2 = \frac{1}{n-1}\hat{\sigma}_F^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}_n\right)^2$$

ist eine erwartungstreue Schätzung für σ_F^2.

Für das Beispiel 2.3 gilt dann:
S_n^2 ist eine erwartungstreue Schätzung für $\frac{1}{\lambda^2}$, $\sqrt{S_n^2}$ ist eine Schätzung für $\frac{1}{\lambda}$.

Beispiel 2.4. Es sei $X = (X_!, X_2, \ldots, X_n)$ eine klassische mathematische Stichprobe aus einer Grundgesamtheit, die eine gemischte Poissonverteilung besitzt:

$$\mathbb{P}_\vartheta(X_1 = k) = a \cdot \frac{\lambda_1^k}{k!}e^{-\lambda_1} + (1-a)\,\frac{\lambda_2^k}{k!}e^{-\lambda_2}, \qquad k \geq 0$$

mit $\vartheta = (a, \lambda_1, \lambda_2)\,, \quad a \in (0,1)\,, \lambda_1, \lambda_2 > 0$. Die entsprechende Verteilungsfunktion werde mit F_ϑ bezeichnet.
Für die momenterzeugende Funktion

$$\varphi_\vartheta(s) := \mathbb{E}_\vartheta\left(s^{X_1}\right) = \sum_{k\geq 0}\mathbb{P}_\vartheta(X_1 = k)s^k = ae^{\lambda_1(s-1)} + (1-a)e^{\lambda_2(s-1)}, \quad s \in [0,1]$$

gilt:

$$\begin{aligned}
T_1(F_\vartheta) &:= \varphi_\vartheta'(1) = \mathbb{E}_\vartheta\, X_1 = a\lambda_1 + (1-a)\lambda_2 \\
T_2(F_\vartheta) &:= \varphi_\vartheta''(1) = \mathbb{E}_\vartheta\, X(X-1) = a\lambda_1^2 + (1-a)\lambda_2^2 \\
T_3(F_\vartheta) &:= \varphi_\vartheta'''(1) = \mathbb{E}_\vartheta\, X(X-1)(X-2) = a\lambda_1^3 + (1-a)\lambda_2^3
\end{aligned}$$

Wir definieren die entsprechenden empirischen Momente

$$\begin{aligned}
T_1(\hat{F}_n) &= \frac{1}{n}\sum_{k=1}^{n} X_k, \qquad T_2(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} X_k(X_k - 1), \\
T_3(\hat{F}_n) &= \frac{1}{n}\sum_{k=1}^{n} X_k(X_k-1)(X_k-2).
\end{aligned}$$

Ist $x = (x_1, x_2, \ldots, x_n)$ eine konkrete Stichprobe aus der nach F_ϑ verteilten Grundgesamtheit, so erhält man folgende Gleichungen, aus denen sich die empirischen Schätzer

$\hat{a}, \hat{\lambda}_1, \hat{\lambda}_2$ für $\vartheta = (a, \lambda_1, \lambda_2)$ berechnen lassen:

$$\begin{aligned}
\hat{a}\hat{\lambda}_1 + (1-\hat{a})\,\hat{\lambda}_2 &= T_1(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} x_k \\
\hat{a}\hat{\lambda}_1^2 + (1-\hat{a})\,\hat{\lambda}_2^2 &= T_2(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} x_k(x_k-1) \\
\hat{a}\hat{\lambda}_1^3 + (1-\hat{a})\,\hat{\lambda}_2^3 &= T_3(\hat{F}_n) = \frac{1}{n}\sum_{k=1}^{n} x_k(x_k-1)(x_k-2).
\end{aligned}$$

b) Schätzung der Schranken des Trägers von F:

$$\begin{aligned}
m &= \sup\{a \in \mathbb{R} : F(a) = 0\}, & M &= \inf\{a \in \mathbb{R} : F(a) = 1\} \\
\hat{m}_n &= \min\{X_k,\ k = 1, \dots, n\}, & \hat{M}_n &= \max\{X_k,\ k = 1, \dots, n\}
\end{aligned}$$

2.5 Eigenschaften von Schätzern

Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Experiment mit dem Stichprobenraum $(E, \mathscr{E})$ und es sei $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$. Weiterhin sei γ wie oben eine meßbare Funktion von Θ in Γ und $g(X)$ ein Schätzer für $\gamma(\vartheta)$.
Der Einfachheit halber nehmen wir an, $\Gamma \subseteq \mathbb{R}$. Zur Beurteilung der Güte des Schätzers $g(X)$ definieren wir die *Risikofunktion*

$$R(g,\, \vartheta) := \mathbb{E}_\vartheta \left(g(X) - \gamma(\vartheta)\right)^2 \quad \vartheta \in \Theta$$

$R(g,\, \vartheta)$ ist also die mittlere quadratische Abweichung des Schätzers $g(X)$ von dem zu schätzenden Wert $g(\vartheta)$, wenn $\mathbb{P}_\vartheta$ die zugrundeliegende Wahrscheinlichkeitsverteilung ist.
Man wähle die Schätzfunktion $g(\cdot)$ so, daß $R(g,\, \vartheta)$ möglichst minimal wird.
Unter der Voraussetzung, daß $R(g,\, \vartheta) < \infty, \quad \forall\, \vartheta \in \Theta$ definiere:

Definition 2.5. Ein Schätzer $h(X)$ für die Funktion $\gamma(\vartheta)$ heißt *besser als* $g(X)$, falls gilt:

$$R(h,\, \vartheta) \leq R(g,\, \vartheta) \text{ für alle } \vartheta \in \Theta \text{ und } R(h,\, \vartheta) < R(g,\, \vartheta) \text{ für mindestens ein } \vartheta \in \Theta.$$

Wenn es zu einem gegebenen Schätzer $g(X)$ für $\gamma(\vartheta)$ einen besseren Schätzer $h(X)$ für $\gamma(\vartheta)$ gibt, so nennt man $g(X)$ *nicht zulässig* (als Schätzer für $\gamma(\vartheta)$).
$g(X)$ heißt *zulässiger Schätzer* für $\gamma(\vartheta)$, falls es keinen besseren Schätzer für $\gamma(\vartheta)$ gibt.

Es ist vernünftig, sich auf zulässige Schätzer zu beschränken.

Definition 2.6. Ein Schätzer $g^*(X)$ für $\gamma(\vartheta)$ heißt *optimal*, falls

$$R(g^*,\,\vartheta) = \inf_g R(g,\vartheta) \qquad \text{für alle } \vartheta \in \Theta,$$

wobei das Infimum über alle (zulässigen) Schätzer $g(X)$ für $\gamma(\vartheta)$ gebildet wird.

Im allgemeinen gibt es keinen optimalen Schätzer für $\gamma(\vartheta)$!

Begründung: Für jedes fest gewählte $\vartheta_0 \in \Theta$ ist $\inf_g R(g,\vartheta_0) = 0$, da der Schätzer $g(X) \equiv \gamma(\vartheta_0)$ unter allen konkurrierenden Schätzern vorkommt.
Dieser Schätzer ist sehr gut, wenn $\vartheta = \vartheta_0$ der wahre Parameter ist, aber für andere ϑ allerdings nicht.

Wir verfolgen unser Ziel, eine vernünftige Schätzfunktion g zu finden, die das mittlere quadratische Risiko möglichst klein hält, durch folgende Überlegungen:
Es gilt

$$\begin{aligned} R(g,\,\vartheta) = &\quad \mathbb{E}_\vartheta\Big(g(X) - \mathbb{E}_\vartheta\big(g(X)\big)\Big)^2 &&+ \mathbb{E}_\vartheta\Big(\mathbb{E}_\vartheta\big(g(X)\big) - \gamma(\vartheta)\Big)^2 \\ =: &\qquad\qquad D^2_\vartheta\, g(X) &&+ b_{g,\gamma}(\vartheta), \qquad \vartheta \in \Theta \\ \text{Risikofunktion} = &\quad \text{zufallsbedingte Streuung} &&+ \text{Verzerrung}^2. \end{aligned} \tag{2.2}$$

Die Größe $b_{g,\gamma}(\vartheta)$ heißt *Verzerrung* oder *Bias* des Schätzers $g(X)$ bezüglich $\gamma(\vartheta)$.
Wenn man das Risiko $R(\cdot)$ für alle $\vartheta \in \Theta$ minimieren will, ist es also vernünftig, unter den erwartungstreuen (unverzerrten) Schätzern, d.h. Schätzern mit $\mathbb{E}_\vartheta\big(g(X)\big) = \gamma(\vartheta),\ \vartheta \in \Theta$ zu suchen.
Wir beschränken uns deshalb darauf, unverzerrte Schätzer mit möglichst kleiner Streuung zu suchen.

Angenommen $H(X)$ ist eine Stichprobenfunktion und $g(X)$ ist ein Schätzer für $\gamma(\vartheta)$. Eine ähnliche Rechnung wie in (2.2) führt auf

$$R(g,\,\vartheta) = \mathbb{E}_\vartheta\Big(g(X) - \mathbb{E}_\vartheta\big(g(X) \mid H(X)\big)\Big)^2 + \mathbb{E}_\vartheta\Big(\mathbb{E}_\vartheta\big(g(X) \mid H(X)\big) - \gamma(\vartheta)\Big)^2$$

Definition 2.7. Die Stichprobenfunktion $H(X)$ heißt eine *suffiziente* oder *erschöpfende* Statistik, falls die Wahrscheinlichkeitsverteilung $\mathbb{P}^X_\vartheta\big(\,\cdot\, \mid H(X)\big)$ nicht von ϑ abhängt.

Wir kommen auf diesen Begriff später ausführlich zurück.

Ist $H(X)$ eine suffiziente Statistik, so ist $g_1(X) := \mathbb{E}_\vartheta \big(g(X) \mid H(X)\big)$ ein neuer Schätzer für $\gamma(\vartheta)$ mit kleinerem Risiko ($g_1(X)$ hängt nicht von ϑ ab):

$$R(g_1,\, \vartheta) = \mathbb{E}_\vartheta \big(g_1(X) - \gamma(\vartheta)\big)^2 \leq R(g,\, \vartheta), \quad \vartheta \in \Theta.$$

Da $g_1(X)$ eine Funktion von $H(X)$ ist, genügt es also, als Schätzer für $\gamma(\vartheta)$ lediglich Funktionen suffizienter Statistiken in Betracht zu ziehen.

Beispiel einer suffizienten Statistik

Ein zufälliger Versuch möge nur die Ausgänge 0 und 1 besitzen, wobei 1 mit der unbekannten Wahrscheinlichkeit $\vartheta \in (0,\, 1) =: \Theta$ als Ergebnis auftritt. Zur Schätzung von ϑ verschafft man sich eine Stichprobe $(i_1, i_2, \ldots, i_n)$ aus n unabhängig und unter gleichartigen Bedingungen durchgeführten Versuchen.
Zur mathematischen Modellierung dieses Sachverhalts führen wir ein:

$$\begin{aligned}
&\Omega = \big\{\omega = (i_1, i_2, \ldots, i_n) \,:\, i_k \in \{0\,,1\},\ k = 1, 2, \ldots, n\big\}, \quad \mathscr{A} = \wp\,(\Omega),\\
&X = (X_1, X_2, \ldots, X_n) \text{ mit } X_k(\omega) = i_k, \quad k = 1, \ldots, n \text{ für } \omega = (i_1, i_2, \ldots, i_n) \in \Omega,\\
&S_n = \sum_{k=1}^{n} X_k \text{ und } \mathbb{P}_\vartheta(\{\omega\}) = \vartheta^{\sum_{k=1}^n i_k} (1-\vartheta)^{n - \sum_{k=1}^n i_k}, \quad \omega \in \Omega.
\end{aligned}$$

Bezüglich jedem $\mathbb{P}_\vartheta$ sind $X_1, X_2, \ldots, X_n$ unabhängige und identisch verteilte Zufallsgrößen mit $\mathbb{P}_\vartheta(X_k = i) = \vartheta^i (1-\vartheta)^{1-i}, \quad i \in \{0\,,1\}$.

Aussage 2.2. *$S_n = \sum_{k=1}^n X_k$ ist eine suffiziente Statistik.*

Beweis:

Es gilt:

$$\mathbb{P}_\vartheta\big(\{\omega\} \mid S_n = m\big) = \begin{cases} \dfrac{1}{\binom{n}{m}}, & \text{falls } \omega = (i_1, i_2, \ldots, i_n) \text{ mit } \sum_{k=1}^n i_k = m \\ 0, & \text{falls } \omega = (i_1, i_2, \ldots, i_n) \text{ mit } \sum_{k=1}^n i_k \neq m \end{cases}$$

□

Zur Schätzung von ϑ beschränken wir uns also auf Funktionen der Form $g(S_n)$.
Soll $g(S_n)$ erwartungstreu sein, so erhalten wir aus der Forderung

$$\mathbb{E}_\vartheta\, g(S_n) = \vartheta, \quad \vartheta \in [0, 1]$$

die Gleichungen

$$\sum_{m=0}^{n} g(m)\binom{n}{m}\vartheta^m(1-\vartheta)^{n-m} = \vartheta, \quad \vartheta \in [0,1],$$

woraus sich durch Koeffizientenvergleich

$$g(m) = \frac{m}{n}, \quad m = 0,1,\dots,n, \quad \text{d.h. } \hat{\vartheta}_n = \frac{S_n}{n}$$

ergibt. Wir werden später zeigen, daß diese Schätzung die minimale Streuung unter allen erwartungstreuen Schätzungen für ϑ hat.
Es gilt:

$$\mathbb{P}_\vartheta\big(\{\omega\}\big) = \sum_{m=0}^{n} \mathbb{P}\big(\{\omega\} \mid S_n = m\big)\, \mathbb{P}_\vartheta(S_n = m)$$

Die Abhängigkeit des Maßes $\mathbb{P}_\vartheta$ von ϑ auf $\mathscr{A}$ wird also allein durch die Abhängigkeit von $\mathbb{P}_\vartheta^{S_n}$ von ϑ vermittelt.

Kapitel 3

Likelihoodschätzer im klassischen Fall

Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Experiment mit dem Stichprobenraum $E = \mathbb{R}^n$, $\mathscr{E} = \mathscr{B}^n$ und $X = (X_1, X_2, \ldots, X_n)$ bestehe aus reellwertigen, bezüglich jedem $\mathbb{P} \in \mathscr{P}$ unabhängigen und identisch verteilten Zufallsgrößen $X_k,\ k = 1, \ldots, n$. Die Familie $\mathscr{P}$ sei parametrisiert:
$\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta),\ \Theta \subseteq \mathbb{R}^k$, und
$F_\vartheta(x) := \mathbb{P}_\vartheta(X_1 \leq x),\ x \in \mathbb{R},\ \vartheta \in \Theta$ bezeichne die Verteilungsfunktion von X_1 bezüglich $\mathbb{P}_\vartheta$.
Dann hat X die Verteilungsfunktion

$$F_\vartheta^X(x_1, \ldots, x_n) = \prod_{m=1}^n F_\vartheta(x_m) = \mathbb{P}_\vartheta\big(X \in (-\infty, x_1] \times \ldots \times (-\infty, x_n]\big).$$

Besitzt F_ϑ eine Dichte f_ϑ, so hat X die Dichte

$$f_\vartheta^X(x_1, \ldots, x_n) = \prod_{m=1}^n f_\vartheta(x_m).$$

Das bedeutet $\mathbb{P}_\vartheta(X \in B) = \int_B f_\vartheta^X\, dx$ für alle Borelmengen $B \in \mathscr{B}$.
Ist X_1 diskret verteilt mit $\mathbb{P}_\vartheta(X_1 = a_m) = p_m(\vartheta)$, so ist auch X diskret verteilt und es gilt

$$\mathbb{P}_\vartheta\Big(X = (a_{m_1}, \ldots, a_{m_n})\Big) = \prod_{r=1}^n p_{m_r}(\vartheta)$$

In beiden Fällen nennt man (bei festgehaltener Stichprobe $x = (x_1, \ldots, x_n)$ bzw. $x = (a_{m_1}, \ldots, a_{m_n})$)

$$L_n(\vartheta; x_1, \ldots, x_n) = \prod_{m=1}^n f_\vartheta(x_m) \ \text{ bzw. } = \prod_{r=1}^n p_{m_r}(\vartheta), \quad \vartheta \in \Theta,$$

die *Likelihood-Funktion* des statistischen Experiments.

Maximum-Likelihood-Schätzer (ML-Schätzer)

Es sei x eine (konkrete) Stichprobe aus $E = \mathbb{R}^n,\ x \in \operatorname{supp}(X)$.

Definition 3.1. Jeder Wert $\hat{\vartheta}(x)$ aus Θ, der die Likelihoodfunktion $L_n(\cdot\,; x)$ bei gegebenem x maximiert, heißt ein *Maximum-Likelihood-Schätzwert* für ϑ auf der Basis der Stichprobe x:

$$\hat{\vartheta}_n(x) := \arg\max_{\vartheta\in\Theta} L_n(\vartheta\,; x) = \arg\max_{\vartheta\in\Theta} \log L_n(\vartheta\,; x)$$

Offenbar ist $\hat{\vartheta}_n(\,\cdot)$ eine Stichprobenfunktion. Setzt man die mathematische Stichprobe X ein, so erhält man einen sogenannten *Maximum-Likelihood-Schätzer* (kurz: ML-Schätzer) $\hat{\vartheta}_n(X)$.

> *R.A. Fisher: Maximum-Likelihood-Prinzip*
> „Finde diejenigen Voraussetzungen, die das Beobachtete mit großer Wahrscheinlichkeit nach sich ziehen und fasse Zutrauen, daß diese Voraussetzungen die wirksamen sind.“

Bemerkung: Maximum–Likelihood-Schätzer sind häufig einfach auszurechnen, haben vielfach gute Eigenschaften, existieren aber nicht immer oder sind nicht eindeutig. ML-Prinzip ist ein sehr allgemeines Prinzip, kann auch bei stochastischen Prozessen angewendet werden.

Maximum-Likelihood-Gleichungen

Unter der Voraussetzung, daß $L_n(\vartheta\,;\ x_1,\ldots,x_n)$ für jedes $(x_1,\ldots,x_n) \in \operatorname{supp}(X)$ bezüglich ϑ differenzierbar ist, sind

$$\frac{\partial}{\partial\vartheta_r} L_n(\hat{\vartheta}_n\,;\ x_1,\ldots,x_n) = 0 \qquad r = 1,2,\ldots,k \qquad \text{(„ML-Gleichungen“)} \quad (3.1)$$

notwendige Bedingungen für $\hat{\vartheta}_n$, ein ML-Schätzer zu sein, sofern das Maximum von L_n im Inneren von Θ angenommen wird.
Anstelle L_n führt man

$$l_n(\vartheta\,;\ x_1,\ldots,x_n) = \log L_n(\vartheta\,;\ x_1,\ldots,x_n) \quad \text{ein.} \qquad \text{(„Loglikelihoodfunktion“)}$$

Äquivalent zu (3.1) ist

$$\frac{\partial}{\partial\vartheta_r} l_n(\hat{\vartheta}_n\,;\ x_1,\ldots,x_n) = 0 \qquad r = 1,2,\ldots,k. \qquad (3.2)$$

Die Funktion

$$\dot{l}_n(\vartheta\,;\ x_1,\ldots,x_n) := \Big(\frac{\partial}{\partial\vartheta_r} l_n(\vartheta\,;\ x_1,\ldots,x_n), \quad r=1,2,\ldots,k\Big), \quad \vartheta\in\Theta$$

nennt man *Scorefunktion* des statistischen Experiments $(\Omega, \mathscr{A}, \mathscr{P}, X)$.
Es gilt

$$\begin{aligned} l_n(\vartheta\,;\ x_1,\ldots,x_n) &= \operatorname{grad} l_n(\vartheta\,;\ x_1,\ldots,x_n) \quad \text{und} \\ \mathbb{E}_\vartheta\Big(\dot{l}_n(\vartheta\,;\ X_1,\ldots,X_n)\Big) &= \mathbb{E}_\vartheta\Big(\frac{\dot{L}_n}{L_n}\Big) = \underset{\mathbb{R}^n}{\int\cdots\int} \dot{L}_n(\vartheta\,;\ x_1,\ldots,x_n)\,dx_1\ldots dx_n \\ &= \operatorname{grad} \underset{\mathbb{R}^n}{\int\cdots\int} L_n\,dx_1\ldots dx_n \\ &= 0 \end{aligned}$$

(Hier haben wir vorausgesetzt, daß Differentiation nach ϑ und Integration bezüglich x vertauschbar sind.)

Bemerkung:

$$l_n(\vartheta\,;\ X_1,\ldots,X_n) = \sum_{m=1}^{n} \log f_\vartheta(X_m)$$

ist eine Summe unabhängiger und identisch verteilter Zufallsgrößen.

$$\dot{l}_n(\vartheta\,;\ X_1,\ldots,X_n) = \operatorname{grad}_\vartheta l_n(\vartheta\,;\ X_1,\ldots,X_n)$$

ist eine zentrierte Summe unabhängiger identisch verteilter Zufallsvektoren.
Für jedes feste ϑ sind $L_n(\vartheta\,;\ X_1,\ldots,X_n)$, $l_n(\vartheta\,;\ X_1,\ldots,X_n)$ und $\dot{l}_n(\vartheta\,;\ X_1,\ldots,X_n)$ Stichprobenfunktionen.

Definition 3.2. Ist $E = \{x\in\mathbb{R} \mid f_\vartheta(x) > 0\}$ (bzw. $E = \{a\in\mathbb{R} \mid \mathbb{P}_\vartheta(X=a) > 0\}$) unabhängig von ϑ, so nennt man für ϑ, $\alpha\in\Theta$ und $x\in E$ den Quotienten

$$\frac{L_n(\alpha\,;\,x)}{L_n(\vartheta\,;\,x)} \qquad \text{den } \textit{Likelihoodquotienten.}$$

Ist $\log\frac{f_\alpha}{f_\vartheta}$ bezüglich F_ϑ integrierbar, so gilt

$$\frac{1}{n}\log\frac{L_n(\alpha\,;\,X_1,\ldots,X_n)}{L_n(\vartheta\,;\,X_1,\ldots,X_n)} \xrightarrow{\mathbb{P}_\vartheta-\text{f.s.}} \int\Big[\log\frac{f_\alpha}{f_\vartheta}\Big] f_\vartheta\,dx =: -K(F_\vartheta\,,\,F_\alpha)$$

$K(F_\vartheta\,,\,F_\alpha)$ heißt *Kullback-Information von F_ϑ bezüglich F_α.*

Lemma 3.3. *Es gilt:*

$$K(F_\vartheta\,,\,F_\alpha) \geq 0$$
$$K(F_\vartheta\,,\,F_\alpha) = 0 \Longleftrightarrow F_\vartheta = F_\alpha$$

Im Fall, daß X_1 eine diskrete Verteilung besitzt, gilt

$$K(F_\vartheta\,,\,F_\alpha) = -\sum_m \Big[\log\frac{p_m(\alpha)}{p_m(\vartheta)}\Big]p_m(\vartheta).$$

Beweis: (nur für den Dichtefall)
Die Funktion $h(x) = x\log x + 1 - x$ ist für $x > 0$ und $x \neq 1$ positiv und nur für $x = 1$ gleich Null. Folglich gilt:

$$\int\Big(\frac{f_\alpha}{f_\vartheta}\log\frac{f_\alpha}{f_\vartheta} + 1 - \frac{f_\alpha}{f_\vartheta}\Big)f_\vartheta\,dx \geq 0$$

und $K(F_\vartheta\,,\,F_\alpha) = 0$ impliziert $f_\alpha = f_\vartheta$.
Also gilt für $\alpha \neq \vartheta$ die Beziehung $-K(F_\vartheta\,,\,F_\alpha) < 0$ und somit

$$\log\frac{L_n(\alpha\,;\,X_1,\ldots,X_n)}{L_n(\vartheta\,;\,X_1,\ldots,X_n)} \xrightarrow{\mathbb{P}_\vartheta-\text{f.s.}} -\infty \quad \text{für } n \longrightarrow \infty,$$

mit anderen Worten, für $\alpha \neq \vartheta$ gilt

$$\frac{L_n(\alpha\,;\,X_1,\ldots,X_n)}{L_n(\vartheta\,;\,X_1,\ldots,X_n)} \xrightarrow{\mathbb{P}_\vartheta-\text{f.s.}} 0 \quad \text{für } n \longrightarrow \infty.$$

Andererseits ist offensichtlich der Quotient für $\alpha = \vartheta$ gleich Eins. □

Lemma 3.3 ist noch einmal ein Argument für die Vernünftigkeit des Maximum-Likelihood-Schätzers: $L_n(\alpha; X_1,\ldots,X_n)$ wird für α fernab von ϑ vergleichsweise zu $L_n(\vartheta; X_1,\ldots,X_n)$ klein sein (mit wachsendem n konvergiert der Quotient ja gegen Null), für α in der Nähe von ϑ auf Grund der Stetigkeit von $\alpha \longrightarrow L_n(\alpha; X_1,\ldots,X_n)$ nahe Eins. Verwendet man $\hat{\vartheta}_n(X_1, X_2,\ldots,X_n)$ als Schätzer für ϑ, so wird man also erwarten können, daß dieser Schätzer in der Nähe von ϑ liegt.

Eigenschaften der Maximum-Likelihood-Schätzer:

Wir geben hier zwei wichtige Eigenschaften von Maximum-Likelihood-Schätzern für den Fall von Stichproben an, die aus unabhängigen, identisch verteilten Zufallsgrößen bestehen. Für die keineswegs einfachen Beweise sei auf die Literatur verwiesen.

a) KONSISTENZ:

Im allgemeinen ist der Maximum-Likelihood-Schätzer nicht erwartungstreu, das heißt, es gilt i.a. nicht $\mathbb{E}_\vartheta\big(\hat\vartheta_n(X_1, X_2, \ldots, X_n)\big) = \vartheta$. Die folgende Eigenschaft der Konsistenz besagt aber, daß man für große Stichprobenumfänge den Schätzer $\hat\vartheta_n$ mit großer Wahrscheinlichkeit in der Nähe von ϑ finden wird. (Wir beschränken uns mit der Formulierung auf den Fall, daß $\mathbb{P}_\vartheta^{X_1}$ eine Dichte $f_\vartheta(x)$ hat.)

- sei Θ eine kompakte Teilmenge von $\mathbb{R}^n$
- es gelte $\{x \in \mathbb{R} \,:\, f_\alpha(x) > 0\}$ unabhängig von $\alpha \in \Theta$
- wenn $\alpha \neq \vartheta$, so $F_\alpha \neq F_\vartheta$ (Identifizierbarkeit des Modells bei ϑ)
- für alle $x \in \mathbb{R}$ sei $\alpha \longrightarrow f_\alpha(x)$ stetig
- es existiere eine $\mathbb{P}_\vartheta$-integrierbare Zufallsgröße H mit

$$\sup_\alpha |\log f(\alpha, X_1)| \leq H(\omega) \qquad \mathbb{P}_\vartheta\text{-f.s.}$$

Aussage 3.1. *Unter den genannten Bedingungen ist jeder Maximum-Likelihood Schätzer $\hat\vartheta_n$ konsistent im Sinne von*

$$\mathbb{P}_\vartheta\big(\,\|\,\hat\vartheta_n - \vartheta\,\| > \epsilon\big) \longrightarrow 0 \qquad \forall \epsilon > 0 \quad \forall \vartheta \in \Theta$$

Beweis:

s. Dacunha-Castelle, Duflo II, S. 126 f.

b) ASYMPTOTISCHE NORMALITÄT:

Wir stellen weiter einige Voraussetzungen an unser statistisches Modell.

Definition 3.4. Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$, $\Theta \subseteq \mathbb{R}^k$ und es sei $\vartheta \in \Theta$.
Dann heißt $(\Omega, \mathscr{A}, \mathscr{P}, X)$ *regulär bei* ϑ, falls Θ eine Umgebung von ϑ ist, und falls $L_n\big(\cdot\,;\, X(\omega)\big)$ wie folgt gewählt werden kann:

H1) In einer Umgebung V von ϑ mit $V \subseteq \Theta$ ist die Funktion $\alpha \longrightarrow L_n\big(\alpha\,;\, x\big)$ für jedes x zweimal stetig differenzierbar.

H2) grad $\log L_n\big(\vartheta\,;\, X(\cdot)\big)$ ist ein zentrierter Zufallsvektor mit endlichen zweiten Momenten bezüglich $\mathbb{P}_\vartheta$.

Außerdem gilt

$$
\begin{aligned}
&\mathbb{E}_\vartheta\Big(\frac{\partial}{\partial\vartheta_i}\log L_n(\vartheta\,;\,X)\cdot\frac{\partial}{\partial\vartheta_j}\log L_n(\vartheta\,;\,X)\Big)\\
&= \mathbb{E}_\vartheta\Big(\frac{\partial^2}{\partial\vartheta_i\,\partial\vartheta_j}\log L_n(\vartheta\,;\,X)\Big)\\
&=: I_n^{(i,j)}(\vartheta)
\end{aligned}
$$

Die $k\times k$ -Matrix $I_n(\vartheta) := \big(I_n^{(i,j)}(\vartheta)\big)_{i,j=1,\dots,k}$ heißt *Fisher-Informationsmatrix* für ϑ auf der Basis von $X=(X_1,X_2,\dots,X_n)$.

H3) $I_n(\vartheta)$ ist invertierbar.

Wir kehren zum ML-Schätzer zurück, betrachten aber nur den Fall, daß X_1 unter jedem $\mathbb{P}_\vartheta$ eine Dichte f_ϑ besitzt.

$$
l_n(\vartheta\,;\,X) = \log L_n(\vartheta\,;\,X) = \sum_{m=1}^{n}\log f_\vartheta(X_m)
$$

Wir setzen

$$
\begin{aligned}
&Y_n^i := \frac{\partial}{\partial\vartheta_i}l_n(\vartheta) = \sum_{m=1}^{n}\frac{\frac{\partial}{\partial\vartheta_i}f_\vartheta(X_m)}{f_\vartheta(X_m)} \qquad \text{und}\\
&Y_n := (Y_n^i)_{i=1,\dots,k} = \operatorname{grad} l_n(\vartheta)
\end{aligned}
$$

Die Vektoren

$$
\Big(\frac{\frac{\partial}{\partial\vartheta_i}f_\vartheta(X_m)}{f_\vartheta(X_m)}\Big)_{i=1,\dots,k}
$$

bilden für $m\geq 1$ bezüglich $\mathbb{P}_\vartheta$ unabhängige, identisch verteilte zentrierte Zufallsvektoren mit der Kovarianzmatrix $I_1(\vartheta)$ (Beachte H3).
Nach dem zentralen Grenzwertsatz für zufällige Vektoren gilt:

$$
\frac{1}{\sqrt{n}}Y_n(\vartheta) \xrightarrow{d(P_\vartheta)} \mathcal{N}_k(0\,,\,I_1(\vartheta))
$$

(Dacunha-Castelle, Duflo I, Seite 225).

Diese Eigenschaft führt nach einer Reihe weiterer Rechnungen auf die folgende

Aussage 3.2. *Es sei $(\Omega,\mathscr{A},\mathscr{P},X)$ ein an der Stelle $\vartheta\in\Theta$ reguläres statistisches Modell. Die Verteilung $\mathbb{P}_\vartheta^X$ habe bezüglich eines dominierenden Maßes μ die Dichte $f_\vartheta(x)$, $x\in E$, $\vartheta\in\Theta$. Es sei weiterhin $(X_1,X_2,\dots,X_n)$ eine klassische mathematische Stichprobe aus einer nach $\mathbb{P}_\vartheta^{X_1}$ verteilten Grundgesamtheit (d.h., $X_1,X_2,\dots,X_n$ seien unabhängige, identisch nach $\mathbb{P}_\vartheta^{X_1}$ verteilte Zufallsvariablen). Weiterhin gelte*

H4) *Es existiert eine Umgebung V von ϑ, $V \subseteq \Theta$, und eine $\mathbb{P}_\vartheta^{X_1}$-integrierbare Funktion H auf $\mathbb{R}^k$ mit*

$$\Big|\frac{\partial}{\partial\vartheta_i}\frac{\partial}{\partial\vartheta_j}\log f_\vartheta(x)\Big| \leq H(x) \quad \vartheta \in V, \quad i,j = 1,\ldots,n$$

Bezeichnet $\hat{\vartheta}_n(X_1, X_2, \ldots X_n)$ einen Maximum-Likelihood-Schätzer für ϑ, so gelte $\hat{\vartheta}_n \xrightarrow{\mathbb{P}_\vartheta} \vartheta$ (Konsistenz).

Dann haben wir:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \xrightarrow{d(P_\vartheta)} \mathcal{N}\big(0, I^{-1}(\vartheta)\big) \text{ und}$$
$$I(\vartheta)\sqrt{n}(\hat{\vartheta}_n - \vartheta) - \frac{1}{\sqrt{n}}\operatorname{grad} l_n(\vartheta) \xrightarrow{P_\vartheta} 0.$$

Zum Beweis dieser Aussage sei ebenfalls auf Dacunha-Castelle, Duflo II, S. 127, verwiesen.

Beispiele 3.1. a) Normalverteilung

Es seien $X_1, \ldots, X_n$ unabhängige, identisch $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsvariablen. Es sei $\vartheta = (\mu, \sigma^2)^T \in \mathbb{R} \times (0, \infty) =: \Theta$
Folglich erhalten wir

$$\log f_\vartheta(x) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2}\frac{(x-\mu)^2}{\sigma^2},$$
$$\operatorname{grad}\log f_\vartheta(x) = \begin{pmatrix} \dfrac{x-\mu}{\sigma^2} \\ -\dfrac{1}{2\sigma^2} + \dfrac{(x-\mu)^2}{2\sigma^4} \end{pmatrix},$$
$$\dot{l}_n(X_1, X_2, \ldots, X_n) = \begin{pmatrix} \displaystyle\sum_{m=1}^n \frac{(X_m - \mu)}{\sigma^2} \\ \displaystyle -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{m=1}^n \frac{(X_m-\mu)^2}{\sigma^4} \end{pmatrix}$$

$\dot{l}_n = 0$ liefert also die Lösung $\hat{\vartheta}_n(X_1, \ldots, X_n) = (\hat{\mu}_n, \hat{\sigma}_n^2)^T$, wobei

$$\hat{\mu}_n(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{m=1}^n X_m =: \bar{X}_n$$
$$\hat{\sigma}_n^2(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{m=1}^n (X_m - \bar{X}_n)^2$$

Dieses Modell ist regulär im oben genannten Sinne.

b) Verschobene Exponentialverteilung

Es seien $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen mit Dichte

$$f_\vartheta(x) = \mathbf{1}_{[\xi,\infty)}(x)\lambda \exp\{-\lambda(x-\xi)\}, \quad x \in \mathbb{R}\,.$$

Es sei $\vartheta = (\xi, \lambda)^T \in \mathbb{R} \times (0, \infty) =: \Theta$
(Skizzieren Sie die Dichte!)
Die Dichte $f_\vartheta(x)$ ist bei festem x nicht bezüglich ϑ differenzierbar. Bei festem x ist $f_\vartheta(x)$ für $\xi = x$ und $\lambda = \frac{1}{x-\xi}$ maximal. Folglich erhalten wir

$$L_n(\vartheta\,;\, X_1, \ldots, X_n) = \mathbf{1}_{[\xi,\,\infty)}\big(\min\{X_1, \ldots, X_n\}\big)\lambda^n \exp\big\{-\lambda \sum_{m=1}^{n} X_m + \lambda\,\xi\, n\big\}$$

$$\text{und somit } \hat{\vartheta}_n(X_1, \ldots, X_n) = \arg\max_{\vartheta\in\Theta} = L_n(\vartheta\,;\, X_1, \ldots, X_n)$$
$$= \begin{pmatrix} \min\{X_1, \ldots, X_n\} \\ \Big(\bar{X}_n - \min\{X_1, \ldots, X_n\}\Big)^{-1} \end{pmatrix},$$

also

$$\hat{\xi}_n = \min\{X_1, \ldots, X_n\} \quad \text{und} \quad \hat{\lambda}_n = \Big(\bar{X}_n - \min\{X_1, \ldots, X_n\}\Big)^{-1}.$$

Der Schätzer $\hat{\vartheta}_n$ ist in diesem Fall konsistent aber nicht asymptotisch normalverteilt. Die Regularitätsvoraussetzung H2) ist verletzt.

Ein einfacher Fall stochastischer Prozesse

1 Autoregressives Schema erster Ordnung:
Es sei $(\epsilon_n,\ n \geq 1)$ eine Folge reellwertiger, unabhängiger, identisch verteilter Zufallsgrößen, $X_0 = x_0$ und x_0, α seien reelle Zahlen. Wir definieren

$$X_n = \alpha X_{n-1} + \epsilon_n, \qquad n \geq 1.$$

Die Folge $(X_n\ n \geq 1)$ heißt *autoregressive Folge erster Ordnung* oder *AR(1)-Folge.*
ϵ_1 habe die Dichte f, die überall auf $\mathbb{R}$ positiv sei.
Dann besitzt auch die Stichprobe $X := (X_1, X_2 \ldots, X_n)$ eine Dichte

$$f^X(x_1, x_2, \ldots, x_n) = \prod_{m=1}^{n} f(x_m - \alpha x_{m-1}) = L_n(\alpha\,;\, x_1, x_2, \ldots, x_n)$$

und es gilt

$$l_n(\alpha\,;\,x_1,x_2,\ldots,x_n)=\sum_{m=1}^{n}\log f(x_m-\alpha x_{m-1}),\quad \dot{l}_n=-\sum_{m=1}^{n}x_{m-1}\frac{\dot{f}(x_m-\alpha x_{m-1})}{f(x_m-\alpha x_{m-1})}.$$

Die ML-Gleichung lautet:

$$\sum_{m=1}^{n}\frac{X_{m-1}\dot{f}(X_m-\hat{\alpha}_n X_{m-1})}{f(X_m-\hat{\alpha}_n X_{m-1})}=0.$$

Im Spezialfall $\epsilon_1\sim\mathcal{N}(0,1)$ gilt

$$l_n(\alpha\,;\,X_1,X_2,\ldots,X_n)=\alpha\sum_{m=1}^{n}X_mX_{m-1}-\frac{\alpha^2}{2}\sum_{m=1}^{n}X_{m-1}^2$$

$$\text{und indem man}\quad \dot{l}_n=\sum_{m=1}^{n}X_mX_{m-1}-\alpha\sum_{m=1}^{n}X_{m-1}^2=0\quad\text{setzt,}$$

bekommt man einen Maximum-Likelihood-Schätzer für α:

$$\hat{\alpha}_n=\frac{\sum_{m=1}^{n}X_mX_{m-1}}{\sum_{m=1}^{n}X_{m-1}^2}.$$

Es gilt

$$\hat{\alpha}_n-\alpha=\frac{\sum_{m=1}^{n}X_{m-1}(X_m-\alpha X_{m-1})}{\sum_{m=1}^{n}X_{m-1}^2}=\frac{\sum_{m=1}^{n}X_{m-1}\epsilon_m}{\sum_{m=1}^{n}X_{m-1}^2}.$$

Man beachte, daß

$$\Big(\sum_{m=1}^{n}X_{m-1}\epsilon_m\Big)_{n\geq 1}\ \text{ein Martingal ist und}\ \Big(\sum_{m=1}^{n}X_{m-1}^2\Big)_{n\geq 1}\ \text{seine bedingte Varianz darstellt:}$$

$$Var_{\wp_n}\big(\sum_{m=1}^{n}X_{m-1}\epsilon_m\big)=\mathbb{E}_\vartheta\Big(\big(\sum_{m=1}^{n}X_{m-1}\epsilon_m\big)^2\Big|\wp_{n-1}\Big)=\sum_{m=1}^{n}X_{m-1}^2.$$

$$\big(\wp_n:=\sigma(X_1,X_2,\ldots,X_n)=\sigma(\epsilon_1,\epsilon_2,\ldots,\epsilon_n),\quad n\geq 1\big)$$

Für die Untersuchung der asymptotischen Eigenschaften von $(\hat{\alpha}_n,\ n\geq 1)$ für $n\longrightarrow\infty$ bietet sich also die Martingaltheorie an.

Kapitel 4

Allgemeine Likelihoodtheorie

Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Modell mit dem Stichprobenraum $(E, \mathscr{E})$. Die Wahrscheinlichkeitsverteilung der Stichprobe X unter dem Wahrscheinlichkeitsmaß $\mathbb{P}_\vartheta$ ist gegeben durch

$$\mathbb{P}_\vartheta^X(B) := \mathbb{P}_\vartheta(X \in B), \quad B \in \mathscr{E}.$$

Im Fall, daß man stochastische Prozesse studiert, ist der Stichprobenraum E in der Regel ein Funktionenraum, z.B. der Raum $\mathbb{C}\big([0,T]\big)$ aller stetigen Funktionen auf dem Beobachtungsintervall $[0,T]$.

In diesem allgemeinen Fall gibt es keine ausgezeichnete „gleichmäßige Verteilung“ wie das Lebesguemaß im $\mathbb{R}^n$, bezüglich der man Dichten $f_\vartheta^X(x)$ bilden kann. Einen Ausweg bietet das Theorem von Radon-Nikodym aus der Maßtheorie, das in bestimmten Fällen die Existenz von Funktionen sichert, die die Rolle von Dichten im klassischen Fall übernehmen.

4.1 Das Theorem von Radon-Nikodym

Wir beginnen mit einem fundamentalen Satz über die Darstellung eines Maßes μ als Integral über eine Funktion f bezüglich eines anderen Maßes ν, dem *Satz von Radon-Nikodym.*

Definition 4.1. Es seien μ und ν zwei σ-finite Maße auf einem meßbaren Raum $(F, \mathscr{F})$. Man sagt, μ *sei absolutstetig bezüglich* ν, falls für jedes $A \in \mathscr{F}$ mit $\nu(A) = 0$ auch $\mu(A) = 0$ gilt.

Symbolisch schreibt man dafür $\mu \ll \nu$. Offenbar folgt aus $\mu \ll \nu$ und $\nu \ll \lambda$ auch $\mu \ll \lambda$ (Transitivität). Gilt sowohl $\mu \ll \nu$ als auch $\nu \ll \mu$, so heißen μ und ν *äquivalent*, im Zeichen: $\mu \equiv \nu$.

Beispiel 4.1. Es sei $(F, \mathscr{F})$ ein meßbarer Raum. Ist f eine $\mathscr{F}$-meßbare reellwertige, nichtnegative Funktion auf F, so wird durch

$$\mu(A) := \int_A f(x)\nu(\,dx), \quad A \in \mathscr{F}$$

ein σ-finites Maß μ auf $\mathscr{F}$ definiert mit $\mu \ll \nu$.

Aussage 4.1 (Satz von Radon-Nikodym). *Es seien μ und ν zwei σ-finite Maße auf $(F, \mathscr{F})$. Ist $\mu \ll \nu$, d.h., folgt für alle $A \in \mathscr{F}$ mit $\nu(A) = 0$ auch $\mu(A) = 0$, so existiert eine nichtnegative Funktion f auf F mit folgenden Eigenschaften:*

1. *f ist $\mathscr{F}$ -meßbar,*
2. *$\mu(A) = \int_A f(x)\nu(\,dx) \quad \forall A \in \mathscr{F}$.*

Die Funktion f ist ν-f.ü. eindeutig bestimmt, d.h., für jede $\mathscr{F}$-meßbare Funktion g mit $\int_A f\,d\nu = \int_A g\,d\nu$, $\forall A \in \mathscr{F}$, gilt $f = g$ ν-f.ü.

Beweis:
Siehe Bauer, H. (1992) Maß- und Integrationstheorie, de Gruyter-Verlag, Kapitel 17. □

Die Funktion f aus dem Satz von Radon-Nikodym heißt *Radon-Nikodym-Ableitung von μ nach ν* und wird mit $\frac{d\mu}{d\nu}$ bezeichnet. Damit kann man die Eigenschaft 2. in der Aussage 4.1 schreiben als:

$$\mu(A) = \int_A \frac{d\mu}{d\nu}(x)\,\nu(\,dx), \quad A \in \mathscr{F}. \tag{4.1}$$

Eigenschaften der Radon-Nikodym-Ableitung

Es gilt (x ist hier stets Element aus F)

$$\text{Falls } \mu \ll \nu, \quad \text{so gilt} \quad \frac{d\mu}{d\nu}(x) > 0 \quad \mu\text{-f.ü.} \tag{4.2}$$

$$\text{Falls } \mu \ll \nu \text{ und } \nu \ll \lambda, \quad \text{so haben wir} \quad \frac{d\mu}{d\lambda}(x) = \frac{d\mu}{d\nu}(x) \cdot \frac{d\nu}{d\lambda}(x) \quad \lambda\text{-f.ü.} \tag{4.3}$$

$$\text{Gilt } \nu \equiv \mu, \quad \text{so ist} \quad \frac{d\nu}{d\mu}(x) = \left(\frac{d\mu}{d\nu}(x)\right)^{-1} \quad \mu\text{- und } \nu\text{-f.ü.} \tag{4.4}$$

Beweis:

1. $\mu\left(\{x : \frac{d\mu}{d\nu}(x) = 0\}\right) = \int_{\{\frac{d\mu}{d\nu} = 0\}} \frac{d\mu}{d\nu}\, d\nu = 0$

2. Wegen $\nu(A) = \int_A \frac{d\nu}{d\lambda}(x)\lambda(\,dx), \quad A \in \mathscr{F},$ gilt für jede nichtnegative Funktion $g : \int_F g(x)\nu(\,dx) = \int_F g(x)\frac{d\nu}{d\lambda}(x)\lambda(\,dx).$

   (Man überlege sich die Gleichung für Indikatorfunktionen, für Linearkombinationen von Indikatorfunktionen und approximiere die erwähnten g durch monotone Folgen solcher Linearkombinationen. Danach wende man den Satz über monotone Konvergenz an.)
   Speziell für $g = \frac{d\mu}{d\nu}\mathbf{1}_A$ gilt: $\mu(A) = \int_A \frac{d\mu}{d\nu}(x)\nu(\,dx) = \int_A \frac{d\mu}{d\nu}(x)\frac{d\nu}{d\lambda}(x)\lambda(\,dx).$

   Die Behauptung folgt auf Grund der λ-fast sicheren Eindeutigkeit der Radon-Nikodym-Ableitung.

3. folgt aus 2. mit $\lambda = \mu :\quad \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\nu} = 1 \quad \mu$- und ν -f.ü. □

Bemerkung:
Wir nennen die Vorgehensweise im Punkt 2. die „Approximationsmethode“ und werden sie später noch mehrfach verwenden.

Zum Begriff „Absolutstetigkeit von Maßen und Funktionen“

Aussage 4.2. *Es gilt folgende Äquivalenz für je zwei endliche Maße μ und ν auf einem meßbaren Raum $(\Omega, \mathscr{A})$:*

$$\mu \ll \nu,\ \text{d.h. } \nu(A) = 0 \Longrightarrow \mu(A) = 0 \quad \forall A \in \mathscr{A}$$

ist äquivalent mit

$$\forall\, \epsilon > 0 : \exists\, \delta > 0 : \forall\, A \ \text{ mit } \ \mu(A) < \delta \ \text{ folgt } \ \mu(A) < \epsilon.$$

Beweis:
Hinlänglichkeit der Bedingung: Gilt $\nu(A) = 0$, so ist $\nu(A) < \delta$ für alle $\delta > 0$, also auch $\mu(A) < \epsilon$ für alle $\epsilon > 0$, somit gilt $\mu(A) = 0$.

Notwendigkeit: Angenommen, die Bedingung gilt nicht. Dann gibt es ein $\epsilon > 0$ und eine Folge $A_n \in \mathscr{A}$ mit $\nu(A_n) \leq 2^{-n}$ und $\mu(A_n) \geq \epsilon$. Wir setzen $A = \limsup_{n\to\infty} A_n$. Dann gilt $\nu(A) = 0$ (Borel-Cantelli). Andererseits ist $\mu(A_n) = \int_\Omega \mathbf{1}_{A_n}\, d\mu$ und

$$\mu(A) = \int_\Omega \limsup_{n\to\infty} \mathbf{1}_{A_n}\, d\mu \geq \limsup_{n\to\infty} \int_\Omega \mathbf{1}_{A_n}\, d\mu \geq \epsilon \qquad \text{(Fatousches Lemma)}$$

Widerspruch zur Annahme. □

Definition 4.2. Eine Verteilungsfunktion F auf $\mathbb{R}$ heißt *absolutstetig bezüglich dem Lebesguemaß* λ auf $\mathbb{R}$, falls das von ihr erzeugte Wahrscheinlichkeitsmaß $\mathbb{P}_F$ absolutstetig bezüglich dem Lebesguemaß λ ist.

In diesem Fall gibt es auf Grund des Radon-Nikodym-Theorems eine Borelmeßbare Funktion f mit

$$\mathbb{P}_F(A) = \int_A f\, d\lambda, \quad A \in \mathscr{B}, \quad \text{insbesondere gilt} \quad F(x) = \int_{(-\infty,x]} f(s)\, ds, \quad x \in \mathbb{R}\,.$$

Nicht jede stetige Funktion F ist absolutstetig (z.B. die Cantorsche Funktion).
Eine Verteilungsfunktion F ist absolutstetig genau dann, wenn für alle $\epsilon > 0$ ein $\delta > 0$ existiert, so daß für jede endliche Menge $\{[\alpha_k, \beta_k],\ k = 1, 2, \ldots, m\}$ paarweise disjunkter Teilintervalle mit $\sum_{k=1}^m (\beta_k - \alpha_k) < \delta$ die Ungleichung $\sum_{k=1}^m F(\beta_k) - F(\alpha_k) < \epsilon$ gilt.

Ist F absolutstetig, so ist F Lebesgue-fast überall differenzierbar, und es gilt

$$\frac{dF}{dx} = f(x) \qquad \lambda\text{ -f.ü.},$$

d.h. die Ableitung von F ist λ -f.ü. gleich der Radon-Nikodym-Ableitung von $\mathbb{P}_F$ nach λ. Außerdem gilt somit für jede absolutstetige Verteilungsfunktion F

$$F(x) = \int_{(-\infty,x]} \frac{dF}{ds}\, ds, \quad \text{d.h. } F \text{ hat eine Dichte } f(s) = \frac{dF}{ds}.$$

Literatur: Natanson, I.P., Theorie der Funktionen einer reellen Veränderlichen, Akademie-Verlag Berlin, 1961

Beispiel 4.2. Es seien F und G zwei absolutstetige Verteilungsfunktionen auf $\mathbb{R}$ mit den Dichten f bzw. g bezüglich dem Lebesguemaß λ. Es bezeichnen $\mathbb{P}_F$ und $\mathbb{P}_G$ die durch F bzw. G erzeugten Wahrscheinlichkeitsmaße auf $\mathscr{B}$, d.h. es gilt

$$\mathbb{P}_F\,(\,(a,b]\,) = F(b) - F(a) = \int_a^b f(s)\, ds \quad \text{und}$$

$$\mathbb{P}_G\,(\,(a,b]\,) = G(b) - G(a) = \int_a^b g(s)\, ds$$

für alle a, b mit $a < b$. Dann haben wir die

Aussage 4.3. *a)* $\mathbb{P}_F(B) = 0 \Longleftrightarrow \lambda\big(\{f > 0\} \cap B\big) = 0$ *(für alle Borelmengen B aus $\mathbb{R}$). Analoges gilt für $\mathbb{P}_G$ und g anstelle von $\mathbb{P}_F$ und f.*

b) $\mathbb{P}_F \ll \mathbb{P}_G \Longleftrightarrow \{x \in \mathbb{R} \mid f(x) > 0\} \overset{\lambda\text{-f.ü.}}{\subseteq} \{x \in \mathbb{R} \mid g(x) > 0\}$ *im Sinne von*

$$\lambda\big(\{f > 0\} \setminus \{g > 0\}\big) = 0. \tag{4.5}$$

c) In diesem Fall ist

$$\frac{d\,\mathbb{P}_F}{d\,\mathbb{P}_G}(x) = \frac{f(x)}{g(x)} \quad \textit{für } \mathbb{P}_G\textit{-fast alle } x \in \mathbb{R}\,. \tag{4.6}$$

Beweis:

a) Es gilt

$$\mathbb{P}_F(B) = 0 \Longleftrightarrow \int_B f(s)\,ds = 0 \Longleftrightarrow \int_{B\cap\{f>0\}} f(s)\,ds = 0 \Longleftrightarrow \lambda\big(B \cap \{f > 0\}\big) = 0.$$

b) Es gelte $\mathbb{P}_F \ll \mathbb{P}_G$. Dann gilt $\mathbb{P}_G\big(\{g = 0\}\big) = 0$ und folglich $\mathbb{P}_F\big(\{g = 0\}\big) = 0$, somit haben wir mit a)

$$\lambda\big(\{g = 0\} \cap \{f > 0\}\big) = 0, \quad \text{d.h.} \quad \lambda\big(\{f > 0\} \setminus \{g > 0\}\big) = 0.$$

Gilt dagegen (4.5) und ist $\mathbb{P}_G(B) = 0$, so haben wir (siehe a)) $\lambda\big(B \cap \{g > 0\}\big) = 0$ und somit wegen

$$\{f > 0\} \cap B = \Big(\{f > 0\} \cap \{g > 0\} \cap B\Big) \cup \Big(\{f > 0\} \cap \{g = 0\} \cap B\Big)$$

auch

$$\lambda\big(\{f > 0\} \cap B\big) \leq \lambda\big(\{g > 0\} \cap B\big) + \lambda\big(\{f > 0\} \setminus \{g > 0\}\big) = 0.$$

Das bedeutet $\mathbb{P}_F(B) = 0$.

c) Weiterhin gilt (mit $\frac{0}{0} := 0$)

$$\mathbb{P}_F(B) = \int_B f\,d\lambda = \int_{B\cap\{f>0\}} f\,d\lambda = \int_{B\cap\{f>0\}} \frac{f(s)}{g(s)} g(s)\,ds. \text{ (wegen } \{f > 0\} \subseteq \{g > 0\}\ \lambda\text{-f.ü.)}$$

Bekanntlich gilt für alle nichtnegativen Borelmeßbaren Funktionen h:

$$\int_{\mathbb{R}} h(x)\,\mathbb{P}_G(\,dx) = \int_{\mathbb{R}} h(x)g(x)\,dx$$

Folglich ist

$$\mathbb{P}_F(B) = \int_{B\cap\{f>0\}} \frac{f(s)}{g(s)}\,\mathbb{P}_G(\,ds) = \int_B \frac{f(s)}{g(s)}\,\mathbb{P}_G(\,ds),$$

$$\text{d.h.}\quad \frac{d\,\mathbb{P}_F}{d\,\mathbb{P}_G}(s) = \frac{f(s)}{g(s)} \qquad \mathbb{P}_G\text{-f.s. bzw. } \lambda\text{-f.ü. auf } \{g>0\}.$$

□

Beispiel 4.3. Es sei $\mathbb{P}$ eine diskrete Wahrscheinlichkeitsverteilung auf $E = \{e_1, e_2, \ldots, e_m, \ldots\}$ mit

$$\mathbb{P}\big(\{e_k\}\big) =: p_k \geq 0, \quad \sum_{k=1}^{\infty} p_k = 1, \quad \mathscr{E} = \wp(E).$$

Es sei weiterhin Q ein σ-finites Maß auf $\mathscr{E}$. Dann kann man die Absolutstetigkeit an Hand der Einzelwahrscheinlichkeiten prüfen:

Aussage 4.4. *Es gilt*

a) $\mathbb{P} \ll Q \Longleftrightarrow \{e_k\,:\,p_k>0\} \subseteq \{e_k\,:\,q_k>0\}$

b) $\dfrac{d\,\mathbb{P}}{dQ}(e_k) = \dfrac{p_k}{q_k}$ *für alle k mit $q_k > 0$.*
(Für e_k mit $q_k = 0$ kann $\frac{d\,\mathbb{P}}{dQ}(e_k)$ beliebig gewählt werden.)

Beweis:
Es gilt

$$\mathbb{P}\big(\{e_k\}\big) = \frac{p_k}{q_k} q_k; \quad P(B) = \sum_{k\,:\,e_k\in B} \frac{p_k}{q_k} q_k, \quad B \subseteq E.$$

□

Q bezeichne das *Zählmaß auf* E, d.h. es gelte

$$Q\big(\{e_k\}\big) = 1,\ \forall\, k, \quad Q(B) = \sum_{k\,:\,e_k\in B} \mathbf{1},\ B \subseteq E.$$

Damit ist $\mathbb{P} \ll Q$ $\quad$ ($Q(B) = 0$ impliziert $B = \emptyset$) und es gilt

$$\frac{d\,\mathbb{P}}{dQ}(e_k) = p_k, \quad k \geq 1.$$

4.2 Likelihood-Funktionen für dominierte statistische Räume

Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ mit $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ ein statistischer Raum.

Definition 4.3. Gilt für ein σ-finites Maß μ auf $(\Omega, \mathscr{A})$ die Beziehung $\mathbb{P}_\vartheta \ll \mu$ für alle $\vartheta \in \Theta$, so heißt die Familie $\mathscr{P}$ *durch das Maß* μ *dominiert* und μ ein *dominierendes Maß für* $\mathscr{P}$.

Bemerkung:
Auf der reellen Achse $\mathbb{R}$ oder im $\mathbb{R}^n$ ist häufig das Lebesguemaß ein dominierendes Maß für eine gegebene Familie $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ von Wahrscheinlichkeitsverteilungen, z.B. für die Familie aller eindimensionalen Normalverteilungen $\mathcal{N}(\mu, \sigma^2)$ mit $\vartheta = (\mu, \sigma^2)^T \in \mathbb{R} \times (0, \infty) =: \Theta$. Jede Familie $\mathbb{P}_\vartheta$, $\vartheta \in \Theta$ diskreter Verteilungen auf einer höchstens abzählbar unendlichen Menge $\{a_1, a_2, \ldots, a_m, \ldots\}$ ist dominiert durch das sogenannte „Zählmaß“, definiert durch $\mu(\{a_m\}) \equiv 1$.

Aussage 4.5. *Ist* μ *ein dominierendes Maß für* $\mathscr{P}$*, so gibt es ein zu* μ *äquivalentes Wahrscheinlichkeitsmaß* μ_0*, das* $\mathscr{P}$ *ebenfalls dominiert.*

Beweis:
Da μ σ-finit ist, gibt es nämlich eine Zerlegung von Ω in meßbare Mengen Z_k, $\quad k \geq 1$, mit $\mu(Z_k) \in (0, \infty)$. Wir setzen

$$\mu_0(A) = \sum_{k=1}^{\infty} 2^{-k} \, \frac{\mu(A \cap Z_k)}{\mu(Z_k)}, \qquad A \in \mathscr{A}.$$

$\square$

Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Experiment, dominiert durch ein σ-finites Maß μ. Für jedes $\vartheta \in \Theta$ definieren wir durch

$$L(\vartheta\,;\, \mathscr{A}) := \frac{d\,\mathbb{P}_\vartheta}{d\mu}$$

eine Zufallsgröße auf $(\Omega, \mathscr{A})$. Sie ist

1. meßbar bezüglich $\mathscr{A}$ und
2. es gilt $\mathbb{P}_\vartheta(A) = \displaystyle\int_A L(\vartheta\,;\, \mathscr{A})(\omega)\mu(\,d\omega)$ für alle $A \in \mathscr{A}$.

Definition 4.4. Als Funktion von ϑ heißt $\vartheta \longrightarrow L(\vartheta\,;\,\mathscr{A})(\omega)$ die *Likelihoodfunktion der Familie $\mathscr{P}$ bezüglich μ oder einfach von $\mathscr{P}$ bezüglich μ.*

Beispiel 4.4. Im Fall der Familie der Normalverteilungen $\mathcal{N}(\mu, \sigma^2)$, $(\mu \in \mathbb{R},\ \sigma^2 > 0)$ auf $\mathbb{R}$ erhält man für $\vartheta = (\mu, \sigma^2)^T$

$$L(\vartheta\,;\,\mathscr{B})(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\omega-\mu)^2}{2\sigma^2}\right\}, \quad \omega \in \mathbb{R}$$

für das dominierende Lebesguemaß λ.

Beispiel 4.5. Ist Ω eine abzählbare Menge, $\Omega = \{\omega_k : k \geq 1\}$ und $\{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ eine durch die Einzelwahrscheinlichkeiten $p_k(\vartheta) = \mathbb{P}_\vartheta(\{\omega_k\})$ definierte Wahrscheinlichkeitsverteilung auf $(\Omega, \wp(\Omega))$, so ist wie bereits erwähnt, das Zählmaß μ, definiert durch $\mu(\{\omega_k\}) \equiv 1$, ein die Familie $\mathscr{P} = \{\mathbb{P}_\vartheta : \vartheta \in \Theta\}$ dominierendes Maß. Es gilt

$$L\big(\vartheta\,;\,\wp(\Omega)\big)(\omega_k) = \frac{p_k(\vartheta)}{\mu(\{\omega_k\})} = p_k(\vartheta), \quad k \geq 1.$$

4.3 Stochastische Likelihoodfunktionen

Definition 4.5. Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistischer Raum mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und es sei $\mathscr{H}$ eine Teil-σ-Algebra von $\mathscr{A}$. Es gelte $\mathbb{P}_\vartheta\big|_{\mathscr{H}} \ll \mu$ für ein σ-finites Maß μ auf $\mathscr{H}$.
($\mathbb{P}_\vartheta\big|_{\mathscr{H}}$ bezeichnet die Einschränkung von $\mathbb{P}_\vartheta$ auf die Teil-σ-Algebra $\mathscr{H}$ von $\mathscr{A}$.)
Dann heißt

$$L(\vartheta\,;\,\mathscr{H})(\omega) := \frac{d\,\mathbb{P}_\vartheta\big|_{\mathscr{H}}}{d\mu}(\omega),$$

die *stochastische Likelihoodfunktion von $\mathscr{P}$ bezüglich μ und $\mathscr{H}$.*
Ist $\mathscr{H} = \mathscr{A}$, so handelt es sich um die in Abschnitt 4.2 definierte Likelihoodfunktion von $\mathscr{P}$ bezüglich μ.

Nach Definition ist $\vartheta \longrightarrow L(\vartheta\,;\,\mathscr{H})$ eine Funktion, deren Werte Zufallsgrößen über $(\Omega, \mathscr{A})$ sind, mit folgenden zwei Eigenschaften:

1. Für jedes $\vartheta \in \Theta$ ist $L(\vartheta\,;\,\mathscr{H})$ ist eine $\mathscr{H}$-meßbare Zufallsgröße,
2. $\int_H L(\vartheta\,;\,\mathscr{H})\,d\mu = \mathbb{P}_\vartheta(H) \quad \forall H \in \mathscr{H},\ \forall \vartheta \in \Theta$

Die Eigenschaften 1. und 2. bestimmen $L(\vartheta\,;\,\mathscr{H})$ eindeutig in dem Sinne, daß für jede $\mathscr{H}$-meßbare Funktion $L_1(\vartheta\,;\,\mathscr{H})$ auf Ω mit der Eigenschaft 2, $L(\vartheta, \cdot) = L_1(\vartheta, \cdot)$ μ-f.ü., für alle $\vartheta \in \Theta$ gilt.

Aussage 4.6. *Ist $\mathscr{H}'$ eine weitere σ-Algebra mit $\mathscr{H} \subseteq \mathscr{H}' \subseteq \mathscr{A}$, und ist $(\mathbb{P}_\vartheta \big|_{\mathscr{H}'},\ \vartheta \in \Theta)$ dominiert durch ein Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $\mathscr{H}'$, so ist (natürlich) auch $(\mathbb{P}_\vartheta \big|_{\mathscr{H}},\ \vartheta \in \Theta)$ dominiert durch $\mathbb{P}$, und es gilt*

$$L(\vartheta\,;\,\mathscr{H}) = \mathbb{E}_{\mathbb{P}}\big(L(\vartheta\,;\,\mathscr{H}')|\mathscr{H}\big). \tag{4.7}$$

Beweis:
Nach Definition ist $L(\vartheta\,;\,\mathscr{H})$ eine $\mathscr{H}$-meßbare Funktion mit

$$\int_H L(\vartheta\,;\,\mathscr{H})\,d\,\mathbb{P} = \mathbb{P}_\vartheta(H), \quad H \in \mathscr{H},\ \vartheta \in \Theta,$$

und da $\mathscr{H} \subseteq \mathscr{H}'$ gilt, ist auch

$$\mathbb{P}_\vartheta(H) = \int_H L(\vartheta\,;\,\mathscr{H}')\,d\,\mathbb{P}, \quad H \in \mathscr{H},\ \vartheta \in \Theta.$$

Somit gilt

$$\int_H L(\vartheta\,;\,\mathscr{H})\,d\,\mathbb{P} = \int_H L(\vartheta\,;\,\mathscr{H}')\,d\,\mathbb{P}, \quad H \in \mathscr{H},\ \vartheta \in \Theta.$$

Daraus ergibt sich die Behauptung auf Grund der Definition der bedingten Erwartung. □

Folgerung 4.6 *(Martingaleigenschaft der Likelihoodfunktion).* Ist $(\mathscr{H}_n,\ n \geq 1)$ eine wachsende Folge von Teil-σ-Algebren von $\mathscr{A}$, d.h. gilt $\mathscr{H}_1 \subseteq \mathscr{H}_2 \subseteq \ldots \subseteq \mathscr{H}_n \subseteq \mathscr{A}, \quad (n \geq 1)$, so ist für jedes $\vartheta \in \Theta$ die Folge $\big(L(\vartheta\,;\,\mathscr{H}_n),\ \mathscr{H}_n\big)_{n\geq 1}$ ein nichtnegatives Martingal bezüglich $\mathbb{P}$.

Auf Grund des Martingalkonvergenzsatzes (siehe z.B. Shiryaev) folgt: Der Grenzwert $\lim\limits_{n\to\infty} L_n(\omega) =: L_\infty$ existiert $\mathbb{P}$-fast sicher und es gilt $\mathbb{E}_{\mathbb{P}}(L_\infty) \leq \overline{\lim}_{n\to\infty} \mathbb{E}_{\mathbb{P}}(L_n) = 1$. (Fatou), L_∞ ist $\mathscr{H}_\infty := \bigvee_{n=1}^\infty \mathscr{H}_n$ meßbar. Hier wurde $L_n = L(\vartheta\,;\,\mathscr{H}_n)$ gesetzt.

Frage: Sind die Wahrscheinlichkeitsmaße $\mathbb{P}_\vartheta \big|_{\mathscr{H}_\infty}$ auch absolutstetig bezüglich $\mathbb{P} \big|_{\mathscr{H}_\infty}$?
Dieser Frage werden wir später (im Abschnitt 4.7) nachgehen.

4.4 Deterministische Likelihoodfunktionen

Im allgemeinen beobachtet man nicht den Ausgang ω des zufälligen Experiments $(\Omega, \mathscr{A}, \mathscr{P})$ als Ganzes, sondern nur Teile davon, z.B. Temperaturmessungen oder Aktienkurse, zu diskreter Zeit (d.h. täglich, wöchentlich, monatlich), im Gegensatz zu kontinuierlicher, also zeitstetiger Beobachtung.

Dieser Sachverhalt wird durch eine Stichprobe $X(\omega)$ modelliert: man beobachtet $X(\omega)$ anstelle von ω.
Folglich hat man nach der Ausführung des Experimentes auch nicht die volle Information, welche der Ereignisse $A \in \mathscr{A}$ eingetreten sind und welche nicht, sondern man weiß es nur von Ereignissen der Form $\{X \in B\} = X^{-1}(B)$ mit $B \in \mathscr{E}$. Da man X beobachtet, kann man entscheiden, ob der beobachtete Wert zu B gehört, d.h., ob $\{X \in B\}$ eingetreten ist.
Die relevante σ-Algebra von Ereignissen ist $\mathscr{A}^X := X^{-1}(\mathscr{E}) \subseteq \mathscr{A}$.
Wir können also auf der Grundlage unserer beobachteten Stichprobe X i.a. die Likelihoodfunktion $L(\vartheta\,;\mathscr{A})(\omega)$ nicht bestimmen. Der Ausweg liegt in der Einführung der Likelihoodfunktion $L(\vartheta\,;\mathscr{A}^X)$. Da $L(\vartheta\,;\mathscr{A}^X)$ eine $\mathscr{A}^X$-meßbare Zufallsgröße ist, muß es für jedes $\vartheta \in \Theta$ eine meßbare Funktion Ψ_ϑ von $(E, \mathscr{E})$ in $(R, \mathscr{B})$ geben mit

$$L(\vartheta\,;\mathscr{A}^X)(\omega) = \Psi_\vartheta\big(X(\omega)\big), \quad \mathbb{P}_\vartheta\text{-f.s.}$$

(Siehe z.B. Bauer, Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie)

Diese Funktion $\Psi_\vartheta(x)$ werden wir hier im Folgenden berechnen.

<u>Vorbereitungen:</u>

a) Die von X erzeugte σ-Algebra $\mathscr{A}^X$
Es seien $(\Omega, \mathscr{A})$ und $(E, \mathscr{E})$ meßbare Räume und X eine meßbare Abbildung von $(\Omega, \mathscr{A})$ in $(E, \mathscr{E})$, es gelte also $\mathscr{A}^X := X^{-1}(\mathscr{E}) \subseteq \mathscr{A}$.
Das Mengensystem $\mathscr{A}^X = X^{-1}(\mathscr{E})$ ist eine σ-Algebra von Teilmengen von Ω, und zwar die kleinste σ-Algebra $\mathscr{H}$, bezüglich der X $(\mathscr{H}\text{-}\mathscr{E})$-meßbar ist. Sie heißt *die von X in Ω erzeugte σ-Algebra* und wird auch mit $\sigma(X)$ bezeichnet.
Die Elemente von $\mathscr{A}^X$ sind genau die Urbilder $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}, \quad B \in \mathscr{E}$, mit anderen Worten,

$$A \in \mathscr{A}^X \Longleftrightarrow \exists\, B \in \mathscr{E} \,:\, A = X^{-1}(B). \tag{4.8}$$

b) Die von X erzeugte Familie $\mathscr{P}^X$
Es sei ν ein σ-finites Maß auf $\mathscr{A}^X$. Dann ist durch $\nu^X(B) := \nu(A)$ mit $A := X^{-1}(B) \in \mathscr{A}^X$ ein σ-finites Maß ν^X auf $\mathscr{E}$ definiert.

Es seien nun $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Modell mit dem Stichprobenraum $(E, \mathscr{E})$

und ν ein σ-finites Maß auf $\mathscr{A}$. Wir setzen $\mathscr{P}\big|_{\mathscr{A}^X} := (\mathbb{P}_\vartheta\big|_{\mathscr{A}^X} : \vartheta \in \Theta)$ und $\mathscr{P}^X := (\mathbb{P}_\vartheta^X : \vartheta \in \Theta)$. Hierbei bezeichnet $Q\big|_{\mathscr{A}^X}$ für jedes Maß Q auf $\mathscr{A}$ die Einschränkung von Q auf $\mathscr{A}^X$.

Lemma 4.7. *Genau dann dominiert $\nu\big|_{\mathscr{A}^X}$ die Familie $\mathscr{P}\big|_{\mathscr{A}^X}$, wenn ν^X die Familie $\mathscr{P}^X$ dominiert.*

Beweis:
Es sei $A \in \mathscr{A}^X$ und $B \in \mathscr{E}$ derart, daß $A = X^{-1}(B)$ gilt. Angenommen, $\nu\big|_{\mathscr{A}^X}$ dominiert $\mathscr{P}\big|_{\mathscr{A}^X}$ und es sei $\nu^X(B) = 0$.
Dann folgt

$$\nu(A) = \nu(X^{-1}(B)) = \nu^X(B) = 0, \text{ folglich } \mathbb{P}_\vartheta(A) = 0 \text{ für alle } \vartheta \in \Theta.$$

Somit gilt $\mathbb{P}_\vartheta^X(B) = 0,\ \ \vartheta \in \Theta$. Also dominiert ν^X die Familie $\mathscr{P}^X$. Die Umkehrung zeigt man analog. □

Nehmen wir an, daß ν^X die Familie $\mathscr{P}^X$ dominiert, so haben wir insbesondere die Likelihoodfunktion

$$L^X(\vartheta\,;\,x) := \frac{d\,\mathbb{P}_\vartheta^X}{d\nu^X}(x), \quad x \in E,\ \vartheta \in \Theta$$

von $\mathscr{P}^X$ bezüglich ν^X. Sie ist nach Definition $\mathscr{E}$-meßbar. Da die Elemente $x \in E$ als konkrete Stichproben angesehen werden, die nichtzufällig sind, bezeichnen wir $L^X(\vartheta\,;\,x)$ als *deterministische Likelihoodfunktion von $\mathscr{P}^X$ bezüglich ν^X*. Setzt man in $L^X(\vartheta\,;\,x)$ an die Stelle x die mathematische Stichprobe X ein, so erhält man eine Zufallsgröße $L^X\big(\vartheta\,;\,X(\omega)\big)$, für die folgende Aussage gilt:

Aussage 4.7. *Es gilt für jedes $\vartheta \in \Theta$*

$$L\big(\vartheta\,;\,\mathscr{A}^X\big)(\omega) = L^X\big(\vartheta\,;\,X(\omega)\big), \quad \textit{für } \nu\textit{- fast alle } \omega. \tag{4.9}$$

Beweis:
Nach Definition ist für jedes $B \in \mathscr{E}$

$$\nu^X(B) := \nu\left(X^{-1}(B)\right), \qquad \text{also} \quad \int_\Omega \mathbf{1}_B\big(X(\omega)\big)\nu(\,d\omega) = \int_E \mathbf{1}_B(x)\nu^X(\,dx).$$

Damit gilt für alle $\mathscr{E}$-meßbaren nichtnegativen Funktionen g (Beweis mit der Approximationsmethode aus Kapitel 4.1),

$$\int_\Omega g\big(X(\omega)\big)\nu(\,d\omega) = \int_E g(x)\nu^X(\,dx) \text{ (Substitutionsformel).}$$

Wir setzen $g(x) := L^X(\vartheta\,;\,x)\mathbf{1}_B(x)$ und erhalten für jedes $B \in \mathscr{E}$ die Gleichung

$$\int_{X^{-1}(B)} L^X\big(\vartheta, X(\omega)\big)\nu(\,d\omega) = \int_B L^X(\vartheta\,;\,x)\nu^X(\,dx) = \mathbb{P}_\vartheta^X(B) = \mathbb{P}_\vartheta\left(X^{-1}(B)\right).$$

Da $\mathscr{A}^X = \{X^{-1}(B) \,|\, B \in \mathscr{E}\}$ gilt, ist die Aussage 4.7 bewiesen. Dabei haben wir benutzt, daß $\omega \longrightarrow L^X\big(\vartheta\,;\,X(\omega)\big)$ $\mathscr{A}^X$-meßbar ist. □

Folgerung 4.8. Ist Y eine Zufallsgröße über $(\Omega, \mathscr{A})$ mit Werten in $(F, \mathscr{F})$, und ist $Y = \Psi(X)$ für eine meßbare Abbildung Ψ von $(E, \mathscr{E})$ in $(F, \mathscr{F})$, so gilt $\mathscr{A}^Y \subseteq \mathscr{A}^X$ (wegen $\mathscr{A}^Y = Y^{-1}(\mathscr{F}) = X^{-1}\big(\Psi^{-1}(\mathscr{F})\big) \subseteq X^{-1}(\mathscr{E}) = \mathscr{A}^X$)
Folglich gilt (siehe Aussage 4.6):

$$\begin{aligned} L(\vartheta\,;\,\mathscr{A}^Y) &= \mathbb{E}_{\mathbb{P}}\big(L(\vartheta\,;\,\mathscr{A}^X) \mid \mathscr{A}^Y\big), \text{ oder anders ausgedrückt,} \\ L^Y(\vartheta\,;\,Y(\omega)) &= \mathbb{E}_{\mathbb{P}}\big(L^X(\vartheta\,;\,X) \mid Y\big)(\omega) \quad \mathbb{P}\text{-f.s.} \end{aligned} \tag{4.10}$$

Diese sehr allgemeinen Folgerungen aus dem Satz von Radon-Nikodym bilden den allgemeinen Rahmen für die Likelihoodtheorie stochastischer Prozesse mit diskreter oder auch mit stetiger Zeit.

4.5 Ausflug in die Welt der Stochastischen Prozesse

Unser statistisches Modell umfaßt insbesondere auch Fälle, bei denen die Stichprobe X aus Trajektorien zeitstetiger stochastischer Prozesse besteht. Der Stichprobenraum E ist also ein Funktionenraum. Die Wahl einer geeigneten σ-Algebra $\mathscr{E}$ von Teilmengen von E erfolgt in Analogie des Falles $E = \mathbb{R}^n$. Wir erinnern uns daran, daß $\mathscr{B}^n$ die kleinste σ-Algebra ist, die alle Rechtecke $B_1 \times \ldots \times B_n$ mit $B_k \in \mathscr{B},\ k = 1, \ldots, n$, enthält. Die Rechtecke werden im Funktionenraum durch die sogenannten Zylindermengen ersetzt. Doch zunächst einige Vorbereitungen.

Es seien $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X := \big(X(t),\, t \in [0,T]\big)$ ein reellwertiger stochastischer Prozeß über $(\Omega, \mathscr{A}, \mathbb{P})$. Für $\mathbb{P}$-fast alle $\omega \in \Omega$ seien die Trajektorien $t \longrightarrow X(t,\,\omega)$ von X an jeder Stelle $t \in [0,T)$ rechtsstetig und für jedes $t \in (0,T]$ existiere der Grenzwert von links und sei endlich. (Praktisch alle gegenwärtig in Anwendungen und in der Theorie vorkommenden stochastischen Prozesse haben diese Eigenschaft.)
Der Vektorraum $E_T := \mathbb{D}\big([0,T]\big)$ aller reellwertigen Funktionen $x = (x_s,\ s \leq T)$ mit den genannten Eigenschaften (cadlag = continues a droite, limites a gauche) ist metrisierbar mit einer Metrik d zu einem vollständigen metrischen Raum (E_T, d). Mit $\mathscr{E}_T$ bezeichnen wir die σ-Algebra der Borelmengen von (E_T, d).

Aussage 4.8. *$\mathscr{E}_T$ ist identisch mit der kleinsten σ-Algebra $\sigma(x_s,\ s \in [0,T])$, bezüglich der alle Abbildungen $x \longrightarrow x_s,\ x \in E_T,\ s \in [0,T]$, Borelmeßbar sind.*

Beweis: (Siehe Billigsley, Convergence of Probability Measures)

Für die komplizierte und nicht unmittelbar zugängliche σ-Algebra $\mathscr{E}_T$ gibt es ein Erzeugendensystem $\mathscr{Z}$, das aus speziellen Teilmengen von E_T, den sogenannten Zylindermengen, besteht.

Definition 4.9. Eine Teilmenge Z von E_T heißt eine *Zylindermenge*, falls es ein $m \geq 1$, Zeitpunkte $t_0, t_1, \ldots, t_m$ mit $0 = t_0 < t_1 < \ldots < t_m = T$ und ein $B \in \mathscr{B}^{m+1}$ gibt, mit

$$Z = \{x = (x_s) \in E_T \,|\, (x_{t_0}, x_{t_1}, \ldots, x_{t_m}) \in B\}$$

Mitunter werden wir auch die Notation $Z = Z_{t_1,t_2,\ldots,t_m\,;\,B}$ verwenden; mit $\mathscr{Z}$ werde die Menge aller Zylindermengen aus E_T bezeichnet.

Lemma 4.10. *$\mathscr{Z}$ ist eine Semialgebra von Teilmengen aus E_T.*

Beweis: Übungsaufgabe

Wegen der Aussage 4.8 gilt $\{x_s \in B\} \in \mathscr{E}$ für alle $B \in \mathscr{B}$ und $s \in [0,T]$. Daraus ergibt sich (ohne Beweis) $\mathscr{Z} \subseteq \mathscr{E}_T$ und somit auch $\sigma(\mathscr{Z}) \subseteq \mathscr{E}_T$. (Mit $\sigma(\mathscr{Z})$ bezeichnen wir die kleinste σ-Algebra von Teilmengen von E_T, in der $\mathscr{Z}$ enthalten ist.) Mittels der obigen Aussage 4.8 erhalten wir $\sigma(\mathscr{Z}) = \mathscr{E}_T$.
Um die Wahrscheinlichkeitsverteilung $\mathbb{P}^X$ von X auf $\mathscr{E}_T$ zu bestimmen, genügt es, sie auf $\mathscr{Z}$ anzugeben, da jede σ-additive σ-finite Mengenfunktion Q auf $\mathscr{Z}$ auf eindeutige Weise von $\mathscr{Z}$ auf $\sigma(\mathscr{Z})$ fortgesetzt werden kann.
Für $Z \in \mathscr{Z}$ gilt offenbar

$$\mathbb{P}^X(Z) := \mathbb{P}(X \in Z) = \mathbb{P}\big((X(t_0), \ldots, X(t_m)) \in B\big),$$

wobei Z von der Gestalt

$$Z = \{x = (x_s) \,|\, (x_{t_0}, \ldots, x_{t_m}) \in B\} \text{ für ein } B \in \mathscr{B}^{m+1}$$

sei.
Wir sehen also, daß die Verteilung $\mathbb{P}^X$ von X auf $\mathscr{E}_T$ eindeutig bestimmt ist durch die Wahrscheinlichkeitsverteilungen aller möglichen endlichdimensionalen Vektoren $\big(X(t_0), \ldots, X(t_m)\big)$ mit $\{t_0, t_1, \ldots, t_m\} \subseteq [0,T]$, wobei o.B.d.A. $t_0 = 0$ und $t_m = T$ gesetzt werden und $0 < t_1 < \ldots < t_m$ gefordert werden kann.

Definition 4.11. Für jede Auswahl $\mathscr{T}_n := \{t_0, t_1, \ldots, t_m\} \subseteq [0, T]$ mit den eben angegebenen Eigenschaften, ist durch

$$\Phi_{\mathscr{T}_n}(B) := \mathbb{P}\big((X(t_0), \ldots, X(t_m)) \in B\big), \quad B \in \mathscr{B}^{m+1}$$

eine Wahrscheinlichkeitsverteilung $\Phi_{\mathscr{T}_n}$ auf $\mathscr{B}^{m+1}$ definiert, die zu $\mathscr{T}_n$ gehörende sogenannte *endlichdimensionale Verteilung von* X.

Die endlichdimensionalen Verteilungen stochastischer Prozesse lassen sich in vielen Fällen explizit bestimmen.

Beispiel 4.6. Poissonscher Prozess $\big(N(t),\ t \geq 0\big)$ mit Intensitätsparameter $\vartheta > 0$.

Definition 4.12. Ein stochastischer Prozess $N = \big(N(t),\ t \geq 0\big)$ auf einem Grundraum $(\Omega, \mathscr{F}, \mathbb{P})$ heißt *Poissonscher Prozess mit Intensitätsparameter* $\vartheta > 0$, falls gilt:

1. $N(0) = 0 \quad \mathbb{P}$-f.s.,
2. für $0 = t_0 < t_1 < \ldots < t_m = T$ sind $\big(N(t_k) - N(t_{k-1})\big),\ k = 1, 2, \ldots, m$ unabhängig,
3. für $0 \leq s < t$ ist $N(t) - N(s)$ Poissonverteilt zum Parameter $\vartheta(t - s)$,
4. Jede Funktion $t \longrightarrow N(t, \omega)$ ist stückweise konstant, ihre Sprungpunkte $T_k,\ k \geq 1$, häufen sich nicht im Endlichen und es gilt

   $$N(T_k) = N(T_k + 0) = N(T_k - 0) + 1.$$

   (Aus 1. - 3. folgt bereits, daß es eine Version von N gibt, die die Eigenschaft 4. besitzt.)

Es sei $0 = t_0 < t_1 < t_2 < \ldots < t_n = T$. Dann ist $\Phi^{N,\vartheta}_{n;t_1,t_2,\ldots,t_n}$ eine diskrete Verteilung auf N^n mit $N = \{0, 1, 2, \ldots\}$ und ihre Einzelwahrscheinlichkeiten sind

$$\begin{aligned}
&\Phi^{N,\vartheta}_{n;t_1,t_2,\ldots,t_n}\Big((i_1, i_2, \ldots, i_n)\Big) = \mathbb{P}_\vartheta\big(N(t_1) = i_1, \ldots, N(t_n) = i_n\big) \\
&= \frac{(\vartheta t_1)^{i_1}}{i_1!} e^{-\lambda t_1} \frac{(\vartheta(t_2 - t_1))^{i_2 - i_1}}{(i_2 - i_1)!} e^{-\lambda(t_2 - t_1)} \cdot \ldots \cdot \frac{(\vartheta(t_n - t_{n-1}))^{i_n - i_{n-1}}}{(i_n - i_{n-1})!} e^{-\lambda(t_n - t_{n-1})} \\
&= \vartheta^{i_n} e^{-\vartheta t_n} \nu\big((i_1, \ldots, i_n)\big) \qquad (4.11)
\end{aligned}$$

mit

$$\nu\big((i_1, \ldots, i_n)\big) = \prod_{k=1}^{n} \frac{((t_k - t_{k-1}))^{i_k - i_{k-1}}}{(i_k - i_{k-1})!}.$$

Beispiel 4.7. Wienerprozess $(W(t),\ t \geq 0)$ mit Drift μ und Diffusion $\sigma^2 > 0$.

Definition 4.13. Ein stochastischer Prozess $W = (W(t),\ t \geq 0)$ auf einem Grundraum $(\Omega, \mathscr{F}, \mathbb{P})$ heißt *Wienerprozess mit den Parametern* μ, σ^2, wenn gilt:

1. $W(0) = 0 \quad \mathbb{P}$-f.s.,
2. für $0 = t_0 < t_1 < \ldots < t_m = T$ sind $\big(W(t_k) - W(t_{k-1})\big),\ k = 1, 2, \ldots, m$ unabhängig,
3. für $0 \leq s < t$ ist $W(t) - W(s) \sim \mathcal{N}\big(\mu(t-s),\ \sigma^2(t-s)\big)$-verteilt,
4. $t \longrightarrow W(t, \omega)$ ist stetig für $\mathbb{P}$-fast alle $\omega \in \Omega$.
   (Aus 1. - 3. folgt bereits, daß es eine Version von W gibt, die die Eigenschaft 4. besitzt.)

μ und σ^2 heißen *Drift* bzw. *Diffusion* von W.
Einen Wienerprozess W mit Drift $\mu = 0$ und Diffusion $\sigma^2 = 1$ nennen wir *Standard-Wienerprozess* und bezeichnen ihn mit W^0.

Bemerkung: Ist W^0 ein Standard-Wienerprozess, so bildet

$$W(t) = \mu t + \sigma W^0(t), \quad t \geq 0$$

einen Wienerprozess mit Drift μ und Diffusion σ^2.

Es sei $W = (W(t),\, t \geq 0)$ ein Wiener-Prozess mit dem Parameter $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty) =: \Theta$. Es sei $0 = t_0 < t_1 < \ldots < t_m = T$. Wir setzen

$$X := (W(t_1), \ldots, W(t_m)), \quad E = \mathbb{R}^m,\ \mathscr{E} = \mathscr{B}^m.$$

Für die Wahrscheinlichkeitsdichte von X gilt

Aussage 4.9. *X hat die Dichte bezüglich des Lebesguemaßes*

$$\varphi^{\vartheta}_{t_1,\ldots,t_m}(x_1, \ldots, x_m) = \prod_{k=1}^{m} \varphi^{\vartheta}_{t_k - t_{k-1}}(x_k - x_{k-1}), \tag{4.12}$$

$$\textit{wobei} \quad \varphi^{\vartheta}_t(x) := (2\pi\sigma^2 t)^{-\frac{1}{2}} \exp\Big\{ -\frac{(x-\mu t)}{2\sigma^2 t} \Big\}, \quad x \in \mathbb{R},\ t > 0 \quad \textit{gesetzt wurde.}$$

Beweis:
Es sei $Z := (W(t_1) - W(t_0), \ldots, W(t_m) - W(t_{m-1}))$ mit $t_0 = 0$. Z hat nach Definition des Wiener-Prozesses die Dichte

$$f_Z(z_1, z_2, \ldots, z_m) = \prod_{k=1}^{m} \varphi^{\vartheta}_{t_k - t_{k-1}}(z_k).$$

Es gilt weiter $X = A \cdot Z$ mit der Matrix $A = (e_{ij})_{i,j=1,\ldots,m}$, $e_{ij} = 1$ für $i \geq j$, $e_{ij} = 0$ für $i < j$. Folglich ist für $x = (x_1, \ldots, x_m)$, $\varphi_{t_1,\ldots,t_m}(x_1, \ldots, x_m) = f_Z(A^{-1}x) \cdot |\det A^{-1}|$, woraus sich wegen

$$A^{-1} = (f_{ij})_{i,j=1,\ldots,n}, \quad f_{ij} = \begin{cases} 1 & \text{für } {\scriptstyle i=j} \\ -1 & \text{für } {\scriptstyle i=j+1} \\ 0 & \text{sonst} \end{cases}$$

die Behauptung ergibt. □

4.6 Likelihood am Beispiel des Wienerprozesses

Es seien $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$, $\Theta = \mathbb{R} \times (0, \infty)$, $\vartheta = (\mu, \sigma^2)^T \in \Theta$ und $W = (W(t),\ t \geq 0)$ sei ein reellwertiger stochastischer Prozess auf $(\Omega, \mathscr{A})$, der für jedes $\vartheta = (\mu, \sigma^2)^T \in \Theta$ bezüglich $\mathbb{P}_\vartheta$ ein Wienerprozess mit Drift μ und Diffusion σ^2 ist. Wir nehmen an, μ und σ^2 seien unbekannt.

Ziel ist es, auf der Grundlage der Beobachtung $X = \big(W(t),\ t \in [0, T]\big)$ des Wienerprozesses bis zur Zeit $T < \infty$ Schätzungen für μ und σ^2 vorzunehmen. Wir konzentrieren uns hier auf die Likelihoodmethode.

Beginnen wir mit dem Fall diskreter Beobachtung, d.h., wir kennen W nur zu den Zeitpunkten $t_1 < \ldots < t_m = T$. Die Menge $\mathscr{T}$ sei die Menge dieser Zeitpunkte: $\mathscr{T} = \{t_1, t_2, \ldots, t_m\}$. Als Stichprobe liegt uns dann der Vektor $X = \big(W(t_1), \ldots, W(t_m)\big)$ vor.
Die Verteilung $\mathbb{P}^X_\vartheta$ von X ist eine Normalverteilung mit der Dichte bezüglich des Lebesguemaßes λ_m (vgl. (4.12))

$$\varphi^{\vartheta}_{\mathscr{T}}(x_1, \ldots, x_m) = \prod_{k=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2(t_k - t_{k-1})}} \exp\left\{ -\frac{\big(x_k - x_{k-1} - \mu(t_k - t_{k-1})\big)^2}{2(t_k - t_{k-1})\sigma^2} \right\}, \tag{4.13}$$

$x = (x_1, \ldots, x_m) \in \mathbb{R}^m$.

Da $\varphi^{\vartheta}_{\mathscr{T}}$ eine streng positive Dichte für $\mathbb{P}^X_\vartheta$ ist, sind $\mathbb{P}^X_\vartheta$ und das Lebesguemaß λ_m auf $\mathbb{R}^m$ äquivalente Maße, für jedes $\vartheta \in \Theta$.

Somit sind für je zwei $\vartheta, \vartheta_0 \in \Theta$ die Wahrscheinlichkeitsmaße $\mathbb{P}^X_\vartheta$ und $\mathbb{P}^X_{\vartheta_0}$ äquivalent. Daraus ergibt sich (siehe (4.3) und (4.4)) :

$$\frac{d\,\mathbb{P}^X_\vartheta}{d\,\mathbb{P}^X_{\vartheta_0}} = \frac{d\,\mathbb{P}^X_\vartheta}{d\lambda} \cdot \frac{d\lambda}{d\,\mathbb{P}^X_{\vartheta_0}} = \frac{d\,\mathbb{P}^X_\vartheta}{d\lambda} \cdot \left(\frac{d\,\mathbb{P}^X_{\vartheta_0}}{d\lambda}\right)^{-1}.$$

Das bedeutet für die deterministische Likelihoodfunktion

$$L^X(\vartheta\,;\,x) = \frac{d\,\mathbb{P}^X_\vartheta}{d\,\mathbb{P}^X_{\vartheta_0}}(x) = \frac{\varphi^{\vartheta}_{\mathscr{T}}(x)}{\varphi^{\vartheta_0}_{\mathscr{T}}(x)} =$$
$$= \left(\frac{\sigma^2}{\sigma_0^2}\right)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\left(\sigma^{-2}-\sigma_0^{-2}\right)\sum_{k=1}^{m}\frac{(x_k-x_{k-1})^2}{t_k-t_{k-1}} + \left(\frac{\mu}{\sigma^2}-\frac{\mu_0}{\sigma_0^2}\right)x_m - \frac{1}{2}\left(\frac{\mu^2}{\sigma^2}-\frac{\mu_0^2}{\sigma_0^2}\right)T\right\}.$$

Da ϑ_0 beliebig aus Θ gewählt werden kann, setzen wir $\mu_0 = 0$.
Unter Beachtung von $t_m = T$ folgt

$$L^X(\vartheta\,;\,x) = \left(\frac{\sigma^2}{\sigma_0^2}\right)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\left(\sigma^{-2}-\sigma_0^{-2}\right)\sum_{k=1}^{m}\frac{(x_k-x_{k-1})^2}{t_k-t_{k-1}} + \frac{\mu}{\sigma^2}x_m - \frac{1}{2}\frac{\mu^2}{\sigma^2}T\right\}$$

mit $x = (x_1, x_2, \ldots, x_m)$.

Für die stochastische Likelihoodfunktion ergibt sich (siehe (4.9)):

$$L(\vartheta\,;\,\mathscr{A}^X)(\omega) = L^X(\vartheta\,;\,X(\omega)) = \frac{d\,\mathbb{P}_\vartheta\big|_{\mathscr{A}^X}}{d\,\mathbb{P}_{\vartheta_0}\big|_{\mathscr{A}^X}} =$$
$$= \left(\frac{\sigma^2}{\sigma_0^2}\right)^{-\frac{m}{2}} \exp\left\{-\frac{1}{2}\left(\sigma^{-2}-\sigma_0^{-2}\right)\sum_{k=1}^{m}\frac{\big(W(t_k)-W(t_{k-1})\big)^2}{t_k-t_{k-1}} + \frac{\mu}{\sigma^2}W(T) - \frac{1}{2}\frac{\mu^2}{\sigma^2}T\right\}$$

Der Maximum-Likelihood-Schätzer für $\vartheta = (\mu, \sigma^2)^T$ ergibt sich aus den Maximum-Likelihood-Gleichungen

$$\frac{\partial}{\partial\mu}\log L(\vartheta\,;\,\mathscr{A}^X)(\omega) = 0 \qquad \text{und}$$
$$\frac{\partial}{\partial\sigma^2}\log L(\vartheta\,;\,\mathscr{A}^X)(\omega) = 0.$$

Wir erhalten $\hat{\vartheta}_{\mathscr{T}} = (\hat{\mu}_{\mathscr{T}}, \hat{\sigma}^2_{\mathscr{T}})^T$ mit

$$\hat{\mu}_{\mathscr{T}} = \frac{W(T)}{T}, \qquad \text{unabhängig von der konkreten Gestalt von } \mathscr{T}\text{, und}$$
$$\hat{\sigma}^2_{\mathscr{T}} = \frac{1}{m}\sum_{k=1}^{m}\frac{\big(W(t_k)-W(t_{k-1})\big)^2}{t_k-t_{k-1}} - \frac{1}{m}\frac{W_T^2}{T}.$$

Wir berechnen den Erwartungswert dieser Schätzer und untersuchen ihre Streuung. Aus der Definition des Prozesses $\big(W(t),\ t \geq 0\big)$ folgt sofort:

$$\mathbb{E}_\vartheta(\hat{\mu}_{\mathscr{T}}) = \mu, \qquad D^2_\vartheta \hat{\mu}_{\mathscr{T}} = \frac{\sigma^2 T}{T^2} = \frac{\sigma^2}{T}, \qquad (\hat{\mu}_{\mathscr{T}} \text{ ist erwartungstreu})$$
$$\mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}}) = \left(1 - \frac{1}{m}\right)\sigma^2 + \frac{\mu^2 T}{m}.$$

$\hat{\sigma}^2_{\mathscr{T}}$ ist ein verfälschter (biased) Schätzer für σ^2 mit dem Bias

$$\mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}}) - \sigma^2 =: b(\vartheta, T, \mathscr{T}) = \frac{\mu^2 T - \sigma^2}{m}.$$

Es gilt

$$D^2_\vartheta \hat{\sigma}^2_{\mathscr{T}} \leq 2 D^2_\vartheta \left(\frac{1}{m} \sum_{k=1}^{m} \frac{\big(W(t_k) - W(t_{k-1})\big)^2}{t_k - t_{k-1}} \right) + 2 D^2_\vartheta \left(\frac{1}{m} \frac{W(T)^2}{T} \right). \tag{4.14}$$

(Zum Beweis beachte man $(a+b)^2 \leq 2(a^2 + b^2),\ a, b \in \mathbb{R}$.)

Untersuchung des asymptotischen Verhaltens des Schätzers $\hat{\sigma}^2_{\mathscr{T}}$ für $m \to \infty$, $T = \text{const}$.

Es seien für jedes $n \geq 1$

$$\mathscr{T}_n := \{0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_{m_n}^{(n)} = T\} \quad \text{und}$$

$\hat{\mu}_{\mathscr{T}_n}$, $\hat{\sigma}^2_{\mathscr{T}_n}$ die eben konstruierten ML-Schätzer für (μ, σ^2).

Voraussetzung: Die Anzahl der Beobachtungszeitpunkte nimmt unbegrenzt zu: $\lim_{n\to\infty} m_n = \infty$.

Bemerkung: Erwartungswert und Streuung von $\hat{\mu}_{\mathscr{T}_n}$ bleiben konstant bei wachsendem n und hängen nicht von der expliziten Form von $\mathscr{T}_n$ ab. Der Schätzer $\hat{\mu}_{\mathscr{T}_n}$ hängt nur vom letzten Beobachtungswert $W(T)$ ab, seine Streuung läßt sich durch häufigere Beobachtung innerhalb des Zeitraumes $[0, T]$ nicht verkleinern.

Für $\mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n})$ gilt offenbar $\lim_{n\to\infty} \mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n}) = \sigma^2$, d.h., $\hat{\sigma}^2_{\mathscr{T}_n}$ ist asymptotisch erwartungstreu.

Aussage 4.10. *Es gilt für jedes* $\vartheta = (\mu, \sigma^2)^T \in \Theta$

$$\lim_{n\to\infty} D_\vartheta^2 \hat{\sigma}^2_{\mathscr{T}_n} = 0 \quad \textit{und insbesondere}$$
$$\mathbb{P}_\vartheta\big(\mid \hat{\sigma}^2_{\mathscr{T}_n} - \sigma^2 \mid > \epsilon\big) \stackrel{n\to\infty}{\longrightarrow} 0 \quad \textit{für jedes } \epsilon > 0.$$

Falls $\sum_{n=1}^\infty \frac{1}{m_n} < \infty$*, so haben wir sogar*

$$\lim_{n\to\infty} \hat{\sigma}^2_{\mathscr{T}_n}(\omega) = \sigma^2, \quad \mathbb{P}_\vartheta \text{ -f.s. für jedes } \vartheta \in \Theta.$$

Beweis:
Aus (4.14) erhalten wir

$$D_\vartheta^2 \hat{\sigma}^2_{\mathscr{T}_n} \leq 2D_\vartheta^2 \left(\frac{1}{m_n} \sum_{k=1}^{m_n} \frac{\left(W(t_k^{(n)}) - W(t_{k-1}^{(n)})\right)^2}{t_k^{(n)} - t_{k-1}^{(n)}} \right) + 2D_\vartheta^2 \left(\frac{1}{m_n} \frac{W(T)^2}{T} \right) \tag{4.15}$$

Wir untersuchen beide Summanden auf der rechten Seite. Offenbar ist

$$D_\vartheta^2 \left(\frac{1}{m_n} \frac{W(T)^2}{T} \right) = \frac{1}{m_n^2 T^2} D_\vartheta^2 \left(W(T)^2\right) \longrightarrow 0 \quad \text{für } m_n \to \infty$$

Wir setzen

$$Z_k^{(n)} := \frac{W(t_k^{(n)}) - W(t_{k-1}^{(n)})}{(t_k^{(n)} - t_{k-1}^{(n)})^{\frac{1}{2}}}.$$

Dann gilt nach Definition des Wienerschen Prozesses $Z_k^{(n)} \sim \mathcal{N}\left(\mu \left(t_k^{(n)} - t_{k-1}^{(n)} \right)^{\frac{1}{2}}, \sigma^2 \right)$ und die $Z_1^{(n)}, Z_2^{(n)}, \dots, Z_{m_n}^{(n)}$ sind voneinander unabhängig. Folglich haben wir

$$\frac{1}{m_n^2} D_\vartheta^2 \left(\sum_{k=1}^{m_n} \frac{\left(W(t_k^{(n)}) - W(t_{k-1}^{(n)})\right)^2}{t_k^{(n)} - t_{k-1}^{(n)}} \right) = \frac{1}{m_n^2} D_\vartheta^2 \Big(\sum_{k=1}^{m_n} (Z_k^{(n)})^2 \Big).$$

Wir zeigen, daß die rechte Seite dieser Gleichung für $m_n \longrightarrow \infty$ gegen Null konvergiert.

Lemma 4.14. *Ist* X *eine normalverteilte Zufallsvariable,* $X \sim \mathcal{N}(\mu, \sigma^2)$*, so gilt*

$$D^2 X^2 = \mathbb{E}\big(X^4\big) - \big(\mathbb{E}(X^2)\big)^2 = 4\sigma^2\mu^2 + 2\sigma^4.$$

Beweis:

Für jede $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsgröße X haben wir
$\mathbb{E}\left(\frac{(X-\mu)^4}{\sigma^4}\right) = 3$ und $\mathbb{E}\left((X-\mu)^3\right) = 0$.
Damit erhalten wir $\mathbb{E}\left(X^3\right) = \mu^3 + 3(\sigma^2+\mu^2)\mu - 3\mu^3 = \mu^3 + 3\sigma^2\mu$
und $\mathbb{E}(X^4) = 4\mu\,\mathbb{E}(X^3) - 6\mu^2(\mu^2+\sigma^2) + 4\mu^4 - \mu^4 + 3\sigma^4 = \mu^4 + 6\sigma^2\mu^2 + 3\sigma^4$.
Mit $\big(\mathbb{E}(X^2)\big)^2 = (\mu^2+\sigma^2)^2 = \mu^4+\sigma^4+2\mu^2\sigma^2$ gilt
$D^2X^2 = \mathbb{E}\big(X^4\big) - \big(\mathbb{E}(X^2)\big)^2 = 4\sigma^2\mu^2 + 2\sigma^4$.

□

Wendet man Lemma 4.14 auf die Zufallsvariablen $Z_k^{(n)}$ an, so erhält man mit

$$D_\vartheta^2 (Z_k^{(n)})^2 = 2\sigma^4 + 4\mu^2\sigma^2(t_k^{(n)} - t_{k-1}^{(n)}) = 2\sigma^2\big(\sigma^2 + 2\mu^2(t_k^{(n)} - t_{k-1}^{(n)})\big)$$

für den ersten Summanden in (4.15):

$$\frac{1}{m_n^2} D_\vartheta^2 \left(\sum_{k=1}^{m_n} \frac{\Big(W(t_k^{(n)}) - W(t_{k-1}^{(n)})\Big)^2}{t_k^{(n)} - t_{k-1}^{(n)}} \right) = \frac{2\sigma^2}{m_n^2} \left(\sum_{k=1}^{m_n} \sigma^2 + 2\mu^2(t_k^{(n)} - t_{k-1}^{(n)}) \right) = \frac{2\sigma^4}{m_n} + \frac{4\mu^2\sigma^2}{m_n^2} \le \frac{K}{m_n} \tag{4.16}$$

für eine Konstante $K > 0$.
Folglich gilt $D_\vartheta^2 \hat{\sigma}^2_{\mathscr{T}_n} \xrightarrow{n\to\infty} 0$. Somit haben wir mittels (4.16)

$$\begin{aligned}
\mathbb{P}_\vartheta\big(\mid \hat{\sigma}^2_{\mathscr{T}_n} - \sigma^2 \mid > \epsilon\big) &= \mathbb{P}_\vartheta\big(\mid \hat{\sigma}^2_{\mathscr{T}_n} - \mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n}) + \mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n}) - \sigma^2 \mid > \epsilon\big) \\
&\le \mathbb{P}_\vartheta\big(\mid \hat{\sigma}^2_{\mathscr{T}_n} - \mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n}) \mid + \mid \mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n}) - \sigma^2 \mid > \epsilon\big) \\
&\le \mathbb{P}_\vartheta\big(\mid \hat{\sigma}^2_{\mathscr{T}_n} - \mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n}) \mid > \frac{\epsilon}{2}\big) \le \frac{4K}{m_n\epsilon^2}
\end{aligned}$$

falls n so groß ist, daß $\dfrac{\mid \mathbb{E}_\vartheta(\hat{\sigma}^2_{\mathscr{T}_n}) - \sigma^2 \mid}{m_n} = \dfrac{\mid \mu^2 T - \sigma^2 \mid}{m_n} < \dfrac{\epsilon}{2}$ gilt.
Mit Hilfe der Tschebyschevschen Ungleichung ergibt sich die Behauptung. Ist nun $\sum_{n\ge1} \frac{1}{m_n} < \infty$, so folgt für jedes $\epsilon > 0$

$$\mathbb{P}_\vartheta\big(\limsup_{n\to\infty} A_n^\epsilon\big) = 0 \;\text{ mit }\; A_n^\epsilon = \big\{\omega \,:\, \mid \hat{\sigma}^2_{\mathscr{T}_n}(\omega) - \sigma^2 \mid > \epsilon\big\}. \qquad \text{(Borel-Cantelli-Lemma)}$$

Das bedeutet $\hat{\sigma}^2_{\mathscr{T}_n} \longrightarrow \sigma^2 \quad \mathbb{P}_\vartheta$-f.s. für jedes $\vartheta \in \Theta$. □

Folgerungen:

a) Bei unbegrenztem Wachstum der Zahl m_n der Beobachtungspunkte läßt sich σ^2 aus $\mathbb{P}_\vartheta$- fast jeder Trajektorie $\big(W(t),\ t \in [0,T]\big)$ beliebig genau bestimmen ($\vartheta = (\mu, \sigma^2)^T$). Das gilt für jedes $T > 0$ (!).

b) Gilt $\sigma^2 = \sigma_0^2$, so ist

$$L^X(\mu\,;\,X) = \exp\left\{\frac{\mu}{\sigma^2}W(T) - \frac{\mu^2}{\sigma^2}T\right\}$$

für jede Stichprobe $X = \big(W(t_1), \ldots, W(t_m)\big)^T$ mit $t_m = T$.

Es gilt also

$$\mathbb{P}_\vartheta(A) = \int_A \exp\left\{\frac{\mu}{\sigma^2}W(T) - \frac{\mu^2}{\sigma^2}T\right\} d\,\mathbb{P}_{\vartheta_0} \quad \text{mit } \vartheta = (\mu, \sigma^2)^T,\ \vartheta_0 = (0, \sigma^2)^T \tag{4.17}$$

für jedes Ereignis A der Form $A = \{\omega\ :\ (W(t_1), \ldots, W(t_m)) \in B\}$, $\quad B \in \mathscr{B}^m$, mit $t_m = T$, d.h. (a) ist gültig für jede Zylindermenge $Z = Z_{t_1,\ldots,t_m\,;\,B}$ mit $0 < t_1 < \ldots < t_m = T$, und somit für jedes A aus $\mathscr{B}\big(\mathbb{C}([0,T])\big)$.
Damit haben wir für die Familie

$$\mathscr{P}_{\sigma^2} := \big(\mathbb{P}_\vartheta, \vartheta = (\mu, \sigma^2)^T,\ \mu \in \mathbb{R}\big)$$

bei fixiertem $\sigma^2 > 0$ erhalten, daß alle $\mathbb{P}_\vartheta^T := \mathbb{P}_\vartheta\big|_{\mathscr{B}(\mathbb{C}([0,T]))}$ zueinander äquivalent sind, und daß die Likelihoodfunktion

$$L(\vartheta\,;\,\mathscr{A}_T) = \frac{\mathbb{P}_\vartheta^T}{\mathbb{P}_{\vartheta_0}^T} \quad \text{für } \vartheta = (\mu, \sigma^2)^T,\ \vartheta_0 = (0, \sigma^2)^T$$

die Form

$$L(\vartheta\,;\,\mathscr{A}_T) = \exp\left\{\frac{\mu}{\sigma^2}W(T) - \frac{\mu^2}{\sigma^2}T\right\}$$

besitzt. Hierbei haben wir die Notation

$$\mathbb{P}_\vartheta^T := \mathbb{P}_\vartheta\big|_{\mathscr{A}_T}, \quad \text{mit } \mathscr{A}_T := \sigma\big(W(s),\ s \in [0,T]\big)$$

verwendet.

Kapitel 5

Asymptotik von Likelihoodfunktionen: allgemeiner Fall

Im vorangegangenen Kapitel haben wir im Fall des Wienerschen Prozesses gesehen, daß bei unbegrenzter Erhöhung der Beobachtungszahl die Likelihoodfunktion konvergiert, wobei als Grenzwerte Null und Unendlich auftreten können. Wir werden in diesem Kapitel zeigen, daß dies in sehr viel allgemeineren Fällen ebenfalls gilt.

5.1 Definitionen, Martingaleigenschaften

Die folgenden Bezeichnungen und Voraussetzungen gelten für alle Abschnitte dieses Kapitels.

Es seien P und Q zwei Wahrscheinlichkeitsmaße auf $(\Omega, \mathscr{A})$ und $(\mathscr{A}_n)_{n\geq 1}$ eine Filtration von σ-Algebren in $\mathscr{A}$, d.h., es gelte $\mathscr{A}_n \subseteq \mathscr{A}_{n+1} \subseteq \mathscr{A}$, $n \geq 1$. Weiterhin sei $\mathscr{A} = \bigvee_{n\geq 1} \mathscr{A}_n := \sigma\left(\bigcup_{n\geq 1} \mathscr{A}_n\right)$. (Wie wir bereits wissen, ist $\bigcup_{n\geq 1} \mathscr{A}_n$ eine Algebra von Teilmengen von Ω, aber i.a. keine σ-Algebra.) Mit P_n bzw. Q_n bezeichnen wir die Einschränkungen von P bzw. Q auf $\mathscr{A}_n$.

Um die Beziehung zu stochastischen Prozessen (hier mit diskreter Zeit) herzustellen, stellen wir uns vor, daß auf $(\Omega, \mathscr{A})$ Zufallsgrößen $(X_n,\ n \geq 1)$ gegeben sind, und daß $\mathscr{A}_n = \sigma\left(X_1, X_2, \ldots, X_n\right)$ gilt.

Definition 5.1. Q heißt *lokal absolutstetig bezüglich* P, falls $Q_n \ll P_n$ für alle $n \geq 1$ gilt.

Es sei Q lokal absolutstetig bezüglich P und es sei $L_n := \dfrac{dQ_n}{dP_n}$, $n \geq 1$. Die Folge $(L_n, n \geq 1)$ heißt der *Likelihoodprozess von Q bezüglich P.*

Lemma 5.2. *$(L_n, \mathscr{A}_n, n \geq 1)$ ist ein nichtnegatives Martingal bezüglich P und $(L_n^{\frac{1}{2}}, \mathscr{A}_n, n \geq 1)$ ist ein gleichgradig integrierbares nichtnegatives Supermartingal bezüglich P.*

Beweis:
L_n und $L_n^{\frac{1}{2}}$ sind nach Definition (vgl. Radon-Nikodym-Theorem) $\mathscr{A}_n$-meßbar. Außerdem gilt für jedes $A \in \mathscr{A}_n$

$$\int_A L_n \, dP = Q(A) = \int_A L_{n+1} \, dP. \qquad \text{(wegen } \mathscr{A}_n \subseteq \mathscr{A}_{n+1}\text{)}$$

Das ist die Martingaleigenschaft für $(L_n, \mathscr{A}_n)$. Daß $L_n^{\frac{1}{2}}$ ein Supermartingal ist, folgt mittels der Jensenschen Ungleichung und der Tatsache, daß $x \longrightarrow -x^{\frac{1}{2}}$, $(x > 0)$ konvex ist:

$$0 \leq \mathbb{E}_P(L_{n+1}^{\frac{1}{2}} \mid \mathscr{A}_n) \leq (\mathbb{E}_P \, (L_{n+1} \mid \mathscr{A}_n))^{\frac{1}{2}} = L_n^{\frac{1}{2}}.$$

□

Es gilt

$$L_n \geq 0 \text{ und } \mathbb{E}_P(L_n) = \int_\Omega \frac{dQ_n}{dP_n} \, dP = 1, \text{ sowie } \mathbb{E}_P(L_n^{\frac{1}{2}}) \leq \mathbb{E}_P(L_n) = 1. \tag{5.1}$$

(Schwarzsche Ungleichung)

Aus dem Martingalkonvergenzsatz folgt, daß (L_n) P-fast sicher gegen eine nichtnegative Zufallsgröße L_∞ konvergiert, für die, wegen $\mathbb{E}_P(L_n) = 1$ und wegen des Fatouschen Lemmas, $\mathbb{E}_P(L_\infty) \leq 1$ gilt. Insbesondere ist $P(L_\infty < \infty) = 1$.

Für die Folge $\rho_n := \mathbb{E}_P(L_n^{\frac{1}{2}})$ erhalten wir

$$0 < \rho_{n+1} \leq \rho_n \leq 1, \quad n \geq 1.$$

Folglich existiert der Grenzwert $\rho := \lim_{n \to \infty} \rho_n$.
Die gleichgradige Integrierbarkeit von $(L_n^{\frac{1}{2}})$ ergibt sich aus

$$\int_{\{L_n^{\frac{1}{2}} > c\}} L_n^{\frac{1}{2}} \, dP \leq P(L_n^{\frac{1}{2}} > c) \cdot \mathbb{E}_P(L_n) = P(L_n^{\frac{1}{2}} > c) \leq \frac{E_P(L_n)}{c^2} = \frac{1}{c^2} \quad \text{für alle } c > 0, n \geq 1.$$

Somit gilt also auch

$$\lim_{n\to\infty} \mathbb{E}_P\left(\left|L_n^{\frac{1}{2}} - L_\infty^{\frac{1}{2}}\right|\right) = 0 \text{ und insbesondere}$$
$$\rho = \lim_{n\to\infty} \mathbb{E}_P(L_n^{\frac{1}{2}}) = \mathbb{E}_P(L_\infty^{\frac{1}{2}}) \tag{5.2}$$

□

5.2 Konvergenz des Likelihoodprozesses

Wir kommen nun zum zentralen Punkt dieses Kapitels, dem folgenden Theorem:

Aussage 5.1. *Es seien P und Q zwei Wahrscheinlichkeitsmaße auf $(\Omega, \mathscr{A}, (\mathscr{A}_n))$, so daß Q lokal absolutstetig bezüglich P ist. Dann gelten folgende Aussagen:*

a) (L_n) konvergiert $(P+Q)$-fast überall gegen eine Zufallsgröße L_∞ mit $0 \le L_\infty \le \infty$, und es gilt:

$$Q(A) = \int_A L_\infty \, dP + Q\left(A \cap \{L_\infty = \infty\}\right) \tag{5.3}$$

Insbesondere konvergiert (L_n) sowohl P- als auch Q-fast sicher gegen L_∞.

b) Jede der folgenden Bedingungen ist notwendig und hinreichend dafür, daß Q absolutstetig bezüglich P (auf $\mathscr{A} = \bigvee_{n\ge1} \mathscr{A}_n$) ist:

i) $Q(L_\infty = \infty) = 0$,

ii) $\mathbb{E}_P(L_\infty) = 1$.

In diesem Fall ist

$$Q(A) = \int_A L_\infty \, dP, \quad A \in \mathscr{A}$$

und L_n konvergiert im $L_1(P)$-Sinne gegen L_∞. (Insbesondere ist (L_n) bezüglich P gleichgradig integrierbar.)

c) Jede der folgenden Bedingungen ist notwendig und hinreichend dafür, daß Q und P auf $\mathscr{A}_\infty$ singulär zueinander sind:

i) $Q(L_\infty = \infty) = 1$,

ii) $\mathbb{E}_P(L_\infty) = 0$.

In diesem Fall gilt

$$\lim_{n\to\infty} L_n = L_\infty = \infty \quad Q\text{-}f.s.,$$
$$\lim_{n\to\infty} L_n = L_\infty = 0 \quad P\text{-}f.s.$$

Bemerkungen:

1. Zwei Wahrscheinlichkeitsmaße $\mathbb{P}_1$ und $\mathbb{P}_2$ heißen *singulär zueinander*, wenn es ein $A_0 \in \mathscr{A}$ mit $P(A_0) = 1,\ Q(\Omega \setminus A_0) = 1$ gibt.

2. (5.3) liefert die Zerlegung von Q in einen bezüglich P absolutstetigen und einen bezüglich P singulären Anteil, man nennt sie die „Lebesguesche Zerlegung“ von Q bezüglich P.

Beweis:
Zu a)
<u>1. Schritt:</u> Einführung des Martingals $(U_n, \mathscr{A}_n)$:
Wir definieren $R_n := \frac{1}{2}(P_n + Q_n)$ und $R := \frac{1}{2}(P + Q)$.
Offenbar gilt

$$Q_n \ll R_n, \quad P_n \ll R_n, \quad Q \ll R, \quad P \ll R.$$

Weiterhin definieren wir für jedes $n \geq 1$

$$U_n := \frac{dQ_n}{dP_n}.$$

Dann ist $(U_n, \mathscr{A}_n)$ ein nichtnegatives Martingal bezüglich R mit

$$0 \leq U_n \leq 2 \quad R\text{-fast sicher.} \tag{5.4}$$

Die Martingaleigenschaft von $(U_n, \mathscr{A}_n)$ beweist man wie im Abschnitt 5.1. Die Ungleichungen (5.4) erhält man aus

$$\int_A U_n \, dR_n = Q_n(A) \leq Q_n(A) + P_n(A) = \int_A 2 \, dR_n, \quad A \in \mathscr{A}_n.$$

Somit existiert der Grenzwert $U_\infty = \lim_{n\to\infty} U_n$ fast sicher bezüglich R, U_∞ ist $\mathscr{A}$-meßbar, und es gilt (Satz von Lebesgue über die majorisierte Konvergenz)

$$\lim_{n\to\infty} \mathbb{E}_P \left(|U_n - U_\infty|\right) = 0. \tag{5.5}$$

2. Schritt: Identifikation von U_∞.
Für jedes $A \in \bigcup_{n\geq 1} \mathscr{A}_n$ gilt

$$Q(A) = \lim_{n\to\infty} Q_n(A) = \lim_{n\to\infty} \int_A U_n \, dR = \int_A U_\infty \, dR,$$

daraus folgt (! Fortsetzungssatz für σ-additive Mengenfunktionen von einer Algebra auf die von ihr erzeugte σ-Algebra)

$$Q \ll R \quad \text{und} \quad \frac{dQ}{dR}(\omega) = U_\infty(\omega) \quad R\text{-fast sicher.}$$

(Es sei daran erinnert, daß Q und R Wahrscheinlichkeitsmaße auf $\mathscr{A}_\infty$ sind.)
Analog ergibt sich mit

$$V_n := \frac{dP_n}{dR_n} = 2 - U_n \quad \text{und} \quad V_\infty := \lim_{n\to\infty} V_n$$

die Beziehung

$$V_\infty = \frac{dP}{dR} = 2 - U_\infty \quad R\text{-fast sicher.}$$

Als Folgerung aus (5.4) erhalten wir für alle $A \in \mathscr{A}$

$$\int_A U_\infty \, dP = \int_A U_\infty \frac{dP}{dR} \, dR = \int_A (2 - U_\infty) \, dQ, \quad A \in \mathscr{A}. \tag{5.6}$$

Weiterhin haben wir

$$P(U_\infty = 2) = \int_{\{U_\infty = 2\}} (2 - U_\infty) \, dR = 0 \tag{5.7}$$

und

$$Q(U_\infty = 0) = \int_{\{U_\infty = 0\}} U_\infty \, dR = 0. \tag{5.8}$$

Lemma 5.3. *Es gilt $Q \ll P$ (auf $\mathscr{A}$) genau dann, wenn $Q(U_\infty = 2) = 0$. In diesem Fall ist*

$$Q(A) = \int_A \frac{U_\infty}{2 - U_\infty} \, dP, \quad A \in \mathscr{A}. \tag{5.9}$$

Beweis des Lemmas:

„⇒“-Richtung: Da $P(U_\infty = 2) = 0$, folgt aus $Q \ll P$ auch $Q(U_\infty = 2) = 0$.

„⇐“-Richtung: Gilt $Q(U_\infty = 2) = 0$, so haben wir für alle $A \in \mathscr{A}$

$$Q(A) = \int_{A\cap\{U_\infty<2\}} U_\infty \, dR = \int_{A\cap\{U_\infty<2\}} U_\infty \cdot \frac{2-U_\infty}{2-U_\infty} \, dR = \int_{A\cap\{U_\infty<2\}} \frac{U_\infty}{2-U_\infty} \, dP.$$

Wegen (5.7) folgt daraus die Formel (5.9). Ist nun $P(A) = 0$, so ist wegen (5.9) auch $Q(A) = 0$. Also gilt $Q \ll P$. □

3. Schritt: Die Zerlegung der Verteilung Q

Wir definieren für alle $\omega \in \Omega$ und alle n mit $1 \le n \le \infty$

$$\Lambda_n(\omega) := \begin{cases} \frac{U_n(\omega)}{2-U_n(\omega)}, & \text{falls } U_n(\omega) < 2, \\ \infty, & \text{falls } U_n(\omega) = 2. \end{cases}$$

Offenbar gilt

$$\{\omega : \Lambda_\infty(\omega) = \infty\} = \{\omega : U_\infty(\omega) = 2\}. \tag{5.10}$$

$$\begin{aligned} \text{Wegen } \lim_{n\to\infty} U_n(\omega) &= U_\infty(\omega) \qquad R\text{-fast sicher,} \\ \text{folgt } \lim_{n\to\infty} \Lambda_n(\omega) &= \Lambda_\infty(\omega) \qquad R\text{-fast sicher.} \end{aligned}$$

Daraus, aus (5.7), sowie (5.10) ergibt sich $P(\Lambda_\infty = \infty) = 0$.
Aus (5.6) und (5.7) ergibt sich für jedes $A \in \mathscr{A}$

$$\begin{aligned} \int_A U_\infty \, dP = \int_A \mathbf{1}_{\{U_\infty<2\}} U_\infty \, dP = \int_A \mathbf{1}_{\{U_\infty<2\}} U_\infty \frac{dP}{dR} \, dR = \\ \int_A \mathbf{1}_{\{U_\infty<2\}} U_\infty (2-U_\infty) \, dR = \int_A \mathbf{1}_{\{U_\infty<2\}} (2-U_\infty) \, dQ. \end{aligned}$$

Daraus folgt für jedes $A \in \mathscr{A}$ und jede nichtnegative $\mathscr{A}$-meßbare Zufallsgröße Z mittels der Approximationsmethode

$$\int_A \mathbf{1}_{\{U_\infty<2\}} Z U_\infty \, dP = \int_A \mathbf{1}_{\{U_\infty<2\}} Z (2-U_\infty) \, dQ$$

und insbesondere für $Z = \mathbf{1}_{\{U_\infty<2\}}(2-U_\infty)^{-1}$ folgt aus (5.9)

$$\int_A \frac{U_\infty}{2-U_\infty} \mathbf{1}_{\{U_\infty<2\}} \, dP = Q(A \cap \{U_\infty < 2\}).$$

Mittels (5.7) ergibt sich für alle $A \in \mathscr{A}$

$$Q(A \cap \{U_\infty < 2\}) = \int_A \Lambda_\infty \, dP \quad \text{und damit}$$

$$Q(A) = Q(A \cap \{\Lambda_\infty = \infty\}) + \int_A \Lambda_\infty \, dP, \quad A \in \mathscr{A}, \tag{5.11}$$

wobei (siehe oben) $Q(\Lambda_\infty = 0) = 0$ und $P(\Lambda_\infty = \infty) = 0$ gelten.

<u>4. Schritt:</u> Identifikation von $\Lambda_n,\ n \leq \infty$
Wegen

$$\frac{dP_n}{dR_n} = 2 - U_n, \quad \text{gilt} \quad P_n(A) = \int_A (2 - U_n) \, dR_n$$

und somit $P_n(U_n = 2) = 0$.
Damit ist

$$\int_A U_n \, dP_n = \int_{A \cap \{U_n < 2\}} U_n \, dP_n = \int_{A \cap \{U_n < 2\}} U_n \frac{dP_n}{dR_n} \, dR_n = \int_{A \cap \{U_n < 2\}} (2 - U_n) \, dQ_n.$$

Mit der gleichen Methode wie im Lemma des 2. Schrittes folgt nun

$$\int_A \frac{U_n}{2 - U_n} \, dP_n = Q_n(A), \quad A \in \mathscr{A}_n.$$

Das bedeutet aber
$L_n = \frac{U_n}{2-U_n}$ P_n-f.s., und wegen $Q_n \ll P_n$ auch R_n-f.s., also auch R-f.s. Insbesondere ist $L_n < \infty$ R-f.s.
Wegen $\lim\limits_{n\to\infty} U_n = U_\infty$ R-f.s., ergibt sich

$$\lim_{n\to\infty} U_n = \Lambda_\infty \ \ R\text{-f.s.} \quad \text{und damit} \quad \lim_{n\to\infty} L_n =: L_\infty \ R\text{-f.s. und } L_\infty = \Lambda_\infty \ R\text{-f.s.}$$

Unter Berücksichtigung von (5.11) ist damit Teil a) bewiesen.
Zu b)
Ist $Q \ll P$, so gilt wegen $P(L_\infty = \infty) = 0$ auch $Q(L_\infty = \infty) = 0$ und

$$1 = Q(\Omega) = \int_\Omega L_\infty \, dP = \mathbb{E}_P(L_\infty).$$

Ist umgekehrt $Q(L_\infty = \infty) = 0$, so folgt aus a), daß $Q \ll P$ gilt und $1 = \mathbb{E}_P(L_\infty)$ erfüllt ist.
Im Fall $\mathbb{E}_P(L_\infty) = 1$ muß $Q(L_\infty = \infty) = 0$ gelten.
Zu c)

Aus $Q(L_\infty = \infty) = 1$ folgt $\mathbb{E}_P(L_\infty) = 0$, also $P(L_\infty = 0) = 1$, somit sind P und Q singulär.
Umgekehrt, sind P und Q singulär, gibt es also ein $A_0 \in \mathscr{A}$ mit $P(A_0) = 1,\ Q(\Omega \setminus A_0) = 1$, so ist

$$1 = Q(\Omega \setminus A_0) = \int_{\Omega \setminus A_0} L_\infty \, dP + Q\Big(\{L_\infty = \infty\} \cap \Omega \setminus A_0\Big) = 0 + Q(L_\infty = \infty).$$

□

Folgerung 5.4. Unter den Voraussetzungen der Aussage 5.1 gilt mit $\rho = \lim_{n\to\infty} \mathbb{E}_P(L_n{}^{\frac{1}{2}})$ (siehe Abschnitt 5.1):
Falls $\rho = 0$, so sind P und Q auf $\mathscr{A}$ zueinander singulär.

Beweis:
Es gilt $\rho = \mathbb{E}_P(L_\infty^{\frac{1}{2}})$, folglich ist $L_\infty = 0$ P-fast sicher, und die Behauptung folgt aus Teil c) der Aussage. □

5.3 Eine Anwendung

Es seien P und Q zwei Wahrscheinlichkeitsmaße auf $(\Omega, \mathscr{A})$ und es sei $(X_n,\ n \geq 1)$ eine Folge reellwertiger Zufallsgrößen über $(\Omega, \mathscr{A})$, die bezüglich P und bezüglich Q voneinander unabhängig seien. Für die Verteilungen jeder der X_n gelte

$$Q^{X_n} \ll P^{X_n}, \text{ d.h. } P(X_n \in B) = 0 \quad \text{impliziert} \quad Q(X_n \in B) = 0, \quad B \in \mathscr{B}.$$

Wir setzen

$$f_n(x) = \frac{dQ^{X_n}}{dP^{X_n}}(x), \quad x \in \mathbb{R}\,.$$

Mit $\mathscr{A}_n$ werde die von $X_1, X_2, \ldots, X_n$ erzeugte Teil-σ-Algebra von $\mathscr{A}$ bezeichnet, und ohne Beschränkung der Allgemeinheit gelte $\mathscr{A} = \mathscr{A}_\infty := \bigvee_{n\geq 1} \mathscr{A}_n$. Mit der Terminologie von Abschnitt 5.1 gilt, daß Q bezüglich P lokal absolutstetig ist. Auf Grund der Unabhängigkeit der $X_n,\ n \geq 1$ unter P und Q gilt

$$L_n(\omega) = \frac{dQ_n}{dP_n}(\omega) = \prod_{k=1}^{n} f_k\left(X_k\left(\omega\right)\right).$$

Wir definieren

$$\rho := \prod_{k=1}^{\infty} \mathbb{E}_p\left(f_k^{\frac{1}{2}}\left(X_k\right)\right).$$

Aussage 5.2 (Null-Eins-Gesetz von Kakutani). *Q ist auf $\mathscr{A}_\infty$ entweder absolutstetig oder singulär zu P, je nachdem, ob $\rho > 0$ oder $\rho = 0$ gilt.*

Beweis:
Ist $\rho > 0$, so bildet $(L_n^{\frac{1}{2}},\ n \geq 1)$ eine Cauchyfolge in $L_2(P)$. Das ergibt sich aus

$$\mathbb{E}_P \left| L_n^{\frac{1}{2}} - L_m^{\frac{1}{2}} \right|^2 = 2 - 2\,\mathbb{E}_P \left(L_n^{\frac{1}{2}} \cdot L_m^{\frac{1}{2}} \right)$$

und (falls $n < m$, wir nutzen die Unabhängigkeit der $(X_k,\ k \geq 1)$)

$$\mathbb{E}_P \left(L_n^{\frac{1}{2}} \cdot L_m^{\frac{1}{2}} \right) = \mathbb{E}_P(L_n) \cdot \mathbb{E}_P \left(\prod_{k=n+1}^{m} f_k^{\frac{1}{2}}\,(X_k) \right) = \prod_{k=n+1}^{m} \mathbb{E}_P \left(f_k^{\frac{1}{2}}\,(X_k) \right) = \frac{\rho_m}{\rho_n}$$

mit der Notation $\rho_n = \mathbb{E}_P \left(L_n^{\frac{1}{2}} \right) = \prod_{k=1}^{n} \mathbb{E}_P \left(f_k^{\frac{1}{2}}\,(X_k) \right)$.
Wegen $\rho > 0$ ist $\lim\limits_{m,n \to \infty} \dfrac{\rho_m}{\rho_n} = 1$.
Da (L_n) P-fast sicher gegen L_∞ konvergiert, muß folglich $(L_n^{\frac{1}{2}})$ im $L_2(P)$-Sinne gegen $L_\infty^{\frac{1}{2}}$ konvergieren. Das bedeutet

$$\lim_{n \to \infty} \mathbb{E}_P \mid L_n - L_\infty \mid = 0,$$

und somit gilt wegen $\mathbb{E}_P(L_n) = 1$ auch $\mathbb{E}_P(L_\infty) = 1$. Nun wenden wir die Aussage 5.1 an und erhalten $Q \ll P$ mit

$$\frac{dQ}{dP}(\omega) = \prod_{k=1}^{\infty} f_k\,(X_k\,(\omega))\,.$$

Ist $\rho = 0$, so ergibt sich die Singularität von P und Q aus der Folgerung 5.4. $\square$

Kapitel 6

Suffiziente Statistiken

In diesem Kapitel untersuchen wir einen weiteren statistischen Begriff, der eng mit Likelihoodfunktionen zusammenhängt und mit der Frage nach eventuell möglicher Datenreduktion zu tun hat. Wir gehen der Frage nach, in welchen Fällen man an Stelle der gesamten Stichprobe X lediglich Stichprobenfunktionen $H(X)$ zu verwenden braucht, ohne Informationen über den zugrunde liegenden Parameter ϑ zu verlieren.

6.1 Vorbetrachtungen

Es sei $(\Omega, \mathscr{A}, \mathscr{P}, X)$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und X als Stichprobe mit Werten in einem meßbaren Raum $(E, \mathscr{E})$. Die Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta^X$ von X auf $(E, \mathscr{E})$ hängt von ϑ ab. Hat man eine konkrete Stichprobe x, also eine Realisierung von X erhalten, so kann man i.a. daraus Schlüsse über die dem Experiment zugrunde liegende Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta$, also den wahren Parameter $\vartheta \in \Theta$ ziehen.

Beispiel 6.1. Es sei $X = (X_1, X_2, \ldots, X_n)$ eine mathematische Stichprobe aus einer auf $[\vartheta, \vartheta + 1]$ gleichmäßig verteilten Grundgesamtheit, $\vartheta \in \mathbb{R}$ sei unbekannt. Hat man eine Realisierung $x = (x_1, \ldots, x_n)$ von X erhalten, so kann man schließen, daß

$$(\max_{k=1,\ldots,n} x_k) - 1 \leq \vartheta \leq \min_{k=1,\ldots,n} x_k$$

gilt. Man erhält also aus x Informationen über ϑ.

Beispiel 6.2. Ist $X = (X_1, X_2, \ldots, X_n)$ eine mathematische Stichprobe aus einer exponentiell mit dem Parameter λ verteilten Grundgesamtheit, so kann man erwarten, daß

$$\bar{X}_n := \frac{1}{n} \sum_{k=1}^{n} X_k$$

in der Nähe von $\mathbb{E}_\lambda(X_1) = \dfrac{1}{\lambda}$ liegt (Gesetz der großen Zahlen).

Beispiel 6.3. Aus der Beobachtung einer Trajektorie $X = \big(W(s),\ s \in [0,T]\big)$ des Wienerschen Prozesses mit den Parametern μ und σ^2 auf dem Intervall $[0,T]$ $(T > 0)$ läßt sich σ^2 exakt berechnen (siehe Kapitel 4).

In jedem der genannten Fälle enthält also die Stichprobe X Informationen über den wahren Parameter ϑ.

Wir kehren zum eingangs formulierten allgemeinen Fall zurück.
Ist $H(\cdot)$ eine Stichprobenfunktion auf $(E, \mathscr{E})$ (d.h. eine Statistik) mit Werten in $(F, \mathscr{F})$, so hängt auch ihre Wahrscheinlichkeitsverteilung $\mathbb{P}_\vartheta^H$ auf $(F, \mathscr{F})$ im allgemeinen von ϑ ab. Wir haben das im Beispiel 6.2 mit $H(X) = \bar{X}_n$ gesehen.

Es entsteht die Frage, ob in $H(X)$ noch genau soviel Information über ϑ enthalten ist, wie in der Stichprobe X selbst. Ist dies der Fall, so nennt man $H(X)$ eine *suffiziente* (oder auch *erschöpfende*) *Statistik für* $\vartheta \in \Theta$ *bzw. für* $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$. Wir gehen darauf im Folgenden näher ein.

6.2 Suffiziente Statistiken und suffiziente σ-Algebren

Wir verwenden die Bezeichnungen des Abschnitts 6.1. Ausgangspunkt der folgenden Überlegungen ist die Formel

$$\mathbb{P}_\vartheta^X(B) = \int_F \mathbb{P}_\vartheta^X(B \mid H = h)\, \mathbb{P}_\vartheta^H(\,dh), \quad B \in \mathscr{E}, \tag{6.1}$$

die man als Verallgemeinerung der Formel der totalen Wahrscheinlichkeit ansehen kann. Die Abhängigkeit der Verteilung $\mathbb{P}_\vartheta^X$ von ϑ wird also aufgespaltet in die Abhängigkeit von ϑ der Verteilung $\mathbb{P}_\vartheta^H$ der Statistik H und die Abhängigkeit von ϑ der bedingten Verteilung $\mathbb{P}_\vartheta^X(\cdot \mid H = h)$ von X unter $H = h$. Beschränkt man sich bei der Datenerhebung auf die Statistik $H(X)$ anstelle X, so geht also ein Teil der Abhängigkeit (und damit ein Teil der Information über ϑ, die in X steckt) verloren.

Beispiel 6.4. Es sei $X = (X_1, X_2, \ldots, X_n)$ eine mathematische Stichprobe aus einer zweipunktverteilten Grundgesamtheit mit dem Parameter $p \in (0,1)$, genauer, $X_1, X_2, \ldots, X_n$ seien unabhängig und identisch verteilt, und es gelte

$$\mathbb{P}_p(X_k = 1) = p, \quad \mathbb{P}_p(X_k = 0) = 1 - p. \qquad \text{(Bernoullischema mit dem Parameter } p\text{)}$$

Wir definieren $H(X) := \max\limits_{k=1,\dots,n} X_k$.
Die Statistik H hat die möglichen Werte 0 und 1 und besitzt die Verteilung

$$\mathbb{P}(H=0) = (1-p)^n, \quad \mathbb{P}(H=1) = 1-(1-p)^n.$$

Für $\mathbb{P}_p(X = x \mid H = i), \quad x = (i_1, i_2, \dots, i_n) \in \{0,1\}^n,\ i \in \{0,1\}$ gilt

$$\begin{aligned}
\mathbb{P}_p(X = x \mid H = 0) =& 1, \quad \text{falls } x = (0,0,\dots,0), \quad \mathbb{P}_p(X = x \mid H = 0) = 0 \ \text{ sonst},\\
\mathbb{P}_p(X = x \mid H = 1) =& \frac{p^{\sum_{j=1}^n i_j}(1-p)^{n-\sum_{j=1}^n i_j}}{1-(1-p)^n}, \quad \text{falls}\\
& x = (i_1, i_2, \dots, i_n) \in \{0,1\}^n \text{ mit } H(x) = 1.
\end{aligned}$$

Offensichtlich enthält $H(X) = \max_{k=1,\dots,n} X_k$ weit weniger Information über den wahren Parameter p als die Stichprobe X selbst, aus der man zum Beispiel durch $\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ eine bessere Schätzung für p gewinnt als es auf der Grundlage von $H(X)$ möglich wäre.

Wir behandeln im Folgenden die am Ende des Abschnitts 6.1 aufgeworfene Frage in einem allgemeineren Rahmen.

Es seien $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und $\mathscr{H}$ eine Teil-σ-Algebra von $\mathscr{A}$.

Definition 6.1. Die σ-Algebra $\mathscr{H}$ heißt *suffizient für $\mathscr{P}$ (bzw. für $\vartheta \in \Theta$) hinsichtlich $\mathscr{A}$*, falls es für alle $A \in \mathscr{A}$ eine nicht von ϑ abhängende $\mathscr{H}$-meßbare Zufallsgröße Z_A gibt mit

$$Z_A = \mathbb{E}_\vartheta(\mathbf{1}_A \mid \mathscr{H}) \quad \mathbb{P}_\vartheta\text{-fast sicher}, \quad \forall \vartheta \in \Theta. \tag{6.2}$$

Anstelle Z_A schreiben wir symbolisch $\mathbb{P}(A \mid \mathscr{H})$ um anzudeuten, daß Z_A für jedes ϑ eine Version der bedingten Wahrscheinlichkeit $\mathbb{P}_\vartheta(A \mid \mathscr{H})$ von A unter $\mathscr{H}$ ist, allerdings aber nicht von ϑ abhängt.
Ist $\mathscr{H}$ suffizient, so gilt also

$$\mathbb{P}_\vartheta(A) = \int_\Omega \mathbb{P}(A \mid \mathscr{H})\, d\,\mathbb{P}_\vartheta\Big|_{\mathscr{H}} \quad \text{für alle } A \in \mathscr{A},\ \vartheta \in \Theta. \tag{6.3}$$

Hier bestimmt bereits die Abhängigkeit der Einschränkung $\mathbb{P}_\vartheta\Big|_{\mathscr{H}}$ von ϑ diejenige der $\mathbb{P}_\vartheta$ auf $\mathscr{A}$.

Definition 6.2. Eine Zufallsgröße T auf $(\Omega, \mathscr{A})$ mit Werten in $(E, \mathscr{E})$ nennt man *suffizient für $\mathscr{P}$ hinsichtlich $\mathscr{A}$*, falls $\mathscr{A}^T := T^{-1}(\mathscr{E})$ suffizient im genannten Sinne ist.

In diesem Fall gilt für alle $A \in \mathscr{A}$ und $\vartheta \in \Theta$:

$$\mathbb{P}_\vartheta(A) = \int_\Omega \mathbb{P}(A \mid \sigma(T))\, d\,\mathbb{P}_\vartheta = \int_E \mathbb{P}(A \mid T = t)\, \mathbb{P}_\vartheta^T(\, dt). \tag{6.4}$$

Hier bestimmt die Abhängigkeit der Verteilung $\mathbb{P}_\vartheta^T$ von ϑ bereits die Abhängigkeit der Maße $\mathbb{P}_\vartheta$ von ϑ.

Bemerkung 6.3. Anstelle der σ-Algebra $\mathscr{A}$ setzt man häufig auch $\mathscr{A}^X$ für eine Stichprobe X. In diesem Fall fordert man in der Definition der Suffizienz von $\mathscr{H}$ die Eigenschaft $\mathscr{H} \subseteq \mathscr{A}^X$, was im Fall $\mathscr{H} = \mathscr{A}^T$ für eine Zufallsgröße T bedeutet, daß T sich als meßbare Funktion H von X darstellen läßt:

$$T(\omega) = H(X(\omega)).$$

Damit haben wir den im Abschnitt 6.1 betrachteten Fall eingeordnet.

Beispiel 6.5. $(X_1, X_2, \ldots, X_n)$ sei ein Bernoullischema mit dem Parameter $p \in (0,1)$. (Die $X_k,\ k = 1, \ldots, n$ sind unabhängig und es gilt $\mathbb{P}_p(X_k = i) = p^i(1-p)^{1-i},\ i = 0, 1$.) Dann ist $S_n := \sum_{k=1}^n X_k$ eine suffiziente Statistik für $(\mathbb{P}_p,\, p \in (0,1)\,)$ hinsichtlich $\mathscr{A}^X$.
Beweis:
Die σ-Algebra $\mathscr{A}^X$ wird erzeugt durch die Zerlegung

$$\{\omega \mid X(\omega) = (i_1, i_2, \ldots, i_n)\}\,, \quad (i_1, i_2, \ldots, i_n) \in \{0,1\}^n.$$

Es gilt

$$\mathbb{P}_p\big(X = (i_1, i_2, \ldots, i_n) \mid S_n = m\big) = \mathbf{1}_{\{m\}}(i_1 + \ldots + i_n) \frac{p^{\sum_{j=1}^n i_j}(1-p)^{n - \sum_{j=1}^n i_j}}{\mathbb{P}_p(S_n = m)}$$
$$= \frac{\mathbf{1}_{\{m\}}(i_1 + \ldots + i_n)}{\binom{n}{m}}$$

und diese bedingte Wahrscheinlichkeit ist unabhängig von p. □

Beispiel 6.6. Es sei $\big(W(t),\, t \in [0,T]\big)$ ein Wienerscher Prozess mit den Parametern μ und σ^2. Der Parameter σ^2 sei bekannt, μ sei unbekannt. Dann ist $W(T)$ eine suffiziente Statistik für μ bezüglich $\mathscr{A}^T = \sigma\big(W(t),\, t \in [0,T]\big)$.

Beweis:
Ist Z eine Zylindermenge der Form

$$Z = \{\omega \in \Omega \mid \big(W(t_1), \ldots, W(t_n)\big) \in B\}$$

für gewisse $t_1 < t_2 < \ldots < t_n = T$ und ein $B \in \mathscr{B}^n$, so gilt

$$\mathbb{P}_\mu\big(Z \mid W(T) = w\big) = \int_B \cdots \int \varphi_{t_1,\ldots,t_n}(x_1, \ldots, x_{n-1}, w)\, dx_1 \ldots\, dx_{n-1}$$

$$\text{mit} \quad \varphi_{t_1,\ldots,t_n}(x_1, \ldots, x_n) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi(t_k - t_{k-1})\sigma^2}} \frac{\exp\left\{-\frac{1}{2}\sum_{k=1}^{n} \frac{(x_k - x_{k-1} - \mu(t_k - t_{k-1}))^2}{\sigma^2(t_k - t_{k-1})}\right\}}{\exp\left\{-\frac{1}{2}\frac{(w-\mu T)^2}{\sigma^2 T}\frac{1}{\sqrt{2\pi T\sigma^2}}\right\}}.$$

Eine kurze Rechnung zeigt, daß dieser Ausdruck nicht vom Parameter μ abhängt. Wenn $\mathbb{P}_\mu(A \mid W(T) = w)$ für jede Zylindermenge A aus $\mathscr{A}^T$ nicht von μ abhängt, so gilt dasselbe für alle $A \in \mathscr{A}^T$.

6.3 Suffiziente σ-Algebren und Statistiken in dominierten Modellen

Es sei $\mathscr{P} = (\mathbb{P}_\vartheta,\, \vartheta \in \Theta)$ eine durch ein σ-finites Maß μ dominierte Familie von Wahrscheinlichkeitsmaßen auf $(\Omega, \mathscr{A})$. Die Likelihoodfunktion $L(\vartheta\,;\,\omega)$ ist definiert durch

$$L(\vartheta\,;\,\omega) = \frac{d\,\mathbb{P}_\vartheta}{d\mu}(\omega),$$

und ist für jedes $\vartheta \in \Theta$ μ-fast überall eindeutig bestimmt.

Die Suffizienz einer σ-Algebra $\mathscr{H}$ bzw. einer Statistik T (mit Werten in einem meßbaren Raum $(E, \mathscr{E})$) läßt sich an Hand von L feststellen. Das ist der Inhalt des folgenden Satzes. Mit seiner Hilfe werden wir weitere Beispiele suffizienter σ-Algebren und Statistiken gewinnen.
Wir beginnen mit einem Lemma technischer Natur.

Lemma 6.4. *Die Familie $\mathscr{P} = (\mathbb{P}_\vartheta,\, \vartheta \in \Theta)$ sei dominiert durch ein σ-finites Maß μ. Dann existiert eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $(\Omega, \mathscr{A})$, so daß gilt:*

a) $\mathbb{P}(A) = 0 \Longleftrightarrow (\mathbb{P}_\vartheta(A) = 0,\ \ \forall \vartheta \in \Theta)$ *für alle* $A \in \mathscr{A}$.

b) Die Verteilung $\mathbb{P}$ aus a) kann als konvexe Linearkombination $\mathbb{P} = \sum_{k\geq 1} c_k\, \mathbb{P}_{\vartheta_k}$ einer höchstens abzählbar unendlichen Parametermenge $\{\vartheta_k,\ k \geq 1\} \subseteq \Theta$ gewählt werden.

(Für einen Beweis siehe z.B. Dacunha-Castelle, Duflo, Band I.)

Eine Wahrscheinlichkeitsverteilung $\mathbb{P}$ mit den Eigenschaften a) und b) nennt man *ein bezüglich* $\mathscr{P} = (\mathbb{P}_\vartheta,\, \vartheta \in \Theta)$ *privilegiertes Maß*. Offenbar gilt $\mathbb{P} \ll \mu$ und

$$L(\vartheta\,;\,\omega) = \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \cdot \frac{d\,\mathbb{P}}{d\mu}. \tag{6.5}$$

Zur Untersuchung der Abhängigkeit der Likelihoodfunktion L von ϑ genügt es also, sich auf privilegierte dominierende Maße zu beschränken.
Sind $\mathbb{P}$ und $\mathbb{P}'$ zwei privilegierte Wahrscheinlichkeitsmaße bezüglich $\mathscr{P}$, so sind sie äquivalent: $\mathbb{P} \equiv \mathbb{P}'$.

Folgerung 6.5. Ist $\mathbb{P}$ ein privilegiertes Wahrscheinlichkeitsmaß bezüglich $\mathscr{P}$, so gilt für jede nichtnegative Zufallsgröße Z auf $(\Omega, \mathscr{A})$ die Gleichung

$$\mathbb{E}_{\mathbb{P}}(Z) = \sum_{k \geq 1} c_k\, \mathbb{E}_{\vartheta_k}(Z) \tag{6.6}$$

mit der Bezeichnung aus Punkt b) des obigen Lemmas.

Beweis:
Die Gleichung (6.6) ist richtig für $Z = \mathbf{1}_A,\;\; A \in \mathscr{A}$, somit auch für $Z = \sum_{k=1}^m \alpha_k \mathbf{1}_{A_k}$. Nun wendet man die Approximationsmethode an. □

Aussage 6.1 (Charakterisierungssatz). *Folgende Eigenschaften sind äquivalent:*

i) $\mathscr{H}$ *(bzw.* T*) ist suffizient für* $\mathscr{P}$ *hinsichtlich* $\mathscr{A}$

ii) für jedes bezüglich $\mathscr{P}$ *privilegiertes Wahrscheinlichkeitsmaß* $\mathbb{P}$ *und jedes* $\vartheta \in \Theta$ *gibt es eine* $\mathbb{P}$*-Version von* $\dfrac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$*, die bezüglich* $\mathscr{H}$ *(bzw.* $\sigma(T)$*) meßbar ist*[1]

iii) für jedes dominierende σ*-finite Maß* μ *gibt es eine* $\mathscr{A}$*-meßbare Zufallsgröße* h *(unabhängig von* ϑ*) und eine* $\mathscr{H}$*-meßbare Zufallsgröße* G_ϑ *(bzw. eine Borelmeßbare Funktion* g_ϑ *von* $(E, \mathscr{E})$ *in* $(\mathbb{R}, \mathscr{B})$*), so daß gilt:*

$$L(\vartheta\,;\,\omega) = h(\omega) G_\vartheta(\omega) \quad \mu\text{-fast überall}\;, \vartheta \in \Theta.$$
$$\text{(bzw. } L(\vartheta\,;\,\omega) = h(\omega) g_\vartheta(T(\omega)) \quad \mu\text{-fast überall}\;, \vartheta \in \Theta\text{)}^1$$

[1] Man beachte: Eine reellwertige Zufallsgröße Y ist $\sigma(T)$-meßbar genau dann, wenn es eine $(\mathscr{E}$-$\mathscr{B})$-meßbare Funktion ψ von E in $\mathbb{R}$ gibt mit $Y(\omega) = \psi(T(\omega)),\ \omega \in \Omega$.

Beweis:
(i)$\Longrightarrow$(ii):
$\mathscr{H}$ sei suffizient für $\mathscr{P}$ hinsichtlich $\mathscr{A}$ und $\mathbb{P}$ sei ein privilegiertes Maß für $\mathscr{P}$. Wir verwenden die Bezeichnungen des Abschnitts 6.2. Nach Definition gibt es für jedes A eine nicht von ϑ abhängende Zufallsgröße Z_A, die $\mathscr{H}$-meßbar ist und für die gilt

$$Z_A = \mathbb{P}_\vartheta(A \mid \mathscr{H}) \quad \mathbb{P}_\vartheta\text{-f.s.}, \quad \forall \vartheta \in \Theta.$$

Das bedeutet

$$\int_H Z_A \, d\,\mathbb{P}_\vartheta = \mathbb{P}_\vartheta(A \cap H) \quad \text{für alle } H \in \mathscr{H} \text{ und alle } \vartheta \in \Theta.$$

Daraus ergibt sich (siehe (6.6))

$$\int_H Z_A \, d\,\mathbb{P} = \mathbb{P}(A \cap H)$$

und somit $\mathbb{E}_{\mathbb{P}}(\mathbf{1}_A \mid \mathscr{H}) = Z_A \quad \mathbb{P}$-fast sicher.
Sind $\vartheta \in \Theta$ und $A \in \mathscr{A}$, so gilt folglich

$$\begin{aligned}
&\mathbb{P}_\vartheta(A) = \int_\Omega \mathbb{E}_{\mathbb{P}}(\mathbf{1}_A \mid \mathscr{H})\, d\,\mathbb{P}_\vartheta = \int_\Omega \mathbb{E}_{\mathbb{P}}(\mathbf{1}_A \mid \mathscr{H}) \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \cdot d\,\mathbb{P} = \\
&\int_\Omega \mathbb{E}_{\mathbb{P}}\Big(\mathbb{E}_{\mathbb{P}}(\mathbf{1}_A \mid \mathscr{H}) \cdot \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}\Big)\, d\,\mathbb{P} = \int_\Omega \mathbb{E}_{\mathbb{P}}\left(\mathbf{1}_A \mid \mathscr{H}\right) \cdot \mathbb{E}_{\mathbb{P}}\left(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}\right) d\,\mathbb{P} = \\
&\int_\Omega \mathbb{E}_{\mathbb{P}}\Big(\mathbf{1}_A\, \mathbb{E}_{\mathbb{P}}(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}) \mid \mathscr{H}\Big)\, d\,\mathbb{P} = \int_A \mathbb{E}_{\mathbb{P}}\Big(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}\Big)\, d\,\mathbb{P}\,.
\end{aligned} \tag{6.7}$$

Andererseits gilt

$$\mathbb{P}_\vartheta(A) = \int_A \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}\, d\,\mathbb{P}, \quad \forall\, A \in \mathscr{A}\,. \tag{6.8}$$

Somit folgt aus (6.7) und (6.8)

$$\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} = \mathbb{E}_{\mathbb{P}}\left(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}\middle|\mathscr{H}\right) \quad \mathbb{P}\text{-fast überall,}$$

d.h., es gibt eine $\mathscr{H}$-meßbare Version von $\dfrac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$.
(ii)$\Longrightarrow$(iii)
Dieser Schritt ergibt sich aus Formel (6.5).
(ii)$\Longrightarrow$(i)

Mit der Formel für die Umrechnung bedingter Erwartungen aus der 4. Übung ergibt sich für jede nichtnegative Zufallsgröße Y

$$\mathbb{E}_\vartheta(Y \mid \mathscr{H}) = \frac{\mathbb{E}_{\mathbb{P}}\left(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} Y \mid \mathscr{H}\right)}{\mathbb{E}_{\mathbb{P}}\left(\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \mid \mathscr{H}\right)} = \mathbb{E}_{\mathbb{P}}(Y \mid \mathscr{H}) \quad \mathbb{P}\text{-fast sicher für alle } \vartheta \in \Theta. \tag{6.9}$$

(iii)$\Longrightarrow$(ii)
Es sei $\mathbb{P}$ privilegiert, folglich gilt $\mathbb{P} \ll \mu$, und

$$\frac{d\,\mathbb{P}}{d\mu} = h \sum c_k G_{\vartheta_k} = h(\omega) \cdot G, \tag{6.10}$$

G ist nach Konstruktion $\mathscr{H}$-meßbar. Nach Voraussetzung ist

$$\frac{d\,\mathbb{P}_\vartheta}{d\mu} = h(\omega) G_\vartheta(\omega). \tag{6.11}$$

Aus den beiden Gleichungen (6.10) und (6.11) erhalten wir

$$h(\omega)G_\vartheta(\omega) = \frac{d\,\mathbb{P}_\vartheta}{d\mu} = \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \cdot \frac{d\,\mathbb{P}}{d\mu} = \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} \cdot h(\omega)G(\omega) \quad \mu\text{-fast sicher.} \tag{6.12}$$

Wegen

$$\mathbb{P}_\vartheta(h=0) = \int_{\{h=0\}} \frac{d\,\mathbb{P}_\vartheta}{d\mu}\, d\mu = \int_{\{h=0\}} h \cdot G_\vartheta \, d\mu = 0, \quad \vartheta \in \Theta,$$

gilt $\mathbb{P}(h=0) = 0$. Analog ergibt sich $\mathbb{P}(G=0) = 0$ aus (6.10). Somit folgt aus (6.12)

$$\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} = \frac{G_\vartheta(\omega)h(\omega)}{G(\omega)h(\omega)} = \frac{G_\vartheta(\omega)}{G(\omega)} \quad \mathbb{P}\text{-fast sicher,}$$

und $\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$ hat somit eine $\mathscr{H}$-meßbare Version. □

6.4 Minimal suffiziente Statistiken und σ-Algebren

Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistischer Raum mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$. $\mathscr{P}$ werde durch ein σ-finites Maß μ dominiert und $\mathscr{H}$ sei eine suffiziente σ-Algebra für $\mathscr{P}$ hinsichtlich $\mathscr{A}$.
Ist $\mathscr{H}'$ eine Teil-σ-Algebra von $\mathscr{A}$ mit $\mathscr{H} \subseteq \mathscr{H}'$, so ist auch $\mathscr{H}'$ suffizient. Das ergibt sich unmittelbar aus dem Faktorisierungssatz. Da andererseits nicht jede Teil-σ-Algebra $\mathscr{H}''$ von $\mathscr{H}$ suffizient ist (man prüfe das für die triviale σ-Algebra $\mathscr{H}'' = \{\emptyset, \Omega\}$), ergibt sich die Frage, ob es eine minimale suffiziente σ-Algebra gibt und ob diese gegebenenfalls

eindeutig ist.

Es seien $\mathbb{P}$ eine privilegierte dominierende Wahrscheinlichkeitsverteilung für $\mathscr{P}$ und

$$L(\vartheta\,;\,\omega) := \frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}(\omega) \text{ die Likelihoodfunktion.}$$

Die σ-Algebra $\mathscr{H}_0$, definiert durch

$$\mathscr{H}_0 := \sigma\big(L\,(\vartheta\,;\,\cdot),\ \vartheta \in \Theta\big)$$

(kleinste σ-Algebra von Teilmengen von Ω, bezüglich der alle $L\,(\vartheta\,;\,\cdot)\,,\ \vartheta \in \Theta$, meßbar sind)

ist eine suffiziente σ-Algebra für $\mathscr{P}$ bezüglich $\mathscr{A}$. Davon überzeugt man sich leicht durch Anwendung von Formel (6.9). Für Z_A ergibt sich dabei $\mathbb{P}(A \mid \mathscr{H}_0)$.
Die σ-Algebra $\mathscr{H}_0$ hängt davon ab, welche Versionen der Zufallsgrößen $L\,(\vartheta\,;\,\cdot)$ zu ihrer Bildung eingesetzt werden. ($L\,(\vartheta\,;\,\cdot)$ ist ja nach Definition nur $\mathbb{P}$-fast sicher bestimmt.)
Um von diesem Sachverhalt unabhängig zu werden, gehen wir zu Vervollständigung $\mathscr{H}_0^{\mathbb{P}}$ von $\mathscr{H}_0$ bezüglich $\mathbb{P}$ über.
(Bezeichnet man mit $\mathscr{M}$ die Menge aller $A \in \mathscr{A}$ mit $\mathbb{P}(A) = 0$, so ist $\mathscr{M}_0^{\mathbb{P}}$ definiert als $\sigma\big(\mathscr{M}_0 \cup \mathscr{M}\big)$. Es gilt: X ist $\mathscr{M}_0^{\mathbb{P}}$-meßbar genau dann, wenn es eine $\mathscr{M}_0$-meßbare Zufallsgröße $\tilde{X}$ gibt mit $\mathbb{P}(X \neq \tilde{X}) = 0$.)

Lemma 6.6. *Sind $\mathscr{H}_0$ und $\tilde{\mathscr{H}_0}$ zwei σ-Algebren , die durch verschiedene Versionen von $L\,(\vartheta\,;\,\cdot)\,,\ \vartheta \in \Theta$, erzeugt werden, so gilt*

$$\mathscr{H}_0^{\mathbb{P}} = \tilde{\mathscr{H}}_0^{\mathbb{P}}.$$

Beweis:
Jede Version von $L\,(\vartheta\,;\,\cdot)$ ist meßbar bezüglich $\mathscr{H}_0^{\mathbb{P}}$ und auch bezüglich $\tilde{\mathscr{H}}_0^{\mathbb{P}}$. Daraus folgt

$$\mathscr{H}_0 \subseteq \tilde{\mathscr{H}}_0^{\mathbb{P}} \text{ und } \tilde{\mathscr{H}_0} \subseteq \mathscr{H}_0^{\mathbb{P}} \text{ und somit } \mathscr{H}_0^{\mathbb{P}} \subseteq \tilde{\mathscr{H}}_0^{\mathbb{P}},\ \text{sowie } \tilde{\mathscr{H}}_0^{\mathbb{P}} \subseteq \mathscr{H}_0^{\mathbb{P}}.$$

□

Die σ-Algebra $\mathscr{H}_0^{\mathbb{P}}$ ist suffizient (siehe die Bemerkung eingangs dieses Abschnittes) und minimal im Sinne der folgenden Aussage

Aussage 6.2. *Ist $\mathscr{H}$ eine suffiziente σ-Algebra für $\mathscr{P}$ bezüglich $\mathscr{A}$, so gilt $\mathscr{H}_0^{\mathbb{P}} \subseteq \mathscr{H}^{\mathbb{P}}$.*

Beweis:
Ist $\mathscr{H}$ suffizient, so besitzt $\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$ auf Grund des Faktorisierungssatzes eine $\mathscr{H}$-meßbare Version, folglich ist $\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}}$ eine $\mathscr{H}^{\mathbb{P}}$-meßbare Zufallsgröße für alle $\vartheta \in \Theta$. Somit gilt $\mathscr{H}_0 \subseteq \mathscr{H}^{\mathbb{P}}$ und damit $\mathscr{H}_0^{\mathbb{P}} \subseteq \mathscr{H}^{\mathbb{P}}$. □

Diese Aussage führt zu folgender

Definition 6.7. Eine suffiziente σ-Algebra $\mathscr{H}$ heißt *minimal suffizient*, falls $\mathscr{H}^{\mathbb{P}} \subseteq \tilde{\mathscr{H}}^{\mathbb{P}}$ für alle suffizienten σ-Algebren $\tilde{\mathscr{H}}$ gilt.

Die obige Aussage erlaubt nun die

Folgerung 6.8. $\mathscr{H}$ ist minimal suffizient genau dann, wenn $\mathscr{H}^{\mathbb{P}} = \mathscr{H}_0^{\mathbb{P}}$ gilt.

Definition 6.9. Eine Zufallsgröße H auf $(\Omega, \mathscr{A})$ heißt *minimal suffizient*, wenn $\sigma(H)$ minimal suffizient ist.

Das bedeutet, daß es zu jedem $\vartheta \in \Theta$ eine $\mathbb{P}$-Version von $L(\vartheta\,;\,\cdot)$ gibt mit

$$\sigma(H) = \sigma\left(L(\vartheta\,;\,\cdot) \mid \vartheta \in \Theta\right).$$

Kurz ausgedrückt heißt das: Ist die Likelihoodfunktion für jedes $\vartheta \in \Theta$ in Abhängigkeit von ω die Funktion einer Statistik $H = H(\omega)$, so ist H minimal suffizient.

Beispiel 6.7. $X = (X_1, \ldots, X_n)$ sei eine mathematische Stichprobe aus einer gleichmäßig auf $[a, b]$ verteilten Grundgesamtheit, a und b seien unbekannt. Wir setzen

$$\Theta = \{\vartheta = (a, b) \,:\, -\infty < a < b < \infty\}.$$

Als dominierendes Maß für $\mathscr{P}^X$ verwenden wir das Lebesguemaß λ_n auf $(\mathbb{R}^n, \mathscr{B}^n)$. Dann gilt

$$L^X(\vartheta\,;\,X(\omega)) = (b-a)^{-n} \prod_{k=1}^{n} \mathbf{1}_{[a,b]}(X_k) = (b-a)^{-n} \mathbf{1}_{[a,b]}(\min_{k=1,\ldots,n} X_k) \cdot \mathbf{1}_{[a,b]}(\max_{k=1,\ldots,n} X_k).$$

Folglich ist die Stichprobenfunktion $T(X) = \left(\min_{k=1,\ldots,n} X_k, \max_{k=1,\ldots,n} X_k\right)$ eine suffiziente Statistik für $\mathscr{P}\big|_{\mathscr{A}^X}$ bezüglich $\mathscr{A}^X$, sie ist sogar minimal suffizient.

Beispiel 6.8. Es sei $X = (X_0, X_1, \ldots, X_n)$ eine Markovsche Kette mit dem Zustandsraum $E = \{e_1, e_2, \ldots, e_m\}$, der Anfangsverteilung ρ und den Übergangswahrscheinlichkeiten $\pi_{ij},\ i, j \in E$. Als unbekannte Parameter werden die π_{ij} angesehen: $\Theta := \left\{\vartheta = (\pi_{ij})_{i,j\in E}\right\}$. Die Anfangsverteilung sei bekannt.

Dann gilt

$$L_n(\vartheta\,;\,X) = \rho(X_0) \cdot \pi_{X_0X_1} \cdot \pi_{X_1X_2} \cdot \ldots \cdot \pi_{X_{n-1}X_n} = \rho(X_0) \prod_{i,j\in E} \pi_{ij}^{N_n^{ij}}$$

mit

$$N_n^{ij} = \sum_{l=0}^{n-1} \mathbf{1}_{\{X_l=i, X_{l+1}=j\}}.$$

Folglich ist $(N_n^{ij},\, i,j \in E)$ eine minimal suffiziente Statistik für $\big(\mathbb{P}_\vartheta^X,\, \vartheta \in \Theta\big)$ bezüglich $\mathscr{A}^X$.

Beispiel 6.9. Es sei $(X(t),\, t \geq 0)$ ein Ornstein-Uhlenbeck-Prozess, d.h. eine Lösung der Differentialgleichung

$$\begin{aligned} &dX(t) = \rho X(t)\,dt + \sigma\,dW(t), \quad t \geq 0 \\ &X(0) = X_0. && \text{(Anfangsbedingung)} \end{aligned}$$

Dann gilt

$$\frac{d\,\mathbb{P}_\rho^T}{d\,\mathbb{P}_0^T}(\omega) = \exp\left\{\rho \int_0^T X(t)\,dX(t) - \frac{\rho^2}{2}\int_0^T X^2(t)\,dt\right\}.$$

Folglich ist

$$H\big((X(t)\,,\ 0 \leq t \leq T)\big) = \left(\int_0^T X(t)\,dX(t),\ \int_0^T X^2(t)\,dt\right)$$

eine minimal suffiziente Statistik für ρ.

Kapitel 7

Exponentialfamilien

Exponentialfamilien sind dominierte statistische Räume, deren Likelihoodfunktion eine besonders einfache Struktur besitzt, ihr Logarithmus ist von affiner Gestalt. Neben der daraus resultierenden sehr guten analytischen Handhabbarkeit zeichnen sie sich durch zahlreiche statistische Eigenschaften aus. Hinzu kommt, daß viele der gängigen klassischen statistischen Räume Exponentialfamilien bilden.
In diesem Kapitel untersuchen wir Exponentialfamilien von Verteilungen und von Lévy-Prozessen. Ausführliche und tiefer gehende Ergebnisse findet man in Küchler, Sørensen, (1997).

7.1 Wahrscheinlichkeitsverteilungen

Es sei μ ein σ-finites Maß auf $(\mathbb{R}^k, \mathscr{B}^k)$. Wir definieren

$$I := \left\{ u \in \mathbb{R}^k \ : \ \int_{\mathbb{R}^k} \exp\left(< u, x >\right) \mu\left(\, dx\right) < \infty \right\}$$

und setzen voraus, daß I einen nichtleeren offenen Kern enthält: $\mathring{I} \neq \emptyset$. Für jedes $u \in I$ definieren wir ein Wahrscheinlichkeitsmaß $\mathbb{P}^{(u)}$ auf $\mathscr{B}^k$ durch

$$\mathbb{P}^{(u)}(A) = \int_A \exp\bigl\{< u, x > -\psi(u)\bigr\}\mu(\, dx), \quad A \in \mathscr{B}^k$$

mit

$$\psi(u) := \ln \int_{\mathbb{R}^k} \exp\bigl\{< u, x >\bigr\}\mu\left(\, dx\right), \quad u \in I.$$

Definition 7.1. Für jede Teilmenge J von I mit mehr als einem Element nennt man $(\mathbb{P}^{(u)},\ u \in J)$ eine *Exponentialfamilie von Wahrscheinlichkeitsmaßen auf* $\mathbb{R}^k$*, erzeugt von* μ. Der Parameter u heißt *kanonischer Parameter* der Exponentialfamilie.

Exponentialfamilien bilden statistische Räume und jedes $\mathbb{P}^{(u_0)}$ aus $\mathscr{P} = (\mathbb{P}^{(u)},\ u \in I)$ mit $u_0 \in I$ ist ein dominierendes privilegiertes Wahrscheinlichkeitsmaß für $\mathscr{P}$.

Beispiel 7.1.

$$\mu = \mathcal{N}(0,1), \quad d\,\mathbb{P}^{(u)} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2} + ux - \frac{u^2}{2}\right\} dx = \exp\left\{ux - \frac{u^2}{2}\right\} d\mu,$$

$$\psi(u) = \frac{u^2}{2}, \quad u \in I = \mathbb{R}. \quad \mathbb{P}^{(u)} \text{ ist eine Normalverteilung mit dem Erwartungswert } u.$$

Beispiel 7.2.

$$\mu(\{k\}) = \frac{\lambda_0^k}{k!} e^{-\lambda_0}, \qquad \mathbb{P}^{(u)}(\{k\}) = \exp\{uk - \psi(u)\}\, \mu(\{k\}),\ u \in I = \mathbb{R},$$

mit $\psi(u) = \lambda_0(e^u - 1)$. $\mathbb{P}^{(u)}$ ist eine Poissonverteilung mit dem Parameter $\lambda_0 e^u$.

Beispiel 7.3.

$$\mu = \Gamma(\alpha_0, \lambda_0), \quad \alpha_0, \lambda_0 > 0, \ \text{ d.h. } \ \mu(\,dx) = \frac{1}{\Gamma(\alpha_0)} x^{\alpha_0 - 1} \lambda_0^{\alpha_0} e^{-\lambda_0 x} \mathbf{1}_{(0,\infty)}(x)\, dx, \quad x \in \mathbb{R}.$$

$$\mathbb{P}^{(u)}(\,dx) = \frac{1}{\Gamma(\alpha_0)} x^{\alpha_0 - 1} \lambda_0^{\alpha_0} e^{-(\lambda_0 - u)x} e^{-\psi(u)} \mathbf{1}_{(0,\infty)}(x)\, dx$$

mit $\psi(u) = \alpha_0 \ln\left(\dfrac{\lambda_0}{\lambda_0 - u}\right)$ und $u \in I = (-\infty, \lambda_0)$.

$\mathbb{P}^{(u)}$ ist eine $\Gamma(\alpha_0, \lambda_0 - u)$-Verteilung.

Ist γ eine Abbildung von $\Theta \subseteq \mathbb{R}^m$ in I, so nennt man auch $\left(\mathbb{P}_\vartheta := \mathbb{P}^{(\gamma(\vartheta))},\ \vartheta \in \Theta\right)$ eine Exponentialfamilie (sofern $\gamma(\Theta)$ mehrelementig ist).

Die folgende Aussage ist eine mehrdimensionale Verallgemeinerung der Aussagen, die wir in der ersten Übung kennengelernt haben.

Aussage 7.1. *Die Menge I ist konvex. Wenn $u \in \mathring{I}$ gilt, so ist $\psi(\cdot)$ in u unendlich oft differenzierbar, und es gilt für alle $i_j \geq 0$, $j = 1, 2, \ldots, k$:*

$$\int |\, x_1^{i_1} x_2^{i_2} \cdots x_k^{i_k} \,|\, e^{<u,x>} \mu(\,dx) < \infty$$

$$\frac{\partial^{i_1 + \ldots + i_k}}{\partial u_1^{i_1} \ldots \partial u_k^{i_k}} \exp\{\psi(u)\} = \int x_1^{i_1} x_2^{i_2} \cdots x_k^{i_k} e^{<u,x>} \mu(\,dx).$$

Für einen Beweis siehe Dacunha-Castelle, Duflo, Band I.

Es seien $H = (H_1, H_2, \ldots, H_k)$ eine $\mathbb{R}^k$-wertige Zufallsgröße über $(\Omega, \mathscr{A})$ und ν ein σ-finites Maß auf $(\Omega, \mathscr{A})$.
Dann ist durch

$$\mathbb{P}^{(u)}(A) = \int_A \exp\{< u, H(\omega) > -\psi(u)\}\,\nu(\,d\omega)$$

mit

$$\psi(u) = \ln \int_\Omega \exp\{< u, H(\omega) >\}\,\nu(\,d\omega), \quad u \in I$$

und

$$I := \Big\{ v \in \mathbb{R}^k \;:\; \int_\Omega \exp\big\{< u, H(\omega) >\big\}\nu(\,d\omega) < \infty \Big\}$$

auf $(\Omega, \mathscr{A})$ eine Familie $\mathscr{P}_H$ von Wahrscheinlichkeitsmaßen $\mathbb{P}^{(u)}$, $u \in I$ gegeben, die wir *die von H und ν erzeugte Exponentialfamilie* nennen. Die Zufallsgröße H heißt *kanonische Zufallsgröße für $\mathscr{P}_H$.*
Die Anwendung der vorangegangenen Aussage auf $\nu^H := \nu \circ H^{-1}$ liefert:

Aussage 7.2. 1. *I ist konvex und $\psi(\cdot)$ ist im Inneren von I unendlich oft differenzierbar,*

2. *Wenn $u \in \overset{\circ}{I}$, so sind alle Momente $\int_\Omega H_1^{i_1} H_2^{i_2} \cdot \ldots \cdot H_k^{i_k} \exp\big\{< u, H > -\psi(u)\big\}\mu(\,d\omega)$ endlich, und es gilt*

$$\operatorname{grad}\psi(u) \cdot \exp\{\psi(u)\} = \int H \exp\{< u, H >\}\, d\mu, \quad \textit{also} \quad \mathbb{E}^{(u)}(H) = \operatorname{grad}\psi(u),$$

3. *Für $u \in \overset{\circ}{I}$ haben wir $Kov(H_i, H_j) = \Big(\dfrac{\partial^2\psi}{\partial u_i \partial u_j}\Big)(u)$.*

Aussage 7.3. *Die Stichprobenfunktion H ist eine minimal suffiziente Statistik für $\mathscr{P}_H$ bezüglich $\mathscr{A}$.*

Beweis:
Wir erinnern zunächst an die Bezeichnung $\mathscr{M}_0 := \sigma\big(L(\vartheta\,;\,\cdot) \mid \vartheta \in \Theta\big)$. Es gilt $\mathscr{M}_0 \subseteq \sigma(H)$, da $e^{<u,H>}$ bezüglich $\sigma(H)$ meßbar für alle $u \in \mathbb{R}^k$ ist.

Andererseits, ist $u \in \overset{\circ}{I}$ und v irgend ein Element von $\mathbb{R}^k$, so gilt

$$< v, H >= \lim_{\substack{h\to 0\\ h\in\mathbb{R}}} \frac{\exp\{< u+vh, H >\}\exp\{< u, H >\}}{h\exp\{< u, H >\}}.$$

Also ist für jedes $v \in \mathbb{R}^k$ die Zufallsgröße $< v, H >$ bezüglich $\mathscr{M}_0$ meßbar, und somit auch H selbst, d.h., es gilt $\sigma(H) \subseteq \mathscr{M}_0$. □

$(\Omega, \mathscr{A}, \mathscr{P})$ bilde eine Exponentialfamilie mit der kanonischen Zufallsgröße H.
Das zufällige Experiment werde n-mal mit demselben wahren Parameter $\vartheta \in \Theta$ unabhängig voneinander durchgeführt. $(\Omega_k, \mathscr{A}_k, \mathbb{P}_\vartheta)$ sei der Wahrscheinlichkeitsraum für das k-te Experiment, ω_k sei das Ergebnis des k-ten Experiments, A_k Ereignisse, die mit dem k-ten Experiment zusammenhängen. Dann wird das Gesamtexperiment, das aus den n Einzelexperimenten besteht, beschrieben durch:

$$\tilde{\Omega} := \prod_{j=1}^n \Omega_j, \quad \tilde{\mathscr{A}} := \bigotimes_{j=1}^n \mathscr{A}_j, \quad \mathscr{P}^{\otimes n} := \left(\mathbb{P}_\vartheta^{\otimes n} \,:\, \vartheta \in \Theta\right)$$

mit

$$\mathbb{P}_\vartheta^{\otimes n} := \mathbb{P}_\vartheta \otimes \cdots \otimes \mathbb{P}_\vartheta, \quad \omega = (\omega_1, \ldots, \omega_n) \in \tilde{\Omega}.$$

Für die diesem Gesamtexperiment entsprechende Wahrscheinlichkeitsverteilung gilt

$$\begin{aligned}
&\mathbb{P}_\vartheta \otimes \cdots \otimes \mathbb{P}_\vartheta(A_1 \times \ldots \times A_n) = \\
&= \prod_{j=1}^n \mathbb{P}_\vartheta(A_j) = \prod_{j=1}^n \int_{A_j} \exp\Big\{< \gamma(\vartheta), H(\omega_j) > -\psi(\vartheta)\Big\}\, d\mu(\omega_j) \\
&\qquad = \int\cdots\int_{A_1\times\ldots\times A_n} \exp\left\{< \gamma(\vartheta), \sum_{j=1}^n H(\omega_j) > -n\psi(\vartheta)\right\} d\mu^{\otimes n}(\omega).
\end{aligned}$$

$$\mathbb{P}^{\otimes n}(A) = \int_A \exp\left\{< \gamma(\vartheta), \sum_{j=1}^n H(\omega_j) > -n\psi(\vartheta)\right\} d\mu^{\otimes n}(\omega), \quad A \in \mathscr{A}.$$

Folgerung 7.2. Ist $X = X = (X_1, \ldots, X_n)$ eine mathematische Stichprobe und bildet $(\mathbb{P}_\vartheta^{X_1})$ eine Exponentialfamilie mit der kanonischen Variablen $H(x)$, so ist $H(X_1) + \ldots + H(X_n)$ eine minimal suffiziente Statistik für $\mathscr{P}^X$.

Beweis:
Es gilt

$$L^X(\vartheta; x) = \prod_{k=1}^n \exp\Big\{< \gamma(\vartheta), H(x_k) > -\psi(\vartheta)\Big\} = \frac{d\,\mathbb{P}_\vartheta^X}{d\mu^{\otimes n}}(x), \quad (L(\vartheta; \omega) = L^X(\vartheta; X(\omega)))$$

d.h., $\sum_{j=1}^{n} H(X_j)$ ist die kanonische Zufallsgröße der Produkt-Exponentialfamilie $(\mathbb{P}_{\vartheta}^{X},\ \vartheta \in \Theta)$. $\square$

7.2 Lévy-Prozesse

Es seien $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(X(t)\ t \geq 0)$ ein Lévy-Prozess über $(\Omega, \mathscr{A}, \mathbb{P})$. Darunter verstehen wir einen stochastisch stetigen reellwertigen zufälligen Prozess mit unabhängigen stationären Zuwächsen, der $\mathbb{P}$-fast sicher bei Null startet und dessen Trajektorien die càdlàg-Eigenschaft haben. Beispiele sind der Wienersche, der Poissonsche, der Gamma-, und jeder zusammengesetzte Poissonsche Prozess.

Für jedes $t > 0$ sei F_t die Verteilungsfunktion von $X(t)$ bezüglich $\mathbb{P}$:

$$F_t(x) = \mathbb{P}(X(t) \leq x), \quad x \in \mathbb{R},\ t > 0.$$

Lemma 7.3.

$$\textit{Ist}\ \ \mathbb{E}_{\mathbb{P}}\Big(\exp\{uX(t)\}\Big) < \infty\ \ \textit{für ein}\ \ u \in \mathbb{R}\ \ \ \textit{und ein}\ t > 0,$$
$$\textit{so gilt}\ \ \mathbb{E}_{\mathbb{P}}\Big(\exp\{uX(s)\}\Big) < \infty\ \ \textit{für alle}\ \ s > 0.$$

Beweis:
Es sei zunächst $s \in (0, t)$. Damit ist $X(t)$ nach Voraussetzung die Summe der zwei unabhängigen Zufallsgrößen $X(s)$ und $X(t) - X(s)$, die die Verteilungsfunktionen F_s bzw. F_{t-s} besitzen. Folglich gilt

$$F_t = F_s \star F_{t-s}. \tag{a}$$

Diese Gleichung schreiben wir in der Form

$$F_t(z) = \int_{\mathbb{R}} F_s(z-y) F_{t-s}(\,dy) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{(-\infty, z-y]}(x) F_s(\,dx) \right) F_{t-s}(\,dy) =$$
$$\int_{\mathbb{R}^2} \mathbf{1}_{(-\infty, z]}(x+y) F_s(\,dx) \otimes F_{t-s}(\,dy), \quad z \in \mathbb{R}\,.$$

(Wir haben die Folgerung von Tonelli-Hobson zum Satz von Fubini benutzt.)
Daraus ergibt sich mit der üblichen Approximationsmethode für alle nichtnegativen meßbaren Funktionen f auf $\mathbb{R}$ die Gleichung

$$\int_{\mathbb{R}} f(z) F_t(\,dz) = \int_{\mathbb{R}^2} f(x+y) F_s(\,dx) \otimes F_{t-s}(\,dy), \tag{b}$$

wobei beide Seiten gleichzeitig endlich oder beide unendlich sind. Setzt man $f(z) = \exp\{uz\}$, so ist nach Voraussetzung

$$\int_{\mathbb{R}} \exp\{uz\} F_t(\,dz) < \infty$$

und aus (b) und dem Satz von Fubini folgt

$$\int_{\mathbb{R}} \exp\{ux\} F_s(\,dx) < \infty.$$

Ist dagegen $s > t$, so wählt man $m \geq 1$ derart, daß $mt > s$ gilt. Nun ergibt sich mit analogen Mitteln und aus der genannten Folgerung des Satzes von Fubini

$$\left(\int_{\mathbb{R}} e^{ux} F_t(\,dx)\right)^m = \int_{\mathbb{R}} e^{u(x_1+\ldots+x_m)} \prod_{j=1}^{m} F_t(\,dx_j) = \int_{\mathbb{R}} e^{uz} F_{mt}(\,dz) < \infty.$$

Daraus folgt, wie bereits gezeigt,

$$\int_{\mathbb{R}} e^{sx} F_s(\,dx) < \infty.$$

□

Das Lemma erlaubt folgende

Definition 7.4. Das Intervall I werde definiert durch

$$I = \left\{ u \in \mathbb{R} \,:\, \int_{\mathbb{R}} \exp\{uX(t)\}\, d\,\mathbb{P} = \int_{\mathbb{R}} \exp\{ux\} F_t(\,dx) < \infty \right\}.$$

Dabei ist $t > 0$ beliebig gewählt, I ist unabhängig von t. Es sei $u \in I$. Dann ist die Funktion $\psi_t(u)$, definiert durch

$$\psi_t(u) = \ln \int_{\mathbb{R}} \exp\{ux\} F_t(\,dx), \quad t > 0$$

endlich. Außerdem gilt

$$\begin{aligned} \psi_{s+t}(u) &= \ln \int_{\mathbb{R}} \exp\{uz\} F_{s+t}(\,dz) = \ln \int_{\mathbb{R}} \exp\{u(x+y)\} F_s(\,dx) \otimes F_t(\,dy) = \\ &\ln \left(\int_{\mathbb{R}} \exp\{ux\} F_s(\,dx) \cdot \int_{\mathbb{R}} \exp\{uy\} F_t(\,dy) \right) = \psi_s(u) + \psi_t(u), \quad s,t > 0. \end{aligned}$$

Wegen der stochastischen Stetigkeit von $X(\cdot)$ ist auch $t \longrightarrow \psi_t(u)$ stetig, und es folgt

$$\psi_t(u) = t\psi_1(u), \quad \text{wir setzen } \ \psi(u) := \psi_1(u).$$

Lemma 7.5. *Für jedes $u \in I$ ist der Prozess $(M_t^{(u)},\ t \geq 0)$, definiert durch*

$$M_t^{(n)} := \exp\{uX(t) - t\psi(u)\}, \quad t \geq 0$$

ein Martingal bezüglich der Filtration $\left(\mathscr{A}_t^X,\ t \geq 0\right)$ mit $\mathscr{A}_t^X := \sigma\left(X(s),\ s \leq t\right), \quad t \geq 0$.

Beweis:
Es gilt

$$\begin{aligned}
\mathbb{E}\left(M_{t+s}^{(u)} \mid \mathscr{A}_s^X\right) &= \mathbb{E}\Big(\exp\big\{u(X_{t+s} - X_s) + uX_s\big\}\big|\mathscr{A}_s^X\Big) \cdot e^{-(t+s)\psi(u)} \\
&= \exp\big\{uX_s - s\psi(u)\big\}\, \mathbb{E}\Big(\exp\big\{u(X_{t+s} - X_s) - t\psi(u)\big\}\big|\mathscr{A}_s^X\Big) \\
&= M_s^{(u)} \cdot \mathbb{E}\big(M_t^{(u)}\big) = M_s^{(u)}.
\end{aligned}$$

$\square$

Wir definieren durch

$$\mathbb{P}_t^{(u)}(A) := \int_A M_t^{(u)}\, d\,\mathbb{P}, \quad A \in \mathscr{A}_t^X$$

ein Wahrscheinlichkeitsmaß $\mathbb{P}_t^{(u)}$ auf $\mathscr{A}_t^X$. Die Familie $\left(\mathbb{P}_t^{(u)},\ t > 0\right)$ ist verträglich in dem Sinne, daß gilt:

Ist $A \in \mathscr{A}_s^X$ und $s < t$, so ist $A \in \mathscr{A}_t^X$ mit $\mathbb{P}_s^{(u)}(A) = \mathbb{P}_t^{(u)}(A)$.

Das ergibt sich aus der Martingaleigenschaft von $\big(M_t^{(u)}\big)$:

$$\mathbb{P}_s^{(u)}(A) = \int_A M_s^{(u)}\, d\,\mathbb{P} = \int_A M_t^{(u)}\, d\,\mathbb{P} = \mathbb{P}_t^{(u)}(A).$$

Somit ist eine Mengenfunktion $\mathbb{P}^{(u)}$ auf $\bigcup_{t>0} \mathscr{A}_t^X$ definiert:

$$\mathbb{P}^{(u)}(A) = \mathbb{P}_t^{(u)}(A), \quad \text{falls } A \in \mathscr{A}_t^X.$$

Sie ist σ-additiv und somit eindeutig erweiterbar zu einer Wahrscheinlichkeitsverteilung $\mathbb{P}^{(u)}$ auf

$$\mathscr{A}^X := \bigvee_{t>0} \mathscr{A}_t^X = \sigma\big(X(t),\ t \geq 0\big).$$

Die eindeutige Fortsetzung von $\mathbb{P}^{(u)}$ auf $\mathscr{A}^X$ werde ebenfalls mit $\mathbb{P}^{(u)}$ bezeichnet.

Aussage 7.4. *Die Wahrscheinlichkeitsmaße* $\mathbb{P}^{(u)}$, $u \in I$, *sind lokal absolutstetig, d.h.* $\mathbb{P}_t^{(u)} \ll \mathbb{P}_t := \mathbb{P}\big|_{\mathscr{A}_t^X}$, $\forall t > 0$, *und es gilt nach Definition*

$$\frac{d\,\mathbb{P}_t^{(u)}}{d\,\mathbb{P}_t} = \exp\{uX(t) - \psi(u)t\}, \quad t > 0,\ u \in I.$$

Die Familie ($\mathbb{P}^{(u)}$, $u \in I$) heißt *die zu* $\mathbb{P}$ *und* $(X(t),\ t \geq 0)$ *gehörende Exponentialfamilie (von Wahrscheinlichkeitsmaßen auf* $\mathscr{A}^X$*).*

Aussage 7.5. *Für jedes* $u \in I$ *ist* $(X(t), t \geq 0)$ *bezüglich* $\mathbb{P}^{(u)}$ *ein Lévy-Prozess. Gilt* $u \in \overset{\circ}{I}$, *so haben wir*

$$\mathbb{E}^{(u)}\big(X(t)\big) = \psi'(u) \cdot t, \qquad Var^{(u)}\big(X(t)\big) = \psi''(u) \cdot t.$$

Beweis:
Sind X und Y zwei n-dimensionale zufällige Vektoren über $(\Omega, \mathscr{A}, \mathbb{P})$ und ist A eine $n \times n$-Matrix mit $X = A \cdot Y$, so gilt offenbar

$$\mathbb{E}\big(f(X)\big) = \mathbb{E}\big(f(A \cdot Y)\big), \quad \text{also}$$
$$\int_{\mathbb{R}^n} f(x)\,\mathbb{P}^X(\,dx) = \int_{\mathbb{R}^n} f(Ay)\,\mathbb{P}^Y(\,dy).$$

Ist nun $X = (X_{t_1}, X_{t_2}, \dots, X_{t_n})$ und $Y = (X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})$, so haben wir mit

$$A = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

für jede beschränkte meßbare Funktion f

$$\mathbb{E}^{(u)}\big(f(X)\big) = \mathbb{E}^{(u)}\big(f(A \cdot Y)\big),$$

also

$$\int_\Omega \exp\Big\{uX(t_n) - t_n\psi(u)\Big\} f(X)\,d\,\mathbb{P} = \int_\Omega \exp\Big\{uX(t_n) - t_n\psi(u)\Big\} f(A \cdot Y)\,d\,\mathbb{P}$$

und somit wegen

$$X_{t_n} = X_{t_1} + (X_{t_2} - X_{t_1}) + \ldots + (X_{t_n} - X_{t_{n-1}})$$
$$\int_{\mathbb{R}^n} e^{ux_n - t_n\psi(u)} f(x_1, x_2, \dots, x_n)\,\mathbb{P}^X(\,dx_1, \dots, \,dx_n)$$
$$= \int_{\mathbb{R}^n} f(y_1, y_1 + y_2, \dots, y_1 + y_2 + \cdots + y_n) e^{u(y_1+y_2+\cdots+y_n) - \psi(u)t_n} \prod_{k=1}^{n}{}^{\otimes} F_{t_k - t_{k-1}}(\,dy_k)$$
$$= \int_{\mathbb{R}^n} f(y_1, y_1 + y_2, \dots, y_1 + y_2 + \cdots + y_n) \prod_{k=1}^{n}{}^{\otimes} F^{(u)}_{t_k - t_{k-1}}(\,dy_k)$$

mit

$$F_t^{(u)}(\,dx) = \exp\bigl\{ux - t\psi(u)\bigr\}F_t(\,dx).$$

Da f beliebig ist, folgt daraus, daß für jedes $u \in I$ der Prozess $\bigl(X(t),\ t \geq 0\bigr)$ bezüglich $\mathbb{P}^{(u)}$ unabhängige und stationäre Zuwächse besitzt. Das bedeutet, die eindimensionalen Verteilungen $\bigl(F_t^{(u)} \mid u \in I\bigr)$ bilden selbst wieder Exponentialfamilien. □

Folgerung 7.6. X_t ist eine minimal suffiziente Statistik für $\mathscr{P} = (\mathbb{P}^{(u)},\ u \in I)$ hinsichtlich $\mathscr{A}_t^X$. Außerdem bildet $\dfrac{X_t}{t}$ eine erwartungstreue Schätzung für $\psi'(u)$. Für ihre Streuung gilt

$$\operatorname{Var}\left(\frac{X_t}{t}\right) = \frac{\psi''(u)}{t}.$$

7.3 Vollständige Statistiken

Es seien $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$ und T eine Statistik über $(\Omega, \mathscr{A})$ mit Werten in einem meßbaren Raum $(E, \mathscr{E})$.

Definition 7.7. Die Statistik T heißt *vollständig (b-vollständig)*, falls jede reellwertige Borelmeßbare Funktion Φ auf $(E, \mathscr{E})$, die $\mathbb{P}_\vartheta^T$-integrierbar ist (bzw. die beschränkt ist), und für die gilt

$$\mathbb{E}_\vartheta\left(\Phi(T)\right) = 0, \quad \text{für alle } \ \vartheta \in \Theta,$$

notwendig Null ist ($\mathbb{P}_\vartheta^T$-f.s., für alle $\vartheta \in \Theta$).

Aussage 7.6. *Es sei U ein erwartungstreuer Schätzer für $\gamma(\vartheta)$ mit $\mathbb{E}_\vartheta(U^2) < \infty$, $\vartheta \in \Theta$. Weiterhin sei T eine vollständige und suffiziente Statistik für ϑ bezüglich $\mathscr{A}$. Dann ist $\mathbb{E}\,(U \mid T)$ ein erwartungstreuer Schätzer für $\gamma(\vartheta)$ mit kleinerer Varianz:*

$$Var\bigl(\mathbb{E}\,(U \mid T)\bigr) \leq Var(U). \tag{7.1}$$

$\mathbb{E}\,(U \mid T)$ ist eine Funktion von T und zwar die einzige Funktion $h(T)$, die einen erwartungstreuen Schätzer für $\gamma(\vartheta)$ darstellt. Folglich ist $\mathbb{E}\,(U \mid T)$ der Schätzer mit minimaler Varianz unter allen erwartungstreuen Schätzern für $\gamma(\vartheta)$: ein MVUE, minimum variance unbiased estimator.

Beweis: Die Ungleichung (7.1) ist eine Eigenschaft bedingter Erwartungen. Ist S ein erwartungstreuer Schätzer für $\gamma(\vartheta)$ und ist S eine Funktion von T, d.h. gilt $S = h(T)$, so haben wir

$$\mathbb{E}_\vartheta\big(h(T)\big) = \mathbb{E}_\vartheta\big(\mathbb{E}(U \mid T)\big) \quad \text{für alle } \vartheta \in \Theta.$$

Aus der Vollständigkeit von T folgt $h(T) = \mathbb{E}(U \mid T)$. □

Beispiel 7.4. Ist $\mathscr{P}$ eine Exponentialfamilie mit der kanonischen Zufallsgröße H, so ist H minimal suffizient und vollständig.
Beweis:
Es gilt

$$0 = \mathbb{E}_\vartheta\left(f(H)\right) = \int_{\mathbb{R}} f(x)\exp\big\{< x, u > -\varphi(u)\big\}\mu^H(\,dx) =$$
$$\exp\big\{-\psi(u)\big\}\int_{\mathbb{R}} f(x)\exp\big\{< x, u >\big\}\mu^H(\,dx).$$

Das bedeutet, die Laplace-Transformierten von f_+ und f_- sind gleich auf Θ. Daraus ergibt sich $f_+ = f_- = 0 \quad \mu^H$-fast überall und somit $f(H) = 0 \quad \mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$. (Wir haben hier benutzt, daß $\gamma(\Theta)$ in I ein nichtleeres Inneres besitzt.) □

Beispiel 7.5. Es seien $(\Omega, \mathscr{A}, \mathscr{P})$ ein dominiertes statistisches Modell, T eine Statistik über $(\Omega, \mathscr{A})$. Ist T suffizient und b-vollständig, so ist T minimal suffizient.
Beweis:
Die σ-Algebra $\sigma(T)$, vervollständigt bezüglich $\mathbb{P}$ ($\mathbb{P}$ ist ein privilegiertes dominierendes Maß), enthält die minimal suffiziente σ-Algebra $\mathscr{M}_0$. Umgekehrt:

$$\text{Ist } C \in \sigma(T), \quad \text{so ist} \quad \mathbb{E}_\vartheta\left(I_C - \mathbb{E}\left(\mathbf{1}_C \mid \mathscr{M}_0\right)\right) = 0 \quad \text{für alle } \vartheta \in \Theta.$$

Damit gilt $\mathbf{1}_C = \mathbb{E}\left(\mathbf{1}_C \mid \mathscr{M}_0\right) \quad \mathbb{P}_\vartheta$-fast sicher für alle $\vartheta \in \Theta$ wegen der Vollständigkeit von T. Somit hat $\mathbf{1}_C$ eine $\mathbb{P}$-Version, die $\mathscr{M}_0$-meßbar ist, d.h. $C \in \mathscr{M}_0^{\mathbb{P}}$. □

Aus der Minimalsuffizienz folgt nicht die b-Vollständigkeit (siehe 6. Übung).

7.4 Die Cramer-Rao-Ungleichung und Exponentialfamilien

Es sei $(\Omega, \mathscr{A}, \mathscr{P})$ ein statistisches Modell mit $\mathscr{P} = (\mathbb{P}_\vartheta,\ \vartheta \in \Theta)$, $\Theta \subseteq \mathbb{R}^k$ und einem dominierenden σ-finiten Maß μ. Wir können o.B.d.A. annehmen, daß μ ein Wahrscheinlichkeitsmaß ist und bezeichnen es deshalb der Gewohnheit gemäß mit $\mathbb{P}$: $\mu = \mathbb{P}$.
Die Likelihoodfunktion $L(\vartheta\,;\,\omega)$ $(\vartheta \in \Theta, \omega \in \Omega)$ ist definiert durch

$$L(\vartheta\,;\,\omega) := \frac{d\,\mathbb{P}_\vartheta}{\mathbb{P}}.$$

Wegen $\mathbb{P}_\vartheta\big(L(\vartheta\,;\,\cdot) > 0\big) = 1$ ist für jedes $\vartheta \in \Theta$

$$l(\vartheta\,;\,\cdot) := \ln L(\vartheta\,;\,\cdot)$$

eine $\mathbb{P}_\vartheta$-fast sicher definierte Zufallsgröße über $(\Omega, \mathscr{A})$.

Voraussetzung: Die Funktion $\vartheta \longrightarrow L(\vartheta\,;\,\omega)$ sei differenzierbar für $\mathbb{P}$-fast alle $\omega \in \Omega$ und es gelte

$$\mathbb{E}_\vartheta\left(\|\operatorname{grad}_\vartheta l(\vartheta\,;\,\cdot)\|^2\right) < \infty, \quad \text{für alle } \vartheta \in \Theta. \ (\|\cdot\| \text{ ist die Euklidische Norm in } \mathbb{R}^k)$$

Definition 7.8. Die Matrix $I(\vartheta)$, definiert durch

$$I(\vartheta) := \mathbb{E}_\vartheta\Big(\operatorname{grad}_\vartheta l(\vartheta\,;\,\cdot)\operatorname{grad}_\vartheta^T l(\vartheta\,;\,\cdot)\Big)$$

heißt *Fisher-Informationsmatrix* (des statistischen Modells $(\Omega, \mathscr{A}, \mathscr{P})$).

Aussage 7.7 (Cramer-Rao-Ungleichung). *Es gelte*

a) Für $\mathbb{P}_\vartheta$-fast alle ω sei $\vartheta \longrightarrow L(\vartheta\,;\,\omega)$ stetig differenzierbar bezüglich ϑ,

b) $\mathbb{E}_\vartheta\left(\|\operatorname{grad}_\vartheta l(\vartheta\,;\,\cdot)\|^2\right) < \infty, \quad \vartheta \in \Theta$

c) Für jede Zufallsgröße Y mit $\mathbb{E}_\vartheta\big(Y^2\big) < \infty$ gelte

$$\operatorname{grad}_\vartheta \mathbb{E}_\vartheta\big(L(\vartheta\,;\,\cdot)Y\big) = \mathbb{E}_\vartheta\big((\operatorname{grad}_\vartheta L(\vartheta\,;\,\cdot)Y\big), \quad \vartheta \in \Theta$$

d) $I(\vartheta)$ sei regulär, $\vartheta \in \Theta$.

Dann gilt für jedes $\vartheta \in \Theta$ und für jede Zufallsgröße Y mit $\mathbb{E}_\vartheta\big(|Y|\big) < \infty$ die Ungleichung

$$\mathbb{E}_\vartheta\left((Y - \mathbb{E}_\vartheta(Y))^2\right) \geq \big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T I^{-1}(\vartheta)\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big).$$

Spezialisierungen:
Wenn Y eine erwartungstreue Schätzung für $\gamma(\vartheta)$ ist, so gilt:

$$D_\vartheta^2 Y \geq \left(\operatorname{grad}_\vartheta \gamma\,(\vartheta)\right)^T I^{-1}(\vartheta)\left(\operatorname{grad}_\vartheta \gamma\,(\vartheta)\right),$$

und ist Θ eindimensional, $\Theta \subseteq \mathbb{R}$, so haben wir

$$D_\vartheta^2 Y \geq \frac{\left(\dot\gamma(\vartheta)\right)^2}{I(\vartheta)}.$$

Beweis der Cramer-Rao-Ungleichung:

1. Schritt: Aus c) folgt

   $$\mathbb{E}\big(\operatorname{grad}_\vartheta L(\vartheta\,;\,\cdot)\big) = 0, \quad \text{d.h.} \quad \mathbb{E}_\vartheta\big(\operatorname{grad}_\vartheta \ln L(\vartheta\,;\,\cdot)\big) = 0 \;\; (\text{Wegen } \mathbb{E}\big(L(\vartheta\,;\,\cdot)\big) = 1)$$

   Es gilt folglich:

   $$\begin{aligned}\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) = \operatorname{grad}_\vartheta \mathbb{E}\big(L(\vartheta\,;\,\cdot)Y\big) = \mathbb{E}\big((\operatorname{grad}_\vartheta L(\vartheta\,;\,\cdot))\,Y\big) = \\ \mathbb{E}_\vartheta\Big((\operatorname{grad}_\vartheta \ln L(\vartheta\,;\,\cdot))\,Y\Big) = \mathbb{E}_\vartheta\Big((\operatorname{grad}_\vartheta \ln L(\vartheta\,;\,\cdot))\,(Y - \mathbb{E}_\vartheta(Y))\Big)\end{aligned} \tag{7.2}$$

2. Schritt: Für alle $u \in \mathbb{R}^k$ gilt mit $< u, v >= \sum_{l=1}^k u_k v_k = u^T v$

   $$< u, \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) > \overset{(7.2)}{=} \mathbb{E}_\vartheta\Big(< u, \big(\operatorname{grad}_\vartheta \ln L(\vartheta\,;\,\cdot)\big) > \cdot \big(Y - \mathbb{E}_\vartheta(Y)\big)\Big),$$

   und somit wegen der Cauchy-Schwarzschen Ungleichung

   $$< u, \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) >^2 \leq \mathbb{E}_\vartheta\Big(\big(< u, \big(\operatorname{grad}_\vartheta \ln L(\vartheta\,;\,\cdot)\big) >\big)^2\Big) \cdot \mathbb{E}_\vartheta\Big(\big(Y - \mathbb{E}_\vartheta(Y)\big)^2\Big).$$

   Das heißt,

   $$\mathbb{E}_\vartheta\Big(\big(Y - \mathbb{E}_\vartheta(Y)\big)^2\Big) \geq \frac{< u, \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) >^2}{\mathbb{E}_\vartheta\big(< u, \operatorname{grad}_\vartheta \ln L(\vartheta\,;\,\cdot) >^2\big)}, \quad \forall u \in \mathbb{R}^k,\ \forall \vartheta \in \Theta. \tag{7.3}$$

   Wir fixieren $\vartheta \in \Theta$ und wählen $u \in \mathbb{R}^k$ so, daß

   $$< u, \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) >= 1 \tag{7.4}$$

   erfüllt ist. Die rechte Seite von (7.3) lautet dann

   $$\frac{1}{u^T I(\vartheta) u}$$

3. Schritt: Bestimmung des Maximums der rechten Seite von (7.3) unter der Nebenbedingung (7.4).

$$u^T I(\vartheta)u - \lambda\big(< u, \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) > -1\big) \quad \text{soll minimiert werden.}$$

Wir erhalten als notwendige Bedingung

$$\operatorname{grad}_\vartheta : \quad u^T I(\vartheta) - \lambda\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T = 0 \quad \text{und}$$

$$\frac{d}{d\lambda} : \quad < u, \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) >= 1.$$

Auflösung nach u und λ: Erste Gleichung nach u^T auflösen und in die zweite einsetzen liefert:

$$1 = u^T \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) = \lambda\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T I^{-1}(\vartheta)\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big).$$

Also:

$$\lambda = \frac{1}{\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T I^{-1}(\vartheta)\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)},$$

$$u^T = \lambda \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) I^{-1}(\vartheta).$$

Damit folgt

$$\frac{1}{u^T I(\vartheta)u} = \frac{1}{\lambda^2\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)^T I^{-1}(\vartheta)\big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\big)}.$$

□

Bemerkung 7.9. Die Fishersche Informationsmatrix ist definiert durch

$$I(\vartheta) := \mathbb{E}_\vartheta\Big(\operatorname{grad}_\vartheta \ln L(\vartheta\,;\,\cdot)\operatorname{grad}_\vartheta^T \ln L(\vartheta\,;\,\cdot)\Big).$$

Unter der Bedingung, daß $\vartheta \to L(\vartheta\,;\,\cdot)$ zweimal differenzierbar ist und die Vertauschungsgleichung

$$\mathbb{E}\Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j} L(\vartheta\,;\,\cdot)\Big) = \frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\mathbb{E}\Big(L(\vartheta\,;\,\cdot)\Big) = 0$$

gilt ($\mathbb{E}$ ist der Erwartungswert bezüglich des dominierenden privilegierten Maßes, $L(\vartheta;\cdot) = \frac{d\mathbb{P}_\vartheta}{d\mathbb{P}}$), haben wir wegen

$$\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\ln L(\vartheta\,;\,\cdot) = \Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}L(\vartheta\,;\,\cdot)\Big)\frac{1}{L(\vartheta\,;\,\cdot)} - \Big(\frac{\partial}{\partial\vartheta_i}L(\vartheta\,;\,\cdot)\Big)\Big(\frac{\partial}{\partial\vartheta_j}L(\vartheta\,;\,\cdot)\Big)\frac{1}{\big(L(\vartheta\,;\,\cdot)\big)^2}$$

und

$$\mathbb{E}_\vartheta\Big(\frac{1}{L(\vartheta\,;\,\cdot)}\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}L(\vartheta\,;\,\cdot)\Big) = \mathbb{E}\Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}L(\vartheta\,;\,\cdot)\Big) = 0$$

die Beziehung

$$I(\vartheta) = -\,\mathbb{E}_\vartheta\Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\ln L(\vartheta\,;\,\cdot)\Big).$$

Beispiel 7.6. $\mathscr{P}$ bilde eine Exponentialfamilie. Dann ist

$$\begin{aligned}\frac{d\,\mathbb{P}_\vartheta}{d\,\mathbb{P}} &= \exp\Big\{<\vartheta, H> -\psi(\vartheta)\Big\}, \quad \vartheta\in\Theta,\ \Theta \text{ sei offen,}\\ I(\vartheta) &= \operatorname{grad}_\vartheta\big(\operatorname{grad}_\vartheta\psi(\vartheta)\big) = \Big(\frac{\partial^2}{\partial\vartheta_i\partial\vartheta_j}\psi(\vartheta)\Big) \quad \text{sei regulär.}\end{aligned}$$

Weiterhin sei

$$\begin{aligned} l(\vartheta) :&= \ln L(\vartheta\,;\,\cdot) =<\vartheta, H> -\psi(\vartheta) \quad \text{und es gilt}\\ \operatorname{grad}_\vartheta l(\vartheta) &= H - \operatorname{grad}_\vartheta\psi(\vartheta).\end{aligned}$$

Es folgt

$$\mathbb{E}_\vartheta(H) = \operatorname{grad}_\vartheta\psi(\vartheta), \quad \text{da } \mathbb{E}_\vartheta\big(\operatorname{grad}_\vartheta l(\vartheta)\big) = 0 \text{ ist.}$$

Als Zufallsgröße Y nehmen wir $< u, H >$ für ein $u \in \mathbb{R}^k$. Es ergibt sich

$$\begin{aligned}\mathbb{E}_\vartheta(Y) &=< u, \operatorname{grad}_\vartheta\psi(u) > \quad \text{und}\\ \operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y) &=< u, I(\vartheta) > .\end{aligned}$$

Folglich gilt:

$$\begin{aligned} D^2_\vartheta Y &= \mathbb{E}_\vartheta\Big(< H - \operatorname{grad}_\vartheta\psi(\vartheta), u >^2\Big)\\ &= \mathbb{E}_\vartheta\Big(u^T\big(H - \operatorname{grad}_\vartheta\psi(\vartheta)\big)\big(H - \operatorname{grad}_\vartheta\psi(\vartheta)\big)^T u\Big)\\ &= u^T I(\vartheta)u.\end{aligned}$$

Somit ist:

$$\begin{aligned} D^2_\vartheta Y &= u^T I(\vartheta)u = u^T I(\vartheta)I^{-1}(\vartheta)I(\vartheta)u\\ &= \Big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\Big)^T I^{-1}(\vartheta)\Big(\operatorname{grad}_\vartheta \mathbb{E}_\vartheta(Y)\Big).\end{aligned}$$

Also wird im Fall der Exponentialfamilie die untere Schranke in der Cramer-Rao-Ungleichung erreicht.

In der Ungleichung von Cramer-Rao muß es keine erwartungstreue Schätzung für $\gamma(\vartheta)$ geben, deren Streuung die untere Schranke erreicht.

Beispiel 7.7. Es sei X eine zum Parameter $\lambda > 0$ Poissonverteilte Zufallsgröße. X ist eine suffiziente und vollständige Statistik für λ, da $L(\lambda\,;\,x) = \lambda e^{-\lambda x}$. Es sei $T(X) := \mathbf{1}_{\{0\}}(X)$, dann gilt $\mathbb{E}_\lambda(T) = e^{-\lambda}$. D.h., T ist eine erwartungstreue Schätzung mit minimaler Varianz für $g(\lambda) = e^{-\lambda}$. Es gilt

$$D^2T = \mathbb{E}_\lambda\Big(\mathbf{1}_{\{0\}}(X) - e^{-\lambda}\Big)^2 = e^{-\lambda} - 2e^{-2\lambda} + e^{-2\lambda} = e^{-\lambda}(1-e^{-\lambda}).$$

Andererseits ist

$$I(\lambda) = \mathbb{E}_\lambda\Big(\big(\frac{\partial}{\partial\lambda}\ln L(\lambda\,;\,\cdot)\big)^2\Big) = \frac{1}{\lambda^2},\ \text{außerdem, mit } g'(\lambda) = -e^{-\lambda} \text{ gilt}$$

$$\frac{\big(g'(\lambda)\big)^2}{I(\lambda)} = \lambda^2 e^{-2\lambda} \ <\ e^{-\lambda}(1-e^{-\lambda}), \quad \text{wegen } 1+\lambda^2 < e^\lambda.$$

Beispiel 7.8. Es seien $X = (X_1,\dots,X_n)$ eine klassische mathematische Stichprobe unter $\mathbb{P}_\vartheta$, $\vartheta\in\Theta$, $Q_\vartheta := \mathbb{P}_\vartheta^{X_1}$, und $\frac{dQ_\vartheta}{d\,\mathbb{P}}(x) =: f_\vartheta(x)$. Es gilt:

$$\begin{aligned}
L(\vartheta\,;\,\omega) &= \prod_{k=1}^n \frac{dQ_\vartheta}{d\,\mathbb{P}}(X_k) = \prod_{k=1}^n f_\vartheta(X_k),\\
l(\vartheta\,;\,\omega) &= \sum_{k=1}^n \ln f_\vartheta(X_k),\\
\operatorname{grad}_\vartheta l(\vartheta\,;\,\omega) &= \sum_{k=1}^n \operatorname{grad}_\vartheta \ln f_\vartheta(X_k) = \sum_{k=1}^n \frac{\operatorname{grad}_\vartheta f_\vartheta(X_k)}{f_\vartheta(X_k)},\\
I(\vartheta) &= \mathbb{E}_\vartheta\big(\operatorname{grad}_\vartheta l(\vartheta\,;\,\omega)\operatorname{grad}_\vartheta l(\vartheta\,;\,\omega)^T\big)\\
&= \sum_{k=1}^n\sum_{l=1}^n \mathbb{E}_\vartheta\Big(\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_k)\big)\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_l)\big)^T\Big)
\end{aligned}$$

$\Big[$für $k\neq l$ sind X_k und X_l unabhängig, also gilt

$$\begin{aligned}
\mathbb{E}_\vartheta\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_k)\big) &= \mathbb{E}_\vartheta\Big(\frac{\operatorname{grad}_\vartheta f_\vartheta(X_k)}{f_\vartheta(X_k)}\Big) = \mathbb{E}_{\mathbb{P}}\Big(\operatorname{grad}_\vartheta \frac{dQ_\vartheta}{d\,\mathbb{P}}(X_k)\Big)\\
&= \operatorname{grad}_\vartheta \mathbb{E}\Big(\frac{dQ_\vartheta}{d\,\mathbb{P}}(X_1)\Big) = \operatorname{grad}_\vartheta 1 = 0.\Big]\\
&= \sum_{k=1}^n \underbrace{\mathbb{E}_\vartheta\Big(\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_k)\big)\big(\operatorname{grad}_\vartheta \ln f_\vartheta(X_k)\big)^T\Big)}_{=:\,I_1(\vartheta)}\\
&= n\cdot I_1(\vartheta).
\end{aligned}$$

Literaturverzeichnis

[1] Barndorff-Nielsen, O.E. (1988): Parametric Statistical Models and Likelihood, Lecture Notes in Statistics 50, Springer-Verlag

[2] *Basawa, I.V., Prahasa Rao, B.L.S. (1980): Statistical Inference for Stochastic Processes, Academic Press

[3] *Dacunha-Castelle, D., Duflo, M.(1983): Probability and Statistics, Springer-Verlag, Vol. I + II

[4] Karr; A. F. (1991): Point Processes and their Statistical Inference; Marcel Dekker, Inc., 2nd ed.

[5] Küchler,U.,Sørensen,M.(1997):Exponential Families of Stochastic Processes, Springer-Verlag

[6] Kutoyants, Yu.A. (2004) Statistical Inference for Ergodic Diffusion Processes, Springer-Verlag

[7] Lehn,J.,Wegmann,H.(1992): Einführung in die Statistik, Teubner Stuttgart

[8] Liptzer, R., Shiryaev,A. (1977, 1978): Statistics of Random Processes, Springer-Verlag (2 Bände)

[9] Löbus, J.-U.(2001): Ökonometrie, (Mathematische Theorie und Anwendungen) Vieweg-Verlag

[10] *Müller, P.H. et. al. (1991): Lexikon der Stochastik, Akademie-Verlag Berlin

[11] *Stahel, W.A. (1999) Statistische Datenanalyse, Vieweg-Verlag, zweite Auflage

[12] Winkler, W. (1983): Vorlesungen zur Mathematischen Statistik, Teubner Verlag Leipzig,

[13] *Witting,H. (1985): Mathematische Statistik I, Teubner Stuttgart

[14] Witting,H., Müller-Funk,U.(1995): Mathematische Statistik II, Teubner Stuttgart

Für die Grundlagen aus der Maßtheorie, der Theorie stochastischer Prozesse einschließlich Martingaltheorie, Markov-Prozesse und Prozessen mit unabhängigen Zuwächsen werden empfohlen:

[15] Jacod, J., Shiryaev, A. (1987): Limit Theorems for Stochastic Processes, Springer-Verlag

[16] *Karlin, S. (1975): A First Course in Probability Theory, Academic Press, New York

[17] Karlin, S., Taylor, H. (1975): A Second Course in Probability Theory, Academic Press, New York

[18] *Shiryaev, A. (1988): Wahrscheinlichkeit, Akademie-Verlag Berlin

[19] Skorochod, A.V. (1991): Random Prozesses with Independent Increments, Kluwer Amsterdam

Ergänzende Literatur:

[20] Andersen, P.K., Borgan, O, Gill, R.D., Keiding (1993): Statistical Models Based on Counting Processes Springer Series in Statistics.Springer-Verlag

[21] Franke,J.,Härdle,W.,Hafner,C.(2001): Einführung in die Statistik der Finanzmärkte, Springer Verlag

[22] Genon-Catalot,V., Picard,D. (1993): Elments de statistique asymptotiques, Springer Verlag, Series Mathematiques & Applications 11

[23] Ibragimov, I.A., Hasminskii, R.Z. (1981): Statistical Estimation, Asymptotic theory, Springer Verlag

[24] Kutoyants, Yu. (1984): Parameter estimation for stochastic processes, Heldermann, Berlin

[25] Kutoyants, Yu. (1994): Identification of Dynamical Systems with Small Noise; Kluwer Academic Publishers Amsterdam

[26] *Leadbetter,M.R., Lindgren,G.,Rootzn,H.(1983): Extremes, and Related Properties of Random Sequences and Processes, Springer Verlag, New York, Heidelberg, Berlin

[27] LeCam, L., Yang, G.L. (1990) Asymptotics in Statistics, Springer Series in Statistics

[28] *Reiss,R.D., Thomas,M.(2001): Statistical Analysis of Extreme Values, Birkhäuser Verlag

*** bedeutet „Besonders empfohlen“**