Albert-Ludwigs-Universität Freiburg

Vorlesungsskript

Mathematische Statistik

apl. Prof. Dr. Stefan Tappe

Wintersemester 2018/19

Abteilung für Mathematische Stochastik

Inhaltsverzeichnis

Kapitel 1

Punktschätzer

1.1 Statistische Modelle

Definition 1.1.1. *Es sei Θ eine Menge.*

(a) Ein statistisches Modell ist ein Tripel $\mathscr{M} = (\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta))$ bestehend aus:

- *Einer Menge Ω; dem Stichprobenraum.*
- *Einer σ-Algebra $\mathscr{F}$ über Ω; der Algebra der Beobachtungen.*
- *Einer Familie von Wahrscheinlichkeitsmaßen $(\mathbb{P}_\vartheta : \vartheta \in \Theta)$ auf $(\Omega, \mathscr{F})$.*

(b) Θ heißt der Parameterraum.

(c) Ist $\Theta \subset \mathbb{R}^d$, so sprechen wir von einem parametrischen Modell.

Definition 1.1.2. *Ist $\mathscr{M} = (\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta))$ ein statistisches Modell, so ist das n-fache Produktmodell gegeben durch*

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big).$$

Zur Erinnerung: Die Normalverteilung mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ ist die absolutstetige Verteilung auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Dichte $f : \mathbb{R} \to \mathbb{R}_+$ gegeben durch

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\bigg(- \frac{(x-\mu)^2}{2\sigma^2}\bigg).$$

Wir bezeichnen sie mit $\mathrm{N}(\mu, \sigma^2)$.

Beispiel 1.1.3 (Gauß-Produktmodell)**.** *Wir betrachten das Gauß-Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Theta = \mathbb{R} \times (0, \infty)$ *und*

$$\begin{aligned} \Omega &= \mathbb{R}, \\ \mathscr{F} &= \mathcal{B}(\mathbb{R}), \\ \mathbb{P}_\vartheta &= \mathrm{N}(\mu, \sigma^2), \quad \vartheta = (\mu, \sigma^2) \in \Theta. \end{aligned}$$

Wir nennen das n-fache Produktmodell

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

gemäß Definition 1.1.2 das *n-fache Gauß-Produktmodell.*

Definition 1.1.4. *Es sei* $(E, \mathscr{E})$ *ein messbarer Raum.*

(a) Eine messbare Abbildung $S : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ *heißt eine* *Statistik.*

(b) Eine Abbildung $\tau : \Theta \to E$ *heißt eine* *Kenngröße.*

(c) Eine Statistik $T : \Omega \to E$ *nennen wir auch einen* *Schätzer* *für* τ*. Häufig benutzen wir die Notation* $T = \widehat{\tau}$*.*

(d) Ist E *ein endlich-dimensionaler Vektorraum, so heißt ein Schätzer* T *erwartungstreu für* τ*, falls*

$$\mathbb{E}_\vartheta[T] = \tau(\vartheta) \quad \textit{für alle } \vartheta \in \Theta.$$

Bemerkung 1.1.5. *Wir bezeichnen mit* X *die Identität* $X : \Omega \to \Omega$ *gegeben durch* $X(\omega) = \omega$ *für alle* $\omega \in \Omega$*. Es gilt* $X \sim \mathbb{P}_\vartheta$ *unter* $\mathbb{P}_\vartheta$ *für alle* $\vartheta \in \Theta$*.*

Satz 1.1.6. *Es seien* $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ *Zufallsvariablen auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathscr{F}, \mathbb{P})$*. Dann sind folgende Aussagen äquivalent:*

(i) Die Zufallsvariablen $X_1, \ldots, X_n$ *sind unabhängig.*

(ii) Es gilt $\mathbb{P} \circ (X_1, \ldots, X_n) = \otimes_{i=1}^n (\mathbb{P} \circ X_i)$*.*

Beispiel 1.1.7. *Es sei* $\mathscr{M} = (\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta))$ *ein statistisches Modell mit* $\Omega \subset \mathbb{R}$*. Wir definieren folgende Statistiken auf dem Produktmodell* $\mathscr{M}^{\otimes n}$*:*

(a) Der *arithmetische Mittelwert* *ist*

$$(\Omega^n, \mathscr{F}^{\otimes n}) \to \mathbb{R}, \quad x \mapsto \bar{x} := \frac{1}{n} \sum_{i=1}^n x_i.$$

(b) Die <u>Stichprobenvarianz</u> ist

$$\widehat{\sigma}^2 : (\Omega^n, \mathscr{F}^{\otimes n}) \to \mathbb{R}, \quad \widehat{\sigma}^2(x) := \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

(c) Die <u>korrigierte Stichprobenvarianz</u> ist

$$s^2 : (\Omega^n, \mathscr{F}^{\otimes n}) \to \mathbb{R}, \quad s^2(x) := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Lemma 1.1.8. *Es gilt*

$$\begin{aligned} \widehat{\sigma}^2(x) &= \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - \bar{x}^2), \\ s^2(x) &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - \bar{x}^2). \end{aligned}$$

Beweis. Übung. □

Zur Erinnerung: Die <u>Bernoulli-Verteilung</u> mit Parameter $p \in [0,1]$ ist die diskrete Verteilung auf $E = \{0,1\}$ mit stochastischem Vektor $\pi : E \to [0,1]$ gegeben durch

$$\begin{aligned} \pi(0) &= 1 - p, \\ \pi(1) &= p. \end{aligned}$$

Wir bezeichnen sie mit $\mathrm{Ber}(p)$.

Beispiel 1.1.9 (Münzwurf)**.** *Wir nennen das statistische Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Theta = (0,1)$ *und*

$$\begin{aligned} \Omega &= \{0,1\}, \\ \mathscr{F} &= \mathfrak{P}(\Omega), \\ \mathbb{P}_\vartheta &= \mathrm{Ber}(\vartheta), \quad \vartheta \in \Theta \end{aligned}$$

das <u>Bernoulli-Modell</u>.

Zur Erinnerung: Die Binomialverteilung mit Parametern $p \in [0,1]$ und $n \in \mathbb{N}$ ist die diskrete Verteilung auf $E = \{0, 1, \ldots, n\}$ mit stochastischem Vektor $\pi : E \to [0,1]$ gegeben durch

$$\pi(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Wir bezeichnen sie mit $\mathrm{Bi}(n,p)$. Offensichtlich gilt $\mathrm{Bi}(1,p) = \mathrm{Ber}(p)$.

Beispiel 1.1.10 (Anzahl der Erfolge beim mehrfachen Münzwurf)**.** *Wir nennen das statistische Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Theta = (0,1)$ *und*

$$\begin{aligned} \Omega &= \{0, 1, \ldots, n\}, \\ \mathscr{F} &= \mathfrak{P}(\Omega), \\ \mathbb{P}_\vartheta &= \mathrm{Bi}(n, \vartheta), \quad \vartheta \in \Theta \end{aligned}$$

das Binomialmodell. Die Statistik

$$T : \Omega \to \mathbb{R}, \quad T(x) = \frac{x}{n}$$

ist ein erwartungstreuer Schätzer für die Kenngröße

$$\tau : \Theta \to \mathbb{R}, \quad \tau(\vartheta) = \vartheta.$$

In der Tat, es sei $\vartheta \in \Theta$ *beliebig. Dann gilt* $X \sim \mathrm{Bi}(n, \vartheta)$ *unter* $\mathbb{P}_\vartheta$, *und daher* $\mathbb{E}_\vartheta[X] = n\vartheta$. *Es folgt*

$$\mathbb{E}_\vartheta[T] = \frac{1}{n}\mathbb{E}_\vartheta[X] = \vartheta = \tau(\vartheta).$$

Beispiel 1.1.11 (Mehrfacher Münzwurf)**.** *Es sei* $\mathscr{M}$ *das Bernoulli-Modell aus Beispiel 1.1.9. Wir betrachten das* n*-fache Produktmodell*

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

gemäß Definition 1.1.2. Weiterhin seien

$$\begin{aligned} E &= \{0, 1, \ldots, n\}, \\ \mathscr{E} &= \mathfrak{P}(E) \end{aligned}$$

und $S : \Omega^n \to E$ die Statistik

$$S(x) = \sum_{i=1}^{n} x_i.$$

Wir setzen $\mathbb{Q}_\vartheta := \mathbb{P}_\vartheta^{\otimes n} \circ S$ für jedes $\vartheta \in \Theta$. Dann ist das statistische Mdell

$$\mathscr{N} = (E, \mathscr{E}, (\mathbb{Q}_\vartheta : \vartheta \in \Theta))$$

das Binomialmodell aus Beispiel 1.1.10.

Zur Erinnerung: Die Exponentialverteilung mit Parameter $\lambda \in (0, \infty)$ ist die absolutstetige Verteilung auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Dichte $f : \mathbb{R} \to \mathbb{R}_+$ gegeben durch

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x).$$

Wir bezeichnen sie mit $\operatorname{Exp}(\lambda)$. Für $X \sim \operatorname{Exp}(\lambda)$ gilt bekanntlich

$$\mathbb{P}(X \geq s + t \,|\, X \geq s) = \mathbb{P}(X \geq t) \quad \text{für alle } s, t \in \mathbb{R}_+.$$

Beispiel 1.1.12 (Glühbirnen)**.** *Wir betrachten das statistische Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Theta = (0, \infty)$ und

$$\begin{aligned} \Omega &= \mathbb{R}, \\ \mathscr{F} &= \mathcal{B}(\mathbb{R}), \\ \mathbb{P}_\vartheta &= \operatorname{Exp}(\vartheta), \quad \vartheta \in \Theta, \end{aligned}$$

und das zugehörige Produktmodell

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

gemäß Definition 1.1.2. Wir betrachten die Kenngröße

$$\tau : \Theta \to \mathbb{R}, \quad \tau(\vartheta) = \frac{1}{\vartheta}$$

und die Statistik

$$S : \Omega \to \mathbb{R}_+, \quad S(x) = \bar{x}.$$

Dann ist S ein erwartungstreuer Schätzer für τ. In der Tat, es sei $\vartheta \in \Theta$ beliebig. Dann gilt $\mathbb{E}_\vartheta[X_1] = \frac{1}{\vartheta}$, und daher

$$\mathbb{E}_\vartheta[S] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_\vartheta[X_i] = \frac{1}{\vartheta} = \tau(\vartheta).$$

Zur Erinnerung: Die Gleichverteilung mit Parametern $a, b \in \mathbb{R}$ mit $a < b$ ist die absolutstetige Verteilung auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Dichte $f : \mathbb{R} \to \mathbb{R}_+$ gegeben durch

$$f(x) = \frac{1}{b-a} \mathbb{1}_{(a,b)}(x).$$

Wir bezeichnen sie mit $\mathrm{UC}(a, b)$.

Beispiel 1.1.13. *Wir betrachten das statistische Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Theta = (0, \infty)$ *und*

$$\begin{aligned} \Omega &= \mathbb{R}, \\ \mathscr{F} &= \mathcal{B}(\mathbb{R}), \\ \mathbb{P}_\vartheta &= \mathrm{UC}(0, \vartheta), \quad \vartheta \in \Theta, \end{aligned}$$

und das zugehörige Produktmodell

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

gemäß Definition 1.1.2. Wir betrachten die Kenngröße

$$\tau : (0, \infty) \to \mathbb{R}, \quad \tau(\vartheta) = \vartheta,$$

und die zwei Schätzer

$$\begin{aligned} &L : \Omega^n \to \mathbb{R}, \quad L(x) = 2\bar{x}, \\ &M : \Omega^n \to \mathbb{R}, \quad M(x) = \max\{x_1, \dots, x_n\}. \end{aligned}$$

Dann ist L *ein erwartungstreuer Schätzer für* τ*, wohingegen*

$$\mathbb{E}_\vartheta[M] = \frac{n}{n+1}\vartheta \neq \vartheta \quad \textit{für alle } \vartheta \in \Theta.$$

Allerdings gilt

$$\lim_{n\to\infty} \frac{n}{n+1}\vartheta = \vartheta \quad \textit{für alle } \vartheta \in \Theta.$$

1.2 Konsistenz von Schätzern

Es seien $(\mathscr{M}_n)_{n\in\mathbb{N}}$ eine Folge von statistischen Modellen der Form

$$\mathscr{M}_n = \big(\Omega_n, \mathscr{F}_n, (\mathbb{P}_\vartheta^n : \vartheta \in \Theta)\big).$$

Definition 1.2.1. *Es seien $\tau : \Theta \to \mathbb{R}$ eine Kenngröße, und $(T_n)_{n\in\mathbb{N}}$ eine Folge von Schätzern $T_n : \Omega_n \to \mathbb{R}$. Die Folge $(T_n)_{n\in\mathbb{N}}$ heißt (schwach) konsistent für τ, falls*

$$\mathbb{P}^n_\vartheta(|T_n - \tau(\vartheta)| > \epsilon) \overset{n\to\infty}{\longrightarrow} 0 \quad \text{für alle } \epsilon > 0 \text{ und alle } \vartheta \in \Theta.$$

Beispiel 1.2.2. *Es sei $(M_n)_{n\in\mathbb{N}}$ die gemäß Beispiel 1.1.13 gegebene Folge von Schätern für $\tau(\vartheta) = \vartheta$. Die Verteilungsfunktionen $F_{M_n} : \mathbb{R} \to [0,1]$ sind gegeben durch*

$$F_{M_n}(x) = \frac{x^n}{\vartheta^n}\mathbb{1}_{(0,\vartheta)} + \mathbb{1}_{[\vartheta,\infty)}.$$

Also gilt für jedes $\epsilon > 0$

$$\mathbb{P}^n_\vartheta(|M_n - \vartheta| > \epsilon) = \mathbb{P}^n_\vartheta(\vartheta - M_n > \epsilon) = \mathbb{P}^n_\vartheta(M_n < \vartheta - \epsilon) = \left(\frac{\vartheta - \epsilon}{\vartheta}\right)^n \to 0.$$

Folglich ist $(M_n)_{n\in\mathbb{N}}$ konsistent für τ.

Satz 1.2.3. *Wir nehmen an, dass für alle $\vartheta \in \Theta$ gilt*

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}^n_\vartheta}[T_n] \to \tau(\vartheta) \quad \text{für } n \to \infty, \\ &\mathrm{Var}_{\mathbb{P}^n_\vartheta}[T_n] \to 0 \quad \text{für } n \to \infty. \end{aligned}$$

Dann ist $(T_n)_{n\in\mathbb{N}}$ konsistent für τ.

Beweis. Wir setzen für alle $n \in \mathbb{N}$ und $\vartheta \in \Theta$

$$\tau_n(\vartheta) := \mathbb{E}_{\mathbb{P}^n_\vartheta}[T_n].$$

Es sei $\epsilon > 0$ beliebig. Mit der Chebyshev-Ungleichung gilt

$$\begin{aligned} \mathbb{P}^n_\vartheta(|T_n - \vartheta| > \epsilon) &\le \mathbb{P}^n_\vartheta\left(|T_n - \tau_n(\vartheta)| > \frac{\epsilon}{2}\right) + \mathbb{P}^n_\vartheta\left(|\tau_n(\vartheta) - \tau(\vartheta)| > \frac{\epsilon}{2}\right) \\ &= \mathbb{P}^n_\vartheta\left(|T_n - \tau_n(\vartheta)| > \frac{\epsilon}{2}\right) + \mathbb{1}_{\{|\tau_n(\vartheta)-\tau(\vartheta)|>\frac{\epsilon}{2}\}} \\ &\le \left(\frac{2}{\epsilon}\right)^2 \mathrm{Var}_{\mathbb{P}^n_\vartheta}[T_n] + \mathbb{1}_{\{|\tau_n(\vartheta)-\tau(\vartheta)|>\frac{\epsilon}{2}\}} \to 0 \end{aligned}$$

für $n \to \infty$. □

Definition 1.2.4. *Es seien $(X_n)_{n\in\mathbb{N}}$ und X reellwertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathscr{F}, \mathbb{P})$.*

(a) Wir sagen, dass $(X_n)_{n\in\mathbb{N}}$ fast sicher gegen X konvergiert (und schreiben $X_n \overset{f.s.}{\to} X$), falls

$$\mathbb{P}(X_n \to X) = 1.$$

(b) *Wir sagen, dass* $(X_n)_{n\in\mathbb{N}}$ *stochastisch (oder in Wahrscheinlichkeit) gegen* X *konvergiert (und schreiben* $X_n \xrightarrow{\mathbb{P}} X$*), falls für jedes* $\epsilon > 0$ *gilt*

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0 \quad \textit{für } n \to \infty.$$

Satz 1.2.5. *Es seien* $(X_n)_{n\in\mathbb{N}}$ *und* X *reellwertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathscr{F}, \mathbb{P})$.

(a) *Falls* $X_n \xrightarrow{f.s.} X$*, dann gilt auch* $X_n \xrightarrow{\mathbb{P}} X$.

(b) *Falls* $X_n \xrightarrow{\mathbb{P}} X$*, dann existiert eine Teilfolge* $(n_k)_{k\in\mathbb{N}}$*, so dass* $X_{n_k} \xrightarrow{f.s.} X$ *für* $k \to \infty$.

Satz 1.2.6 (Starkes Gesetz der großen Zahlen von Kolmogorov)**.** *Es sei* $(X_n)_{n\in\mathbb{N}} \subset \mathscr{L}^1$ *eine Folge unabhängiger, identisch verteilter Zufallsvariablen auf einem Wahrscheinlichkeitsraum* $(\Omega, \mathscr{F}, \mathbb{P})$*. Dann gilt*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{f.s.} \mu \quad \textit{für } n \to \infty,$$

wobei $\mu := \mathbb{E}[X_1]$.

Definition 1.2.7. *Es seien* $(\mu_n)_{n\in\mathbb{N}}$ *und* μ *Wahrscheinlichkeitsmaße auf* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*. Wir sagen, dass* $(\mu_n)_{n\in\mathbb{N}}$ *schwach gegen* μ *konvergiert (und schreiben* $\mu_n \xrightarrow{w} \mu$*), falls*

$$\int_{\mathbb{R}} f\,d\mu_n \to \int_{\mathbb{R}} f\,d\mu$$

für jede stetige, beschränkte Funktion $f : \mathbb{R} \to \mathbb{R}$.

Satz 1.2.8. *Es seien* $(X_n)_{n\in\mathbb{N}}$ *und* X *Zufallsvariablen, so dass* $X_n \xrightarrow{\mathbb{P}} X$*. Dann gilt* $\mathbb{P} \circ X_n \xrightarrow{w} \mathbb{P} \circ X$.

Satz 1.2.9 (Zentraler Grenzwertsatz)**.** *Es sei* $(X_n)_{n\in\mathbb{N}} \subset \mathscr{L}^2$ *eine Folge unabhängiger, identisch verteilter Zufallsvariablen mit* $\mu := \mathbb{E}[X_1]$ *und* $\sigma^2 := \mathrm{Var}[X_1] > 0$*. Wir setzen*

$$S_n := \sum_{i=1}^{n} X_i \quad \textit{und} \quad Y_n := \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \quad \textit{für } n \in \mathbb{N}.$$

Dann gilt $\mathbb{P} \circ Y_n \xrightarrow{w} \mathrm{N}(0,1)$.

Nun betrachten wir etwas spezieller ein statistisches Modell

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Omega \subset \mathbb{R}$. Es sei $\Omega^{\mathbb{N}}$ der Raum aller Folgen mit Werten in Ω, versehen mit der Produkt-σ-Algebra

$$\mathscr{F}^{\otimes \mathbb{N}} := \sigma(X_n : n \in \mathbb{N}).$$

Hierbei ist $X_n : \Omega^{\mathbb{N}} \to \Omega$ die n-te Koordinatenabbildung. Nach dem Erweiterungssatz von Kolmogorov existiert für jedes $\vartheta \in \Theta$ ein eindeutig bestimmtes Wahrscheinlichkeitsmaß $\mathbb{P}_\vartheta$ auf $(\Omega^{\mathbb{N}}, \mathscr{F}^{\otimes \mathbb{N}})$, so dass

$$\mathbb{P}_\vartheta^{\otimes n} = \mathbb{P}_\vartheta \circ (X_1, \ldots, X_n) \quad \text{für jedes } n \in \mathbb{N}.$$

Wir nehmen an, dass die statistischen Modelle $(\mathscr{M}_n)_{n\in\mathbb{N}}$ gegeben sind durch die konstante Folge

$$\mathscr{M}_n = \big(\Omega^{\mathbb{N}}, \mathscr{F}^{\otimes \mathbb{N}}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big).$$

Definition 1.2.10. *Es seien $\tau : \Theta \to \mathbb{R}$ eine Kenngröße, und $(T_n)_{n\in\mathbb{N}}$ eine Folge von Schätzern $T_n : \Omega^{\mathbb{N}} \to \mathbb{R}$.*

(a) Die Folge $(T_n)_{n\in\mathbb{N}}$ heißt <u>(schwach) konsistent für τ</u>, falls

$$T_n \overset{\mathbb{P}_\vartheta}{\to} \tau(\vartheta) \quad \textit{für alle } \vartheta \in \Theta.$$

(b) Die Folge $(T_n)_{n\in\mathbb{N}}$ heißt <u>stark konsistent für τ</u>, falls

$$T_n \overset{f.s.}{\to} \tau(\vartheta) \textit{ bezüglich } \mathbb{P}_\vartheta \quad \textit{für alle } \vartheta \in \Theta.$$

(c) Die Folge $(T_n)_{n\in\mathbb{N}}$ heißt <u>asymptotisch normalverteilt</u>, falls für jedes $\vartheta \in \Theta$ Folgen $(\mu_n(\vartheta))_{n\in\mathbb{N}} \subset \mathbb{R}$ und $(\sigma_n^2(\vartheta))_{n\in\mathbb{N}} \subset (0, \infty)$ existieren, so dass für

$$Y_n(\vartheta) := \frac{T_n - \mu_n(\vartheta)}{\sigma_n(\vartheta)}$$

gilt $\mathbb{P}_\vartheta \circ Y_n(\vartheta) \overset{w}{\to} \mathrm{N}(0,1)$ für alle $\vartheta \in \Theta$.

Bemerkung 1.2.11. *Ist eine Folge $(T_n)_{n\in\mathbb{N}}$ von Schätzern asymptotisch normalverteilt, dann gilt für alle $\vartheta \in \Theta$*

$$\mathbb{P}_\vartheta \circ T_n \approx \mathrm{N}(\mu_n(\vartheta), \sigma_n^2(\vartheta)) \quad \textit{für große } n \in \mathbb{N}.$$

Beispiel 1.2.12. *Nun betrachten wir das Bernoulli-Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Theta = (0,1)$ *und*

$$\begin{aligned} \Omega &= \{0,1\}, \\ \mathscr{F} &= \mathfrak{P}(\Omega), \\ \mathbb{P}_\vartheta &= \mathrm{Ber}(\vartheta), \quad \vartheta \in \Theta \end{aligned}$$

aus Beispiel 1.1.9. Wir definieren die Kenngröße

$$\tau : \Theta \to \mathbb{R}, \quad \tau(\vartheta) = \vartheta$$

und die Folge $(T_n)_{n\in\mathbb{N}}$ *von Schätzern*

$$T_n : \Omega^{\mathbb{N}} \to \mathbb{R}, \quad T_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(a) Für jedes $n \in \mathbb{N}$ *ist* T_n *ein erwartungstreuer Schätzer für* τ*.*

(b) Die Folge $(T_n)_{n\in\mathbb{N}}$ *ist stark konsistent für* τ*.*

(c) Die Folge $(T_n)_{n\in\mathbb{N}}$ *ist asymptotisch normalverteilt*

Beweis. Es sei $\vartheta \in \Theta$ beliebig.

(a) Für jedes $n \in \mathbb{N}$ gilt

$$\mathbb{E}_\vartheta[T_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_\vartheta[X_i] = \vartheta = \tau(\vartheta).$$

(b) Nach dem starken Gesetz der großen Zahlen (Satz 1.2.6) gilt

$$T_n = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{f.s.}{\to} \vartheta = \tau(\vartheta) \quad \text{bezüglich } \mathbb{P}_\vartheta.$$

(c) Wir wählen

$$\mu_n(\vartheta) := \vartheta \quad \text{und} \quad \sigma_n^2(\vartheta) := \frac{\vartheta(1-\vartheta)}{n}.$$

Weiterhin setzen wir $S_n := \sum_{i=1}^{n} X_i$. Dann gilt

$$Y_n(\vartheta) := \frac{T_n - \mu_n(\vartheta)}{\sigma_n(\vartheta)} = \frac{\frac{S_n}{n} - \vartheta}{\sqrt{\frac{\vartheta(1-\vartheta)}{n}}} = \frac{S_n - n\vartheta}{\sqrt{n\vartheta(1-\vartheta)}},$$

und nach dem zentralen Grenzwertsatz (Satz 1.2.9) folgt $\mathbb{P}_\vartheta \circ Y_n(\vartheta) \overset{w}{\to} \mathrm{N}(0,1)$ für alle $\vartheta \in \Theta$.

□

1.3 Momentenschätzer

Wir betrachten ein statistisches Modell

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Omega \subset \mathbb{R}$, und das zugehörige Produktmodell

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

gemäß Definition 1.1.2. Es sei $d \in \mathbb{N}$, so dass $X_1 \in \mathscr{L}^d$.

Definition 1.3.1. *Es sei $k \in \{1, \ldots, d\}$ beliebig.*

(a) Wir definieren das k-te Moment

$$m_k : \Theta \to \mathbb{R}, \quad m_k(\vartheta) := \mathbb{E}_\vartheta[X_1^k].$$

(b) Wir definieren das k-te Stichprobenmoment

$$\widehat{m}_k : \Omega^n \to \mathbb{R}, \quad \widehat{m}_k(x) := \frac{1}{n} \sum_{i=1}^n x_i^k.$$

Lemma 1.3.2.

(a) Es gilt $\bar{x} = \widehat{m}_1(x)$.

(b) Es gilt $\widehat{\sigma}^2 = \widehat{m}_2 - \widehat{m}_1^2$.

Beweis. Übung. □

Definition 1.3.3. *Es sei $\tau : \Theta \to \mathbb{R}^e$ eine Kenngröße der Form*

$$\tau = f \circ (m_1, \ldots, m_d)$$

mit einer stetigen Funktion $f : D \to \mathbb{R}^e$, wobei $D \subset \mathbb{R}^d$. Wir nehmen hierbei an, dass

$$\begin{aligned} &(m_1(\vartheta), \ldots, m_d(\vartheta)) \in D \quad \textit{für alle } \vartheta \in \Theta \quad \textit{und} \\ &(\widehat{m}_1(x), \ldots, \widehat{m}_d(x)) \in D \quad \textit{für alle } x \in \Omega^n. \end{aligned}$$

Dann nennen wir $T : \Omega^n \to \mathbb{R}^e$ gegeben durch

$$T := f \circ (\widehat{m}_1, \ldots, \widehat{m}_d).$$

einen Momentenschätzer für τ.

Satz 1.3.4. *Momentenschätzer sind stark konsistent.*

Beweis. Es sei $\vartheta \in \Theta$ beliebig. Nach dem starken Gesetz der großen Zahlen (Satz 1.2.6) gilt für $n \to \infty$

$$\frac{1}{n} \sum_{i=1}^{n} X_i^k \overset{\text{f.s.}}{\to} m_k(\vartheta) \quad \text{bezüglich } \mathbb{P}_\vartheta$$

für alle $k = 1, \dots, d$. Wegen der Stetigkeit von f folgt für $n \to \infty$

$$f\left(\left(\frac{1}{n} \sum_{i=1}^{n} X_i^k\right)_{k=1,\dots,d}\right) \overset{\text{f.s.}}{\to} \tau(\vartheta) \quad \text{bezüglich } \mathbb{P}_\vartheta.$$

□

Beispiel 1.3.5. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

das Gauß-Modell aus Beispiel 1.1.3, wobei $\Theta = \mathbb{R} \times (0, \infty)$. Wir betrachten das zugehörige n-fache Gauß-Produktmodell, und die Kenngröße

$$\tau : \Theta \to \mathbb{R}^2, \quad \tau(\mu, \sigma^2) = (\mu, \sigma^2).$$

Dann gilt $\tau = f \circ (m_1, m_2)$ mit

$$f : D \to \mathbb{R}^2, \quad f(x, y) = (x, y - x^2),$$

wobei $D \subset \mathbb{R}^2$ gegeben ist durch $D = \Theta$. Also ist der Momentenschätzer gegeben durch

$$T = (\widehat{m}_1, \widehat{m}_2 - \widehat{m}_1^2) = (\bar{X}, \widehat{\sigma}^2(X)).$$

1.4 Maximum-Likelihood-Schätzer

Definition 1.4.1. *Ein Maß μ auf einem messbaren Raum $(\Omega, \mathscr{F})$ heißt <u>σ-endlich</u>, falls eine aufsteigende Folge $(B_n)_{n \in \mathbb{N}} \subset \mathscr{F}$ existiert, so dass $B_n \uparrow \Omega$ und $\mu(B_n) < \infty$ für alle $n \in \mathbb{N}$.*

Beispiele 1.4.2.

(a) Für jede abzählbare Menge E ist das Zählmaß ζ auf $(E, \mathfrak{P}(E))$ ein σ-endliches Maß.

(b) Das Lebesguemaß λ auf $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ ist σ-endlich.

Definition 1.4.3. *Es seien* $(\Omega, \mathscr{F})$ *ein messbarer Raum, und* μ, ν *zwei Maße auf* $(\Omega, \mathscr{F})$.

(a) *Wir nennen* ν *absolutstetig bezüglich* μ *(symbolisch* $\nu \ll \mu$*), falls*

$$\nu(A) = 0 \quad \textit{für alle } A \in \mathscr{F} \textit{ mit } \mu(A) = 0.$$

(b) *Wir nennen* ν *und* μ *äquivalent (symbolisch* $\nu \approx \mu$*), falls* $\nu \ll \mu$ *und* $\mu \ll \nu$*; das heißt, für jedes* $A \in \mathscr{F}$ *gilt* $\mu(A) = 0$ *genau dann, wenn* $\nu(A) = 0$.

Satz 1.4.4 (Satz von Radon-Nikodym)**.** *Es seien* μ *ein* σ*-endliches Maß und* ν *ein Maß auf einem messbaren Raum* $(\Omega, \mathscr{F})$*. Dann sind folgende Aussagen äquivalent:*

(i) *Es gilt* $\nu \ll \mu$.

(ii) *Es existiert eine nichtnegative, messbare Funktion* $f : \Omega \to \mathbb{R}_+$*, so dass*

$$\nu(A) = \int_A f \, d\mu \quad \textit{für alle } A \in \mathscr{F}.$$

In diesem Fall ist die Funktion f *fast sicher (bezüglich* μ*) eindeutig bestimmt, und wir bezeichnen Sie mit*

$$f = \frac{d\nu}{d\mu}.$$

Lemma 1.4.5. *Es seien* μ, ν, ρ σ*-endliche Maße auf einem messbaren Raum* $(\Omega, \mathscr{F})$.

(a) *Falls* $\varrho \ll \nu \ll \mu$*, dann gilt*

$$\frac{d\varrho}{d\mu} = \frac{d\varrho}{d\nu} \cdot \frac{d\nu}{d\mu}.$$

(b) *Falls* $\nu \approx \mu$*, dann gilt*

$$\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}.$$

Definition 1.4.6.

(a) *Ein statistisches Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

heißt ein Standardmodell, falls ein dominierendes Maß μ *auf* $(\Omega, \mathscr{F})$ *existiert; das heißt,* μ *ist* σ*-endlich und es gilt* $\mathbb{P}_\vartheta \ll \mu$ *für alle* $\vartheta \in \Theta$.

(b) *Die Funktion*

$$L : \Theta \times \Omega \to \mathbb{R}_+, \quad L(\vartheta, \cdot) = \frac{d\mathbb{P}_\vartheta}{d\mu}$$

heißt *Likelihood-Funktion* *des Parameters ϑ für die Beobachtung x.*

(c) *Gilt sogar $\mathbb{P}_\vartheta \approx \mu$ für alle $\vartheta \in \Theta$, so nennen wir*

$$\ell : \Theta \times \Omega \to \mathbb{R}, \quad \ell(\vartheta, x) = \ln L(\vartheta, x)$$

die *Log-Likelihood-Funktion.*

Bemerkung 1.4.7. *Es sei $\widetilde{\mu}$ ein σ-endliches Maß auf $(\Omega, \mathscr{F})$, so dass $\widetilde{\mu} \approx \mu$. Dann ist $\widetilde{\mu}$ ebenfalls ein dominierendes Maß, und für die Likelihood-Funktion gilt*

$$\widetilde{L}(\vartheta, x) = L(\vartheta, x) \cdot h(x),$$

wobei

$$h = \frac{d\mu}{d\widetilde{\mu}}.$$

Beispiele 1.4.8.

(a) *Beim Binomialmodell aus Beispiel 1.1.10 erhalten wir bezüglich $\mu = \zeta$ die Likelihood-Funktion*

$$L : (0,1) \times \{0, 1, \ldots, n\} \to (0, \infty), \quad L(\vartheta, x) = \binom{n}{x} \vartheta^x (1 - \vartheta)^{n-x}.$$

Die Log-Likelihood-Funktion ist

$$\ell(\vartheta, x) = \ln \binom{n}{x} + x \ln \vartheta + (n - x) \ln(1 - \vartheta).$$

(b) *Beim statistischen Modell $\mathscr{M}$ aus Beispiel 1.1.13 (mit der Gleichverteilung) erhalten wir bezüglich $\mu = \lambda$ die Likelihood-Funktion*

$$L : (0, \infty) \times \mathbb{R} \to \mathbb{R}_+, \quad L(\vartheta, x) = \frac{1}{\vartheta} \mathbb{1}_{(0,\vartheta)}(x).$$

Definition 1.4.9. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein Standardmodell mit Likelihood-Funktion $L : \Theta \times \Omega \to \mathbb{R}_+$. Ein Schätzer $T : \Omega \to \Theta$ für den Parameter ϑ heißt ein *Maximum-Likelihood-Schätzer,* *falls*

$$L(T(x), x) = \max_{\vartheta \in \Theta} L(\vartheta, x) \quad \text{für alle } x \in \Omega.$$

Bemerkung 1.4.10. *Unter geeigneten Voraussetzungen läßt sich beweisen, dass Maximum-Likelihood-Schätzer stark konsistent sind.*

Definition 1.4.11.

(a) Die Likelihood-Gleichung ist gegeben durch

$$\nabla_\vartheta L(\vartheta, x) = 0.$$

(b) Sofern $\mathbb{P}_\vartheta \approx \mu$ für alle $\vartheta \in \Theta$, so ist die Log-Likelihood-Gleichung gegeben durch

$$\nabla_\vartheta \ell(\vartheta, x) = 0$$

Beispiel 1.4.12. *Beim Binomialmodell lautet die Log-Likelihood-Gleichung*

$$\frac{x}{\vartheta} - \frac{n-x}{1-\vartheta} = 0.$$

Der Maximum-Likelihood-Schätzer ist

$$T = \frac{X}{n}.$$

Dies ist der Schätzer aus Beispiel 1.1.10.

Lemma 1.4.13. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein Standardmodell mit Likelihood-Funktion $L : \Theta \times \Omega \to \mathbb{R}_+$ bezüglich μ.

(a) Das statistische Modell

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

ist ebenfalls ein Standardmodell, und zwar mit Likelihood-Funktion

$$L^{\otimes n} : \Theta \times \Omega^n \to \mathbb{R}_+, \quad L^{\otimes n}(\vartheta, x) = \prod_{i=1}^n L(\vartheta, x_i).$$

bezüglich $\mu^{\otimes n}$.

(b) Gilt sogar $\mathbb{P}_\vartheta \approx \mu$ für alle $\vartheta \in \Theta$, dann gilt auch $\mathbb{P}_\vartheta^{\otimes n} \approx \mu^{\otimes n}$, und die Log-Likelihood-Funktion ist gegeben durch

$$\ell^{\otimes n} : \Theta \times \Omega^n \to \mathbb{R}_+, \quad \ell^{\otimes n}(\vartheta, x) = \sum_{i=1}^n \ell(\vartheta, x_i).$$

Beweis. Übung. □

Beispiel 1.4.14. *Beim Gauß-Modell aus Beispiel 1.1.3 erhalten wir bezüglich $\mu = \lambda$ die Likelihood-Funktion*

$$L : \big(\mathbb{R} \times (0, \infty)\big) \times \mathbb{R} \to (0, \infty), \quad L((\mu, \sigma^2), x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right).$$

Die Log-Likelihood-Funktion ist gegeben durch

$$\ell : \big(\mathbb{R} \times (0, \infty)\big) \times \mathbb{R} \to (0, \infty), \quad \ell((\mu, \sigma^2), x) = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Also ist die Log-Likelihood-Funktion beim n-fachen Gauß-Produktmodell gegeben durch

$$\begin{aligned} &\ell^{\otimes n} : \big(\mathbb{R} \times (0, \infty)\big) \times \mathbb{R}^n \to (0, \infty), \\ &\ell^{\otimes n}((\mu, \sigma^2), x) = n \cdot \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}, \end{aligned}$$

Die Log-Likelihood-Gleichung lautet

$$\left(\sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2}, \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} \right) = 0.$$

Also ist der Maximum-Likelihood-Schätzer

$$T = (\bar{X}, \widehat{\sigma}^2(X)).$$

Lemma 1.4.15. *Es gelten folgende Aussagen:*

(a) $\overline{x + \mu\mathbb{1}} = \overline{x} + \mu$ *für alle* $x \in \mathbb{R}^n$ *und* $\mu \in \mathbb{R}$.

(b) $\widehat{\sigma}^2(x + \mu\mathbb{1}) = \widehat{\sigma}^2(x)$ *für alle* $x \in \mathbb{R}^n$ *und* $\mu \in \mathbb{R}$.

Beweis. Übung. □

Zur Erinnerung: Die Gammaverteilung mit Parametern $\alpha, \lambda \in (0, \infty)$ ist die absolutstetige Verteilung auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Dichte $f : \mathbb{R} \to \mathbb{R}_+$ gegeben durch

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \mathbb{1}_{(0,\infty)}(x).$$

Wir bezeichnen sie mit $\Gamma(\alpha, \lambda)$. Es gilt $\Gamma(1, \lambda) = \mathrm{Exp}(\lambda)$.

Für $n \in \mathbb{N}$ nennen wir $\chi_n^2 = \Gamma(\frac{n}{2}, \frac{1}{2})$ eine Chi-Quadrat-Verteilung mit n Freiheitsgraden.

Lemma 1.4.16. *Es seien $X_1, \ldots, X_n \sim \mathrm{N}(0,1)$ unabhängige, standardnormalverteilte Zufallsvariablen und $S := X_1^2 + \ldots + X_n^2$. Dann gilt $S \sim \chi_n^2$.*

Satz 1.4.17. *Für das n-fache Gauß-Produktmodell gelten folgende Aussagen:*

(a) Es gilt

$$\bar{X} \sim \mathrm{N}\Big(\mu, \frac{\sigma^2}{n}\Big) \quad \text{und} \quad \frac{n}{\sigma^2}\widehat{\sigma}^2(X) = \frac{n-1}{\sigma^2}s^2(X) \sim \chi_{n-1}^2.$$

(b) $\bar{X}$ und $\widehat{\sigma}^2(X)$ sind unabhängig.

(c) $\bar{X}$ ist ein erwartungstreuer Schätzer für μ.

(d) Für alle $\vartheta = (\mu, \sigma^2) \in \Theta$ gilt

$$\mathbb{E}_\vartheta[\widehat{\sigma}^2(X)] = \frac{n-1}{n}\sigma^2.$$

(e) $s^2(X)$ ist ein erwartungstreuer Schätzer für σ^2.

Beweis. Es ist leicht nachzuweisen, dass $\bar{X}$ ist ein erwartungstreuer Schätzer für μ ist, und dass

$$\bar{X} \sim \mathrm{N}\Big(\mu, \frac{\sigma^2}{n}\Big).$$

Nach Lemma 1.4.15 dürfen wir nun annehmen, dass $\mu = 0$. Wir setzen

$$v_1 := \frac{1}{\sqrt{n}}\mathbb{1} \in \mathbb{R}^n,$$

und wählen $v_2, \ldots, v_n \in \mathbb{R}^n$, so dass $\{v_1, \ldots, v_n\}$ eine ONB von $\mathbb{R}^n$ ist. Wir definieren die orthogonale Matrix

$$A := \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ & v_2 & \\ & \vdots & \\ & v_n & \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Dann gilt $Av_i = e_i$ für alle $i = 1, \ldots, n$. Wir setzen $Y := AX$. Dann sind $Y_1, \ldots, Y_n \sim \mathrm{N}(0, \sigma^2)$ unabhängig. (Kleine Übung.) Wir setzen $Z_1 := Y_1 \in \mathbb{R}$ und $Z_2 := (Y_2, \ldots, Y_n) \in \mathbb{R}^{n-1}$. Es gilt

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{1}{\sqrt{n}}X_i = \frac{1}{\sqrt{n}}Z_1.$$

Weiterhin gilt

$$\begin{aligned}\|X - \bar{X}\mathbb{1}\|^2 &= \|A(X - \bar{X}\mathbb{1})\|^2 = \|Y - \bar{X}A\mathbb{1}\|^2 = \|Y - Z_1 A v_1\|^2 \\ &= \|Y - Z_1 e_1\|^2 = \|Y - (Z_1, 0)\|^2 = \|(0, Z_2)\|^2\end{aligned}$$

Wegen

$$\widehat{\sigma}^2(X) = \frac{1}{n}\|X - \bar{X}\mathbb{1}\|^2 = \frac{1}{n}\|(0, Z_2)\|^2$$

folgt die Unabhängigkeit von $\bar{X}$ und $\widehat{\sigma}^2(X)$. Außerdem gilt für jedes $\vartheta \in \Theta$

$$\mathbb{E}_\vartheta[\|(0, Z_2)\|^2] = \sum_{i=2}^{n} \mathbb{E}_\vartheta[Y_i^2] = (n-1)\sigma^2,$$

und nach Lemma 1.4.16 gilt

$$\frac{\|(0, Z_2)\|^2}{\sigma^2} \sim \chi^2_{n-1}.$$

□

1.5 Mittlerer quadratischer Fehler

Definition 1.5.1. *Es seien*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell, $\tau : \Theta \to \mathbb{R}$ *eine Kenngröße und* $T : \Omega \to \mathbb{R}$ *ein Schätzer für* τ*.*

(a) *Wir nennen*

$$\mathbb{B}_\vartheta(T) := \mathbb{E}_\vartheta[T] - \tau(\vartheta), \quad \vartheta \in \Theta$$

den systematischen Fehler (englisch: bias) des Schätzers T*.*

(b) *Wir nennen*

$$\mathrm{Var}_\vartheta[T] := \mathbb{E}_\vartheta[(T - \mathbb{E}_\vartheta[T])^2], \quad \vartheta \in \Theta$$

die Varianz (bzw. Streuung) von T*.*

(c) *Wir nennen*

$$\mathbb{F}_\vartheta(T) := \mathbb{E}_\vartheta[(T-\tau(\vartheta))^2], \quad \vartheta \in \Theta$$

den mittleren quadratischen Fehler (oder kurz Fehler) von T.

Lemma 1.5.2. *Für alle* $\vartheta \in \Theta$ *gilt die Zerlegung*

$$\mathbb{F}_\vartheta(T) = \mathrm{Var}_\vartheta[T] + \mathbb{B}_\vartheta(T)^2.$$

Beweis. Es gilt

$$\begin{aligned}
\mathbb{F}_\vartheta(T) &= \mathbb{E}_\vartheta[(T-\tau(\vartheta))^2] \\
&= \mathbb{E}_\vartheta[(T-\mathbb{E}_\vartheta[T]+\mathbb{E}_\vartheta[T]-\tau(\vartheta))^2] \\
&= \mathbb{E}_\vartheta[(T-\mathbb{E}_\vartheta[T]+\mathbb{B}_\vartheta(T))^2] \\
&= \mathbb{E}_\vartheta\big[(T-\mathbb{E}_\vartheta[T])^2 + 2(T-\mathbb{E}_\vartheta[T])\mathbb{B}_\vartheta(T) + \mathbb{B}_\vartheta(T)^2\big] \\
&= \mathrm{Var}_\vartheta[T] + 2\mathbb{B}_\vartheta(T)\underbrace{\mathbb{E}_\vartheta[T-\mathbb{E}_\vartheta[T]]}_{=0} + \mathbb{B}_\vartheta(T)^2 \\
&= \mathrm{Var}_\vartheta[T] + \mathbb{B}_\vartheta(T)^2.
\end{aligned}$$

□

Lemma 1.5.3. *Die folgenden Aussagen sind äquivalent:*

(i) T *ist erwartungstreu.*

(ii) $B_\vartheta(T) = 0$ *für alle* $\vartheta \in \Theta$.

(iii) $\mathbb{F}_\vartheta(T) = \mathrm{Var}_\vartheta[T]$ *für alle* $\vartheta \in \Theta$.

Beweis. Folgt aus Lemma 1.5.2. □

Beispiel 1.5.4. *Wir betrachten das Binomialmodell aus Beispiel 1.1.10, die Kenngröße* $\tau : \Theta = (0,1) \to \mathbb{R}$, $\tau(\vartheta) = \vartheta$ *und die Schätzer*

$$T = \frac{X}{n} \quad \textit{und} \quad S = \frac{X+1}{n+2}.$$

Dann ist T *erwartungstreu für* τ. *Weiter gilt*

$$\mathbb{F}_\vartheta(T) = \mathrm{Var}_\vartheta[T] = \frac{1}{n^2}\mathrm{Var}_\vartheta[X] = \frac{1}{n^2}n\vartheta(1-\vartheta) = \frac{\vartheta(1-\vartheta)}{n}.$$

Also gilt auf alle Fälle

$$F_\vartheta \leq \frac{1}{4n} \quad \textit{für alle } \vartheta \in \Theta.$$

Weiterhin gilt

$$\mathbb{B}_\vartheta(S) = \mathbb{E}_\vartheta\left[\frac{X+1}{n+2}\right] - \vartheta = \frac{1}{n+2}(\mathbb{E}_\vartheta[X]+1) - \vartheta = \frac{n\vartheta+1}{n+2} - \vartheta = \frac{1-2\vartheta}{n+2}.$$

Also ist S nicht erwartungstreu. Außerdem gilt

$$\begin{aligned}\mathrm{Var}_\vartheta[S] = \mathrm{Var}_\vartheta\left[\frac{X+1}{n+2}\right] &= \frac{1}{(n+2)^2}\mathrm{Var}_\vartheta[X] \\ &= \frac{1}{(n+2)^2} \cdot n\vartheta(1-\vartheta).\end{aligned}$$

Also folgt

$$\mathbb{F}_\vartheta(S) = \mathrm{Var}_\vartheta[T] + \mathbb{B}_\vartheta(T)^2 = \frac{n\vartheta(1-\vartheta)}{(n+2)^2} + \left(\frac{1-2\vartheta}{n+2}\right)^2.$$

Es gibt ein Intervall $I_n \subset \Theta = (0,1)$, so dass

$$\mathbb{F}_\vartheta(S) \le \mathbb{F}_\vartheta(T) \quad \textit{für alle } \vartheta \in I_n.$$

Lemma 1.5.5. *Es seien $p \in (0,\infty)$ und $X \ge 0$ eine nichtnegative Zufallsvariable. Dann gilt*

$$\mathbb{E}[X^p] = \int_0^\infty p\lambda^{p-1}\mathbb{P}(X > \lambda)d\lambda.$$

Beweis. Mit Hilfe des Satzes von Fubini erhalten wir

$$\begin{aligned}\mathbb{E}[X^p] &= \mathbb{E}\left[\int_0^X p\lambda^{p-1}d\lambda\right] = \mathbb{E}\left[\int_0^\infty p\lambda^{p-1}\mathbb{1}_{\{X>\lambda\}}d\lambda\right] \\ &= \int_0^\infty p\lambda^{p-1}\mathbb{E}[\mathbb{1}_{\{X>\lambda\}}]d\lambda = \int_0^\infty p\lambda^{p-1}\mathbb{P}(X>\lambda)d\lambda.\end{aligned}$$

□

Beispiel 1.5.6. *Wie in Beispiel 1.1.13 betrachten wir das statistische Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Theta = (0,\infty)$ und

$$\begin{aligned}\Omega &= \mathbb{R}, \\ \mathscr{F} &= \mathcal{B}(\mathbb{R}), \\ \mathbb{P}_\vartheta &= \mathrm{UC}(0,\vartheta), \quad \vartheta \in \Theta,\end{aligned}$$

und das zugehörige Produktmodell

$$\mathscr{M}^{\otimes n} = \left(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\right)$$

gemäß Definition 1.1.2. Wir betrachten die Kenngröße

$$\tau : (0, \infty) \to \mathbb{R}, \quad \tau(\vartheta) = \vartheta.$$

Dann gilt $\tau = f \circ m_1$ *mit*

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = 2x.$$

Also ist der Momentenschätzer gegeben durch

$$L = 2\widehat{m}_1 = 2\bar{X}.$$

Die Likelihood-Funktion bezüglich $\mu = \lambda$ *ist*

$$L : \Theta \times \Omega \to \mathbb{R}_+, \quad L(\vartheta, x) = \frac{1}{\vartheta^n} \prod_{i=1}^n \mathbb{1}_{[0,\vartheta]}(x_i).$$

Also ist der Mamimum-Likelihood-Schätzer gegeben durch

$$M = \max\{X_1, \ldots, X_n\}.$$

Es gilt

$$\mathrm{Var}_\vartheta[X_1] = \frac{\vartheta^2}{12}.$$

Da L *erwartungstreu ist, folgt*

$$\mathbb{F}_\vartheta(L) = \mathrm{Var}_\vartheta[L] = \mathrm{Var}_\vartheta\left[\frac{2}{n}\sum_{i=1}^n X_i\right] = \frac{4}{n^2} \cdot n \cdot \frac{\vartheta^2}{12} = \frac{\vartheta^2}{3n}.$$

Wir hatten bereits gezeigt, dass

$$\mathbb{E}_\vartheta[M] = \frac{n}{n+1}\vartheta.$$

Also folgt

$$\mathbb{B}_\vartheta(M) = \mathbb{E}_\vartheta[M] - \tau(\vartheta) = \frac{n}{n+1}\vartheta - \vartheta = \frac{n}{n+1}\vartheta - \frac{n+1}{n+1}\vartheta = -\frac{\vartheta}{n+1}.$$

Außerdem gilt

$$\mathbb{P}_\vartheta^{\otimes n}(M > \lambda) = \left[1 - \left(\frac{\lambda}{\vartheta}\right)^n\right] \mathbb{1}_{[0,\vartheta]}(\lambda), \quad \lambda \in \mathbb{R}_+.$$

Mit Lemma 1.5.5 folgt

$$\begin{aligned}
\mathbb{E}_\vartheta[M^2] &= 2\int_0^\infty \lambda \mathbb{P}_\vartheta^{\otimes n}(X > \lambda) d\lambda = 2\int_0^\vartheta \lambda\left[1 - \left(\frac{\lambda}{\vartheta}\right)^n\right] d\lambda \\
&= 2\int_0^\vartheta \lambda d\lambda - 2\int_0^\vartheta \frac{\lambda^{n+1}}{\vartheta^n} d\lambda = \vartheta^2 - \frac{2}{\vartheta^n}\frac{\vartheta^{n+2}}{n+2} \\
&= \frac{n+2}{n+2}\vartheta^2 - \frac{2}{n+2}\vartheta^2 = \frac{n}{n+2}\vartheta^2.
\end{aligned}$$

Es folgt

$$\begin{aligned}
\mathbb{F}_\vartheta(M) &= \mathrm{Var}_\vartheta[M] + B_\vartheta(M)^2 \\
&= \mathbb{E}_\vartheta[M^2] - \mathbb{E}_\vartheta[M]^2 + B_\vartheta(M)^2 \\
&= \left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2} + \frac{1}{(n+1)^2}\right)\vartheta^2 = \frac{2\vartheta^2}{(n+2)(n+1)}.
\end{aligned}$$

Also gilt $\mathbb{F}_\vartheta(L) \overset{n\to\infty}{\longrightarrow} 0$ *linear, und* $\mathbb{F}_\vartheta(M) \overset{n\to\infty}{\longrightarrow} 0$ *quadratisch. Sogesehen ist der Schätzer* M *vorzuziehen.*

Beispiel 1.5.7. *Wir betrachten das n-fache Gauß-Produktmodell mit Parameterraum* $\Theta = \mathbb{R} \times (0, \infty)$.

(a) $\bar{X}$ *ist ein erwartungstreuer Schätzer für die Kenngröße* μ, *und es gilt*

$$\mathbb{F}_\vartheta(\bar{X}) = \frac{\sigma^2}{n},$$

wobei $\vartheta = (\mu, \sigma^2) \in \Theta$.

(b) *Für die Kenngröße* σ^2 *gilt*

$$\begin{aligned}
\mathbb{B}_\vartheta(\widehat{\sigma}^2(X)) &= -\frac{\sigma^2}{n}, \\
\mathrm{Var}_\vartheta[\widehat{\sigma}^2(X)] &= \frac{2(n-1)}{n^2}\sigma^4, \\
\mathbb{F}_\vartheta(\widehat{\sigma}^2(X)) &= \frac{2n-1}{n^2}\sigma^4.
\end{aligned}$$

(c) $s^2(X)$ ist ein erwartungstreuer Schätzer für die Kenngröße σ^2, und gilt

$$\mathbb{F}_\vartheta(s^2(X)) = \frac{2}{n-1}\sigma^4.$$

Beweis.

(a) Nach Satz 1.4.17 gilt

$$\bar{X} \sim \mathrm{N}\Big(\mu, \frac{\sigma^2}{n}\Big).$$

(b) Nach Satz 1.4.17 gilt

$$\mathbb{B}_\vartheta(\widehat{\sigma}^2(X)) = \mathbb{E}_\vartheta[\widehat{\sigma}^2(X)] - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}.$$

Für $Y \sim \chi^2_n$ gilt bekanntlich

$$\mathbb{E}[Y] = n \quad \text{und} \quad \mathrm{Var}[Y] = 2n.$$

Also folgt weiter

$$\mathrm{Var}_\vartheta[\widehat{\sigma}^2(X)] = \frac{\sigma^4}{n^2} \underbrace{\mathrm{Var}_\vartheta\left[\frac{n}{\sigma^2}\widehat{\sigma}^2(X)\right]}_{=2(n-1)} = \frac{2(n-1)}{n^2}\sigma^4,$$

und damit

$$\mathbb{F}_\vartheta(\widehat{\sigma}^2(X)) = \mathrm{Var}_\vartheta[\widehat{\sigma}^2(X)] + (\mathbb{B}_\vartheta(\widehat{\sigma}^2(X)))^2 = \frac{2(n-1)}{n^2}\sigma^4 + \frac{\sigma^4}{n^2} = \frac{2n-1}{n^2}\sigma^4.$$

(c) Nach Satz 1.4.17 gilt

$$\mathbb{F}_\vartheta(s^2(X)) = \mathrm{Var}_\vartheta[s^2(X)] = \frac{\sigma^4}{(n-1)^2} \underbrace{\mathrm{Var}_\vartheta\left[\frac{n-1}{\sigma^2}s^2(X)\right]}_{=2(n-1)} = \frac{2}{n-1}\sigma^4.$$

□

Definition 1.5.8. *Es seien*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell, $\tau : \Theta \to \mathbb{R}$ eine Kenngröße und $T : \Omega \to \mathbb{R}$ ein erwartungstreuer Schätzer für τ. Dann heißt T varianzminimierend (bzw. ein gleichmäßig bester Schätzer oder UMV für "uniformly minimizing variance"), wenn

$$\mathrm{Var}_\vartheta[T] \leq \mathrm{Var}_\vartheta[S] \quad \Leftrightarrow \quad \mathbb{F}_\vartheta(T) \leq \mathbb{F}_\vartheta(S)$$

für alle $\vartheta \in \Theta$ und für jeden erwartungstreuen Schätzer $S : \Omega \to \mathbb{R}$ für τ.

1.6 Die Informationsungleichung

Definition 1.6.1. *Ein Standardmodell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

heißt regulär, falls gilt:

(a) $\Theta \subset \mathbb{R}$ *ist ein offenes Intervall.*

(b) $\mathbb{P}_\vartheta \approx \mu$ *für alle* $\vartheta \in \Theta$.

(c) Für jeden Schätzer $S : \Omega \to \mathbb{R}$ *und jedes* $\vartheta \in \Theta$ *mit* $S \in \mathscr{L}^1(\mathbb{P}_\vartheta) < \infty$ *gilt*

$$\frac{\partial}{\partial\vartheta} \int_\Omega S(x) L(\vartheta, x) \mu(dx) = \int_\Omega S(x) \frac{\partial}{\partial\vartheta} L(\vartheta, x) \mu(dx).$$

Bemerkung 1.6.2. *Es gilt*

$$\int_\Omega S(x) L(\vartheta, x) \mu(dx) = \mathbb{E}_\vartheta[S].$$

Definition 1.6.3. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein reguläres statistisches Modell.

(a) Wir nennen

$$u : \Theta \times \Omega \to \mathbb{R}, \quad u(\vartheta, x) = \frac{\partial}{\partial\vartheta} \ell(\vartheta, x) = \frac{\frac{\partial}{\partial\vartheta} L(\vartheta, x)}{L(\vartheta, x)}$$

die Einflussfunktion (oder Score-Funktion).

(b) Wir nennen

$$I : \Theta \to \mathbb{R}_+, \quad I(\vartheta) = \mathbb{E}_\vartheta[u(\vartheta, X)^2]$$

die Fischer-Information.

Die Einflussfunktion gibt also Auskunft über den Einfluss von Veränderungen des Parameters. Die Fischer-Information ist der gemittelte quadratische Einfluss; siehe auch Lemma 1.6.5 weiter unten.

Bemerkung 1.6.4. *Sowohl die Einflussfunktion als auch die Fischer-Information sind unabhängig von der Wahl des dominierenden Maßes. In der Tat, sei $\widetilde{\mu}$ ein σ-endliches Maß auf $(\Omega, \mathscr{F})$, so dass $\widetilde{\mu} \approx \mu$. Dann ist $\widetilde{\mu}$ ebenfalls ein dominierendes Maß, und für die Likelihood-Funktion gilt*

$$\widetilde{L}(\vartheta, x) = L(\vartheta, x) \cdot h(x),$$

wobei

$$h = \frac{d\mu}{d\widetilde{\mu}}.$$

Also gilt

$$\frac{\partial}{\partial\vartheta}\widetilde{\ell}(\vartheta, x) = \frac{\frac{\partial}{\partial\vartheta}\widetilde{L}(\vartheta, x)}{\widetilde{L}(\vartheta, x)} = \frac{\frac{\partial}{\partial\vartheta}L(\vartheta, x)}{L(\vartheta, x)} = \frac{\partial}{\partial\vartheta}\ell(\vartheta, x).$$

Lemma 1.6.5.

(a) Es gilt

$$\mathbb{E}_\vartheta[u(\vartheta, X)] = 0 \quad \textit{für alle } \vartheta \in \Theta.$$

(b) Es gilt

$$I(\vartheta) = \operatorname{Var}_\vartheta[u(\vartheta, X)] \quad \textit{für alle } \vartheta \in \Theta.$$

Beweis.

(a) Es gilt

$$\begin{aligned}\mathbb{E}_\vartheta[u(\vartheta, X)] &= \int_\Omega u(\vartheta, x)L(\vartheta, x)\mu(dx) \\ &= \int_\Omega \frac{\partial}{\partial\vartheta}L(\vartheta, x)\mu(dx) = \frac{\partial}{\partial\vartheta}\underbrace{\int_\Omega L(\vartheta, x)\mu(dx)}_{=1} = 0.\end{aligned}$$

(b) Folgt aus Teil (a).

□

Wie wir gesehen haben, gilt $I(\vartheta) \geq 0$ für alle $\vartheta \in \Theta$.

Satz 1.6.6 (Informationsungleichung von Cramer-Rao). *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein reguläres statistisches Modell mit $I(\vartheta) > 0$ *für alle* $\vartheta \in \Theta$*. Weiterhin seien* $\tau \in C^1(\Theta; \mathbb{R})$ *eine Kenngröße und* $T : \Omega \to \mathbb{R}$ *ein erwartungstreuer Schätzer für* τ*. Dann gilt*

$$\mathrm{Var}_\vartheta[T] \geq \frac{|\tau'(\vartheta)|^2}{I(\vartheta)} \quad \text{für alle } \vartheta \in \Theta.$$

Beweis. Es sei $\vartheta \in \Theta$ beliebig. Dann gilt

$$\begin{aligned}
\tau'(\vartheta) &= \frac{\partial}{\partial \vartheta} \tau(\vartheta) = \frac{\partial}{\partial \vartheta} \mathbb{E}_\vartheta[T] = \frac{\partial}{\partial \vartheta} \int_\Omega T(x) L(\vartheta, x) \mu(dx) \\
&= \int_\Omega T(x) \frac{\partial}{\partial \vartheta} L(\vartheta, x) \mu(dx) = \int_\Omega T(x) u(\vartheta, x) L(\vartheta, x) \mu(dx) \\
&= \mathbb{E}_\vartheta[T\, u(\vartheta, \cdot)] = \mathrm{Cov}_\vartheta(T, u(\vartheta, \cdot)).
\end{aligned}$$

Nach der Ungleichung von Cauchy-Schwarz gilt

$$\mathrm{Cov}_\vartheta(T, u(\vartheta, \cdot)) \leq \sqrt{\mathrm{Var}_\vartheta[T]} \cdot \sqrt{\mathrm{Var}_\vartheta[u(\vartheta, \cdot)]} = \sqrt{\mathrm{Var}_\vartheta[T]} \cdot \sqrt{I(\vartheta)}.$$

Also gilt

$$|\tau'(\vartheta)|^2 \leq \mathrm{Var}_\vartheta[T] \cdot I(\vartheta).$$

□

Definition 1.6.7. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein reguläres statistisches Modell mit $I(\vartheta) > 0$ *für alle* $\vartheta \in \Theta$*. Weiterhin seien* $\tau \in C^1(\Theta; \mathbb{R})$ *eine Kenngröße und* $T : \Omega \to \mathbb{R}$ *ein erwartungstreuer Schätzer für* τ*. Dann heißt* T *Cramer-Rao-optimal, falls*

$$\mathrm{Var}_\vartheta[T] = \frac{|\tau'(\vartheta)|^2}{I(\vartheta)} \quad \text{für alle } \vartheta \in \Theta.$$

Bemerkung 1.6.8. *Ein Cramer-Rao-optimaler Schätzer ist ein gleichmäßig bester Schätzer.*

Wie der Beweis von Satz 1.6.6 zeigt, gilt auch ohne die Annahme $I(\vartheta) > 0$ für alle $\vartheta \in \Theta$ die Ungleichung

$$|\tau'(\vartheta)|^2 \leq \mathrm{Var}_\vartheta[T] \cdot I(\vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

Also gilt die Implikation (i) $\Rightarrow$ (ii) folgender Aussagen:

(i) $\tau'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$.

(ii) $\mathrm{Var}_\vartheta[T] > 0$ und $I(\vartheta) > 0$ für alle $\vartheta \in \Theta$.

Die Äquivalenz (i) $\Leftrightarrow$ (ii) gilt, sofern

$$|\tau'(\vartheta)|^2 = \mathrm{Var}_\vartheta[T] \cdot I(\vartheta) \quad \text{für alle } \vartheta \in \Theta.$$

Dies ist insbesondere bei einem Cramer-Rao-optimalen Schätzer T der Fall.

Beispiel 1.6.9. *Wir betrachten das Binomialmodell mit bekanntem $n \in \mathbb{N}$; vergleiche Beispiel 1.1.10. Dann gilt $\Theta = (0,1)$ und*

$$\mathbb{P}_\vartheta = \mathrm{Bi}(n, \vartheta), \quad \vartheta \in \Theta.$$

Wir erhalten die Log-Likelihood-Funktion

$$\ell(\vartheta, x) = \ln \binom{n}{x} + x \ln \vartheta + (n-x) \ln(1-\vartheta),$$

siehe Beispiel 1.4.8(a). Es folgt

$$u(\vartheta, x) = \frac{\partial}{\partial \vartheta} \ell(\vartheta, x) = \frac{x}{\vartheta} - \frac{n-x}{1-\vartheta} = \frac{x(1-\vartheta) - (n-x)\vartheta}{\vartheta(1-\vartheta)} = \frac{x - n\vartheta}{\vartheta(1-\vartheta)}.$$

Weiter folgt

$$I(\vartheta) = \mathrm{Var}_\vartheta[u(\vartheta, X)] = \mathrm{Var}_\vartheta\left[\frac{X}{\vartheta(1-\vartheta)}\right] = \frac{1}{(\vartheta(1-\vartheta))^2} \cdot n\vartheta(1-\vartheta) = \frac{n}{\vartheta(1-\vartheta)}.$$

Nach Satz 1.6.6 gilt für jeden erwartungstreuen Schätzer T für die Kenngröße ϑ, dass

$$\mathrm{Var}_\vartheta[T] \geq \frac{\vartheta(1-\vartheta)}{n}.$$

Für den Schätzer $T = \frac{X}{n}$ gilt

$$\mathrm{Var}_\vartheta[T] = \frac{1}{n^2} \cdot n\vartheta(1-\vartheta) = \frac{\vartheta(1-\vartheta)}{n}.$$

Also ist T Cramer-Rao-optimal, und damit ein gleichmäßig bester Schätzer.

Lemma 1.6.10. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein reguläres statistisches Modell mit Fischer-Information $I : \Theta \to \mathbb{R}_+$.

(a) *Das statistische Modell*

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

ist ebenfalls regulär.

(b) *Die Fischer-Information* $I^{\otimes n} : \Omega^n \to \mathbb{R}_+$ *ist gegeben durch*

$$I^{\otimes n}(\vartheta) = n \cdot I(\vartheta), \quad \vartheta \in \Theta.$$

Beweis.

(a) Übung.

(b) Es gilt

$$I^{\otimes n}(\vartheta) = \mathrm{Var}_{\mathbb{P}_\vartheta^{\otimes n}}[u^{\otimes n}(\vartheta, X)] = \mathrm{Var}_{\mathbb{P}_\vartheta^{\otimes n}}\left[\frac{\partial}{\partial\vartheta}\ell^{\otimes n}(\vartheta, X)\right] = \mathrm{Var}_{\mathbb{P}_\vartheta^{\otimes n}}\left[\sum_{i=1}^n \frac{\partial}{\partial\vartheta}\ell(\vartheta, X_i)\right]$$
$$= \sum_{i=1}^n \mathrm{Var}_{\mathbb{P}_\vartheta^{\otimes n}}\left[\frac{\partial}{\partial\vartheta}\ell(\vartheta, X_i)\right] = \sum_{i=1}^n \mathrm{Var}_{\mathbb{P}_\vartheta^{\otimes n}}[u(\vartheta, X_i)] = \sum_{i=1}^n \mathrm{Var}_\vartheta[u(\vartheta, X)] = n \cdot I(\vartheta).$$

Schauen wir uns die vorletzte Gleichung noch etwas genauer an. Nach dem Satz von Fubini gilt für jedes $i = 1, \dots, n$

$$\begin{aligned}\mathrm{Var}_{\mathbb{P}_\vartheta^{\otimes n}}[u(\vartheta, X_i)] &= \mathbb{E}_{\mathbb{P}_\vartheta^{\otimes n}}[u(\vartheta, X_i)^2] = \int_{\Omega^n} u(\vartheta, x_i)^2 \mathbb{P}_\vartheta^{\otimes n}(dx)\\ &= \int_\Omega \cdots \int_\Omega u(\vartheta, x_i)^2 \mathbb{P}_\vartheta(dx_1) \dots \mathbb{P}_\vartheta(dx_n) = \int_\Omega u(\vartheta, x_i)^2 \mathbb{P}_\vartheta(dx_i)\\ &= \int_\Omega u(\vartheta, x)^2 \mathbb{P}_\vartheta(dx) = \mathbb{E}_\vartheta[u(\vartheta, X)^2] = \mathrm{Var}_\vartheta[u(\vartheta, X)].\end{aligned}$$

□

Beispiel 1.6.11. *Wir betrachten das Gaußmodell mit bekannter Varianz* $\sigma^2 > 0$. *Hierbei gilt* $\Theta = \mathbb{R}$ *und*

$$\mathbb{P}_\vartheta = \mathrm{N}(\vartheta, \sigma^2), \quad \vartheta \in \Theta.$$

Wir erhalten die Log-Likelihood-Funktion

$$\ell(\vartheta, x) = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\vartheta)^2}{2\sigma^2}.$$

Es folgt

$$u(\vartheta, x) = \frac{\partial}{\partial\vartheta}\ell(\vartheta, x) = \frac{x-\vartheta}{\sigma^2}.$$

Weiter folgt

$$I(\vartheta) = \mathrm{Var}_\vartheta[u(\vartheta, X)] = \mathrm{Var}_\vartheta\left[\frac{X-\vartheta}{\sigma^2}\right] = \frac{1}{\sigma^4}\mathrm{Var}_\vartheta[X] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}.$$

Für das n-fache Produktmodell folgt

$$I^{\otimes n}(\vartheta) = \frac{n}{\sigma^2}.$$

Nach Satz 1.6.6 gilt für jeden erwartungstreuen Schätzer T für die Kenngröße ϑ, dass

$$\mathrm{Var}_\vartheta[T] \geq \frac{\sigma^2}{n}.$$

Für das arithmetische Mittel $\bar{X}$ gilt (vergleiche Beispiel 1.5.7)

$$\mathrm{Var}_\vartheta[\bar{X}] = \frac{\sigma^2}{n}.$$

Also ist $\bar{X}$ Cramer-Rao-optimal, und damit ein gleichmäßig bester Schätzer.

Beispiel 1.6.12. *Wir betrachten das Poisson-Modell. Hierbei gilt $\Theta = (0, \infty)$ und*

$$\mathbb{P}_\vartheta = \mathrm{Pois}(\vartheta), \quad \vartheta \in \Theta.$$

Wir erhalten die Likelihood-Funktion

$$L(\vartheta, x) = e^{-\vartheta}\frac{\vartheta^x}{x!} = \frac{1}{x!}\exp\big(x\ln\vartheta - \vartheta\big),$$

und die Log-Likelihood-Funktion

$$\ell(\vartheta, x) = \ln\left(\frac{1}{x!}\right) + x\ln\vartheta - \vartheta.$$

Es folgt

$$u(\vartheta, x) = \frac{\partial}{\partial\vartheta}\ell(\vartheta, x) = \frac{x}{\vartheta} - 1.$$

Weiter folgt

$$I(\vartheta) = \mathrm{Var}_\vartheta[u(\vartheta, X)] = \mathrm{Var}_\vartheta\left[\frac{X}{\vartheta} - 1\right] = \frac{1}{\vartheta^2}\mathrm{Var}_\vartheta[X] = \frac{\vartheta}{\vartheta^2} = \frac{1}{\vartheta}.$$

Für das n-fache Produktmodell folgt

$$I^{\otimes n}(\vartheta) = \frac{n}{\vartheta}.$$

Nach Satz 1.6.6 gilt für jeden erwartungstreuen Schätzer T für die Kenngröße ϑ, dass

$$\mathrm{Var}_\vartheta[T] \geq \frac{\vartheta}{n}.$$

Für das arithmetische Mittel $\bar{X}$ gilt

$$\mathrm{Var}_\vartheta[\bar{X}] = \mathrm{Var}_\vartheta\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \underbrace{\mathrm{Var}_\vartheta[X_i]}_{=\vartheta} = \frac{\vartheta}{n}.$$

Also ist $\bar{X}$ Cramer-Rao-optimal, und damit ein gleichmäßig bester Schätzer.

1.7 Exponentielle Familien

Definition 1.7.1. *Ein reguläres Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

heißt eine exponentielle Familie, falls messbare Funktionen $a, b : \Theta \to \mathbb{R}$ mit $a \in C^1(\Theta; \mathbb{R})$, eine Statistik $T : \Omega \to \mathbb{R}$ und eine messbare Funktion $h : \Omega \to (0, \infty)$ existieren, so dass

$$L(\vartheta, x) = \exp\big(a(\vartheta)T(x) - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega.$$

In diesem Fall nennen wir T eine kanonische Statistik.

Bemerkung 1.7.2. *Es sei $\widetilde{\mu}$ ein σ-endliches Maß auf $(\Omega, \mathscr{F})$, so dass $\widetilde{\mu} \approx \mu$. Dann ist $\widetilde{\mu}$ ebenfalls ein dominierendes Maß, und für die Likelihood-Funktion gilt*

$$\widetilde{L}(\vartheta, x) = \exp\big(a(\vartheta)T(x) - b(\vartheta)\big) \cdot \tilde{h}(x),$$

wobei

$$\tilde{h} = h \cdot \frac{d\mu}{d\widetilde{\mu}}.$$

Bemerkung 1.7.3. *Ein reguläres Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ist genau dann eine exponentielle Familie, wenn messbare Funktionen $a, b : \Theta \to \mathbb{R}$ mit $a \in C^1(\Theta; \mathbb{R})$, eine Statistik $T : \Omega \to \mathbb{R}$ und eine messbare Funktion $k : \Omega \to \mathbb{R}$ existieren, so dass

$$\ell(\vartheta, x) = a(\vartheta)T(x) - b(\vartheta) + k(x) \quad \textit{für alle } (\vartheta, x) \in \Theta \times \Omega.$$

Lemma 1.7.4. *Die Funktion $b : \Theta \to \mathbb{R}$ ist gegeben durch*

$$b(\vartheta) = \ln\bigg(\int_\Omega \exp\big(a(\vartheta)T(x)\big)h(x)\mu(dx)\bigg).$$

Beweis. Es sei $\vartheta \in \Theta$ beliebig. Dann gilt

$$\begin{aligned}\exp(-b(\vartheta))\int_\Omega \exp\big(a(\vartheta)T(x)\big)h(x)\mu(dx) &= \int_\Omega \exp\big(a(\vartheta)T(x) - b(\vartheta)\big)h(x)\mu(dx)\\ &= \int_\Omega L(\vartheta, x)\mu(dx) = 1.\end{aligned}$$

Daraus folgt

$$\exp(b(\vartheta)) = \int_\Omega \exp\big(a(\vartheta)T(x)\big)h(x)\mu(dx).$$

□

Lemma 1.7.5. *Für jedes $\vartheta \in \Theta$ und jedes $S \in \mathscr{L}^1(\mathbb{P}_\vartheta)$ gilt*

$$\exp(b(\vartheta))\mathbb{E}_\vartheta[S] = \int_\Omega S(x)\exp(a(\vartheta)T(x))h(x)\mu(dx).$$

Beweis. Es gilt

$$\begin{aligned}\mathbb{E}_\vartheta[S] &= \int_\Omega S(x)L(\vartheta, x)\mu(dx) = \int_\Omega S(x)\exp\big(a(\vartheta)T(x) - b(\vartheta)\big)h(x)\mu(dx)\\ &= \exp(-b(\vartheta))\int_\Omega S(x)\exp\big(a(\vartheta)T(x)\big)h(x)\mu(dx).\end{aligned}$$

□

Bemerkung 1.7.6. *Es gelte $a'(\vartheta_0) \neq 0$ für ein $\vartheta_0 \in \Theta$. Dann existiert ein $\epsilon > 0$, so dass a auf $U = (\vartheta_0 - \epsilon, \vartheta_0 + \epsilon) \subset \Theta$ injektiv ist, und es gilt*

$$L(a^{-1}(\eta), x) = \exp\big(\eta T(x) - \bar{b}(\eta)\big) \cdot h(x) \quad \textit{für alle } \eta \in a(U),$$

wobei $\bar{b} = b \circ a^{-1}$.

Satz 1.7.7. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

eine exponentielle Familie mit Likelihood-Funktion

$$L(\vartheta, x) = \exp\big(a(\vartheta)T(x) - b(\vartheta)\big) \cdot h(x), \quad (\vartheta, x) \in \Theta \times \Omega,$$

so dass $a'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$.

(a) *Es gilt $b \in C^1(\Theta; \mathbb{R})$ mit*

$$b'(\vartheta) = a'(\vartheta)\mathbb{E}_\vartheta[T].$$

(b) *Für die Kenngröße $\tau : \Theta \to \mathbb{R}$, $\tau(\vartheta) := \mathbb{E}_\vartheta[T]$ gilt $\tau \in C^1(\Theta; \mathbb{R})$ mit*

$$\tau'(\vartheta) = a'(\vartheta)\mathrm{Var}_\vartheta[T].$$

(c) *Es gilt $I(\vartheta) = a'(\vartheta) \cdot \tau'(\vartheta)$.*

Beweis. Nach Bemerkung 1.7.6 dürfen wir annehmen, dass $a(\vartheta) = \vartheta$ für alle $\vartheta \in \Theta$. Etwa für Teil (a) werden wir in dem Fall zeigen, dass

$$b'(\vartheta) = \mathbb{E}_\vartheta[T].$$

Nach Bemerkung 1.7.6 gilt dann für eine allgemeine exponentielle Familie

$$\bar{b}'(\eta) = \mathbb{E}_{a^{-1}(\eta)}[T] \quad \text{für alle } \eta \in a(U).$$

Also gilt

$$(b \circ a^{-1})'(\eta) = \mathbb{E}_{a^{-1}(\eta)}[T] \quad \text{für alle } \eta \in a(U).$$

Nach der Kettenregel erhalten wir

$$(b \circ a^{-1})'(\eta) = b'(a^{-1}(\eta))(a^{-1})'(\eta) = \frac{b'(a^{-1}(\eta))}{a'(a^{-1}(\eta))} \quad \text{für alle } \eta \in a(U).$$

Daraus folgt

$$\frac{b'(\vartheta)}{a'(\vartheta)} = \mathbb{E}_\vartheta[T] \quad \text{für alle } \vartheta \in U.$$

Nun sei S eine Statistik mit $S \in \mathscr{L}^1(\mathbb{P}_\vartheta)$ für alle $\vartheta \in \Theta$. Wir definieren

$$u_S : \Theta \to \mathbb{R}, \quad u_S(\vartheta) := \exp(b(\vartheta))\mathbb{E}_\vartheta[S].$$

Nach Lemma 1.7.5 gilt

$$u_S(\vartheta) = \int_\Omega S(x)\exp(\vartheta T(x))h(x)\mu(dx).$$

Es gilt $u_S \in C^\infty(\Theta;\mathbb{R})$ mit

$$u_S^{(k)}(\vartheta) = \exp(b(\vartheta))\mathbb{E}[ST^k] \quad \text{für alle } k \in \mathbb{N}_0.$$

In der Tat, etwa für $k = 1$ gilt

$$\begin{aligned}\frac{\partial}{\partial\vartheta}u_S(\vartheta) &= \frac{\partial}{\partial\vartheta}\int_\Omega S(x)\exp(\vartheta T(x))h(x)\mu(dx)\\ &= \int_\Omega S(x)\frac{\partial}{\partial\vartheta}\exp(\vartheta T(x))h(x)\mu(dx)\\ &= \int_\Omega S(x)T(x)\exp(\vartheta T(x))h(x)\mu(dx) = \exp(b(\vartheta))\mathbb{E}[ST].\end{aligned}$$

(a) Wir setzen $S = 1$. Nach Lemma 1.7.4 gilt

$$b(\vartheta) = \ln(u_1(\vartheta)).$$

Es folgt $b \in C^1(\Theta;\mathbb{R})$ mit

$$b'(\vartheta) = \frac{u_1'(\vartheta)}{u_1(\vartheta)} = \mathbb{E}_\vartheta[T].$$

(b) Nach Teil (a) gilt für $\tau : \Theta \to \mathbb{R}$, $\tau(\vartheta) := \mathbb{E}_\vartheta[T]$, dass $\tau \in C^1(\Theta;\mathbb{R})$ mit

$$\begin{aligned}u_1'(\vartheta) &= u_1(\vartheta)\mathbb{E}_\vartheta[T],\\ u_1''(\vartheta) &= u_1(\vartheta)\mathbb{E}_\vartheta[T^2].\end{aligned}$$

Also folgt

$$\tau'(\vartheta) = b''(\vartheta) = \frac{u_1''(\vartheta)}{u_1(\vartheta)} - \left(\frac{u_1'(\vartheta)}{u_1(\vartheta)}\right)^2 = \mathbb{E}_\vartheta[T^2] - \mathbb{E}_\vartheta[T]^2 = \mathrm{Var}_\vartheta[T].$$

(c) Wegen $b'(\vartheta) = \mathbb{E}_\vartheta[T]$ gilt

$$\begin{aligned}\frac{\partial}{\partial\vartheta}\mathbb{E}_\vartheta[S] &= \frac{\partial}{\partial\vartheta}[u_S(\vartheta)\exp(-b(\vartheta))] = [u_S'(\vartheta) - u_S(\vartheta)b'(\vartheta)]\exp(-b(\vartheta))\\ &= \mathbb{E}_\vartheta[ST] - \mathbb{E}_\vartheta[S]b'(\vartheta) = \mathbb{E}_\vartheta[S(T - b'(\vartheta))] = \mathrm{Cov}_\vartheta(S,T).\end{aligned}$$

Mit $S = T$ folgt

$$\mathrm{Var}_\vartheta[T] = \frac{\partial}{\partial\vartheta}\mathbb{E}_\vartheta[T] = \tau'(\vartheta).$$

Nach Bemerkung 1.7.3 gilt

$$\ell(\vartheta, x) = \vartheta T(x) - b(\vartheta) + k(x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega.$$

Also folgt

$$u(\vartheta, x) = \frac{\partial}{\partial\vartheta}\ell(\vartheta, x) = T(x) - b'(\vartheta).$$

Nun erhalten wir

$$I(\vartheta) = \mathrm{Var}_\vartheta[u(\vartheta, X)] = \mathrm{Var}_\vartheta[T] = \tau'(\vartheta).$$

□

Wie Satz 1.7.7 zeigt, sind für eine exponentielle Familie mit $a'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$ und die Kenngröße $\tau(\vartheta) = \mathbb{E}_\vartheta[T]$ folgende Aussagen äquivalent:

(i) $\tau'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$.

(ii) $\mathrm{Var}_\vartheta[T] > 0$ für alle $\vartheta \in \Theta$.

(iii) $I(\vartheta) > 0$ für alle $\vartheta \in \Theta$.

Wir sehen ferner, dass T in diesem Fall Cramer-Rao-optimal ist.

Satz 1.7.8 (Cauchy-Schwarzsche Ungleichung)**.** *Ist H ein Hilbertraum, so gilt*

$$\langle x, y\rangle \leq \|x\| \cdot \|y\| \quad \textit{für alle } x, y \in H.$$

Gleichheit gilt genau dann, wenn x und y linear abhängig sind.

Bemerkung 1.7.9. *Wir werden die Cauchy-Schwarzsche Ungleichung auf $H = L^2(\Omega, \mathscr{F}, \mathbb{P})$ anwenden. Hier haben wir das Skalarprodukt*

$$\langle X, Y\rangle = \mathbb{E}[XY].$$

Die Cauchy-Schwarzsche Ungleichung besagt dann

$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}.$$

Insbesondere gilt

$$\mathrm{Cov}(X, Y) \leq \sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Y]}.$$

Satz 1.7.10 (Cramer-Rao, scharfe Version). *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein reguläres statistisches Modell. Weiterhin seien $\tau \in C^1(\Theta; \mathbb{R})$ eine Kenngröße mit $\tau'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$, und $T : \Omega \to \mathbb{R}$ ein erwartungstreuer Schätzer für τ. Dann sind folgende Aussagen äquivalent:

(i) T ist Cramer-Rao-optimal.

(ii) $\mathscr{M}$ ist eine exponentielle Familie mit kanonischer Statistik T, es gilt $a'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$ und

$$\tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)} \quad \textit{für alle } \vartheta \in \Theta.$$

Beweis. Wir der Beweis von Satz 1.6.6 zeigt, gilt wegen der Voraussetzung $\tau'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$ auch $\mathrm{Var}_\vartheta[T] > 0$ und $I(\vartheta) > 0$ für alle $\vartheta \in \Theta$. Die Gleichheit in Satz 1.6.6 gilt genau dann, wenn

$$\mathrm{Cov}_\vartheta(T, u(\vartheta, \cdot)) = \sqrt{\mathrm{Var}_\vartheta[T]} \cdot \sqrt{\mathrm{Var}_\vartheta[u(\vartheta, \cdot)]} \quad \text{für alle } \vartheta \in \Theta.$$

Nach der Cauchy-Schwarzschen Ungleichung (Satz 1.7.8) bedeutet dies, dass die Zufallsvariablen $T - \tau(\vartheta)$ und $u(\vartheta, \cdot)$ im Hilbertraum $L^2(\mathbb{P}_\vartheta)$ linear abhängig sind. Wegen $I(\vartheta) > 0$ gilt $u(\vartheta, \cdot) \neq 0$ in $L^2(\mathbb{P}_\vartheta)$, und wegen $\mathrm{Var}_\vartheta[T] > 0$ gilt $T - \tau(\vartheta) \neq 0$ in $L^2(\mathbb{P}_\vartheta)$. Also ist T genau dann Cramer-Rao-optimal, wenn eine messbare Funktion $c : \Theta \to \mathbb{R} \setminus \{0\}$ existiert, so dass

$$T - \tau(\vartheta) = c(\vartheta) u(\vartheta, \cdot) \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für alle } \vartheta \in \Theta.$$

Dies bedeutet gerade

$$\frac{T}{c(\vartheta)} - \frac{\tau(\vartheta)}{c(\vartheta)} = \frac{\partial}{\partial \vartheta} \ell(\vartheta, \cdot) \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für alle } \vartheta \in \Theta.$$

Für ein gewähltes $\vartheta_0 \in \Theta$ is dies gleichbedeutend mit

$$T(x) \int_{\vartheta_0}^{\vartheta} \frac{1}{c(s)} ds - \int_{\vartheta_0}^{\vartheta} \frac{\tau(s)}{c(s)} ds = \ell(\vartheta, x) - \ell(\vartheta_0, x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega,$$

wobei die Log-Likelihood-Funktion gegebenenfalls auf Nullmengen abgeändert wird. Die Gleichung oben bedeutet gerade, dass

$$L(\vartheta, x) = \exp\left(T(x) \int_{\vartheta_0}^{\vartheta} \frac{1}{c(s)} ds - \int_{\vartheta_0}^{\vartheta} \frac{\tau(s)}{c(s)} ds \right) \cdot L(\vartheta_0, x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega.$$

Dann ist die Likelihood-Funktion von der Form

$$L(\vartheta, x) = \exp\big(a(\vartheta)T(x) - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega,$$

wobei

$$a(\vartheta) = \int_{\vartheta_0}^{\vartheta} \frac{1}{c(s)} ds, \quad b(\vartheta) = \int_{\vartheta_0}^{\vartheta} \frac{\tau(s)}{c(s)} ds \quad \text{und} \quad h(x) = L(\vartheta_0, x).$$

Insbesondere gilt $a'(\vartheta) = \frac{1}{c(\vartheta)} \neq 0$ und

$$\tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)} \quad \text{für alle } \vartheta \in \Theta.$$

Gilt umgekehrt

$$L(\vartheta, x) = \exp\big(a(\vartheta)T(x) - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega,$$

dann gilt

$$L(\vartheta, x) = \exp\left(T(x) \int_{\vartheta_0}^{\vartheta} \frac{1}{c(s)} ds - \int_{\vartheta_0}^{\vartheta} \frac{\tau(s)}{c(s)} ds\right) \cdot L(\vartheta_0, x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega,$$

wobei

$$c(\vartheta) = \frac{1}{a'(\vartheta)} \quad \text{und} \quad \tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)}.$$

□

Falls $T \equiv t$ für ein $t \in \mathbb{R}$, so ist $\tau(\vartheta) = t$ ein Cramer-Rao-optimaler Schätzer. Satz 1.7.10 ist in dieser Situation jedoch nicht anwendbar, da $\tau'(\vartheta) = 0$ für alle $\vartheta \in \Theta$.

Korollar 1.7.11. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

eine exponentielle Familie mit kanonischer Statistik T, so dass $a'(\vartheta) \neq 0$ und $I(\vartheta) > 0$ für alle $\vartheta \in \Theta$.

(a) T ist ein Cramer-Rao-optimaler (und damit gleichmäßig bester) Schätzer für die Kenngröße

$$\tau : \Theta \to \mathbb{R}, \quad \tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)}.$$

(b) Es gilt

$$\operatorname{Var}_\vartheta[T] = \frac{\tau'(\vartheta)}{a'(\vartheta)} \quad \textit{für alle } \vartheta \in \Theta.$$

(c) Es gilt

$$I(\vartheta) = a'(\vartheta) \cdot \tau'(\vartheta) \quad \textit{für alle } \vartheta \in \Theta.$$

Beweis. Nach Satz 1.7.7 gilt $\tau(\vartheta) = \mathbb{E}_\vartheta[T]$ für alle $\vartheta \in \Theta$, und es folgen die Teile (b) und (c). Wegen $I(\vartheta) > 0$ für alle $\vartheta \in \Theta$ gilt $\tau'(\vartheta) \neq 0$ für alle $\vartheta \in \Theta$. Also folgt Teil (a) mit Satz 1.7.10. □

Beispiel 1.7.12. *Beim Binomialmodell gilt*

$$\begin{aligned} L(\vartheta, x) &= \binom{n}{x} \vartheta^x (1-\vartheta)^{n-x} \\ &= \binom{n}{x} \exp\big(x \ln \vartheta + (n-x)\ln(1-\vartheta)\big) \\ &= \binom{n}{x} \exp\left(x \ln\left(\frac{\vartheta}{1-\vartheta}\right) + n\ln(1-\vartheta)\right). \end{aligned}$$

Also gilt

$$L(\vartheta, x) = \exp\big(a(\vartheta)T(x) - b(\vartheta)\big) \cdot h(x) \quad \textit{für alle } (\vartheta, x) \in \Theta \times \Omega,$$

wobei

$$\begin{aligned} a(\vartheta) &= n \ln\left(\frac{\vartheta}{1-\vartheta}\right), \\ T(x) &= \frac{x}{n}, \\ b(\vartheta) &= -n\ln(1-\vartheta), \\ h(x) &= \binom{n}{x}. \end{aligned}$$

Weiterhin gilt

$$\begin{aligned} a'(\vartheta) &= n\left(\frac{1}{\vartheta} + \frac{1}{1-\vartheta}\right) = \frac{n}{\vartheta(1-\vartheta)}, \\ b'(\vartheta) &= \frac{n}{1-\vartheta}. \end{aligned}$$

Also ist T nach Korollar 1.7.11 ein gleichmäßig bester Schätzer für die Kenngröße

$$\tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)} = \vartheta.$$

Korollar 1.7.13. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

eine exponentielle Familie mit kanonischer Statistik T.

(a) Dann ist

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

eine exponentielle Familie mit kanonischer Statistik

$$T_n = \frac{1}{n}\sum_{i=1}^n T(X_i).$$

(b) T_n ist ein Cramer-Rao-optimaler (und damit gleichmäßig bester) Schätzer für die Kenngröße

$$\tau : \Theta \to \mathbb{R}, \quad \tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)}.$$

Beweis.

(a) Es gilt

$$L(\vartheta, x) = \exp\big(a(\vartheta)T(x) - b(\vartheta)\big)h(x).$$

Mit Lemma 1.4.13 folgt

$$\begin{aligned} L^{\otimes n}(\vartheta, x) &= \prod_{i=1}^n \exp\big(a(\vartheta)T(x_i) - b(\vartheta)\big)h(x_i) \\ &= \exp\left(a(\vartheta)\sum_{i=1}^n T(x_i) - nb(\vartheta)\right)\prod_{i=1}^n h(x_i) \\ &= \exp\big(a_n(\vartheta)T_n(x) - b_n(\vartheta)\big)h_n(x), \end{aligned}$$

wobei

$$a_n(\vartheta) = na(\vartheta), \quad b_n(\vartheta) = nb(\vartheta) \quad \text{und} \quad h_n(x) = \prod_{i=1}^n h(x_i).$$

(b) Folgt nun aus Korollar 1.7.11, da

$$\tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)} = \frac{b_n'(\vartheta)}{a_n'(\vartheta)} \quad \text{für alle } \vartheta \in \Theta.$$

□

Beispiel 1.7.14. *Beim Poisson-Modell gilt*

$$L(\vartheta, x) = e^{-\vartheta} \frac{\vartheta^x}{x!} = \exp\big(x \ln \vartheta - \vartheta\big) \frac{1}{x!}.$$

Also gilt

$$L(\vartheta, x) = \exp\big(a(\vartheta) T(x) - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega,$$

wobei

$$\begin{aligned} a(\vartheta) &= \ln \vartheta, \\ T(x) &= x, \\ b(\vartheta) &= \vartheta, \\ h(x) &= \frac{1}{x!}. \end{aligned}$$

Also ist im n-fachen Produktmodell nach Korollar 1.7.13 der Schätzer

$$T_n = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$$

ein gleichmäßig bester Schätzer für die Kenngröße

$$\tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)} = \vartheta.$$

Beispiel 1.7.15. *Beim Gaußmodell*

$$\mathbb{P}_\vartheta = \mathrm{N}(\vartheta, \sigma^2), \quad \vartheta \in \Theta = \mathbb{R}$$

mit bekannter Varianz $\sigma^2 > 0$ gilt

$$\begin{aligned} L(\vartheta, x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x - \vartheta)^2}{2\sigma^2} \Big) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{x^2}{2\sigma^2} + \frac{x\vartheta}{\sigma^2} - \frac{\vartheta^2}{2\sigma^2} \Big) \end{aligned}$$

Also gilt

$$L(\vartheta, x) = \exp\big(a(\vartheta) T(x) - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega.$$

mit

$$\begin{aligned} a(\vartheta) &= \frac{\vartheta}{\sigma^2}, \\ T(x) &= x, \\ b(\vartheta) &= \frac{\vartheta^2}{2\sigma^2}, \\ h(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{x^2}{2\sigma^2}\Big). \end{aligned}$$

Also ist im n-fachen Produktmodell nach Korollar 1.7.13 der Schätzer

$$T_n = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$$

ein gleichmäßig bester Schätzer für die Kenngröße

$$\tau(\vartheta) = \frac{b'(\vartheta)}{a'(\vartheta)} = \vartheta.$$

1.8 Suffiziente Statistiken

Definition 1.8.1. *Es seien*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell, und $S : \Omega \to E$ eine Statistik mit Werten in einem messbaren Raum $(E, \mathscr{E})$. Dann heißt S suffizient, falls für jede messbare, beschränkte Funktion $f : \Omega \to \mathbb{R}$ eine $\sigma(S)$-messbare Zufallsvariable $g : \Omega \to \mathbb{R}$ existiert, so dass

$$\mathbb{E}_\vartheta[f \mid S] = g \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für jedes } \vartheta \in \Theta.$$

Satz 1.8.2 (Faktorisierungssatz). *Es seien $Y : \Omega \to E$ eine Zufallsvariable mit Werten in einem messbaren Raum $(E, \mathscr{E})$, und $X : \Omega \to \mathbb{R}$ eine $\sigma(S)$-messbare Zufallsvariable. Dann existiert eine messbare Funktion $h : E \to \mathbb{R}$, so dass $X = h(Y)$.*

Lemma 1.8.3. *Für eine Statistik $S : \Omega \to E$ sind folgende Aussagen äquivalent:*

(i) S ist suffizient.

(ii) Für jede messbare, beschränkte Funktion $f : \Omega \to \mathbb{R}$ existiert eine messbare Funktion $h : E \to \mathbb{R}$, so dass

$$\mathbb{E}_\vartheta[f \mid S] = h(S) \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für jedes } \vartheta \in \Theta.$$

Beweis. Folgt aus dem Faktorisierungssatz (Satz 1.8.2). □

Beispiel 1.8.4. *Die Statistik $S = X : \Omega \to \Omega$ ist suffizient, da*

$$\mathbb{E}_\vartheta[f \,|\, X] = f(X) \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für jedes } \vartheta \in \Theta.$$

Lemma 1.8.5. *Es sei $(B_i)_{i\in I}$ eine Zerlegung von Ω. Für jede beschränkte Zufallsvariable X und $\mathscr{G} = \sigma(\{B_i : i \in I\})$ gilt*

$$\mathbb{E}[X \,|\, \mathscr{G}] = \sum_{\substack{i\in I \\ \mathbb{P}(B_i)>0}} \mathbb{E}[X \,|\, B_i]\mathbb{1}_{B_i} \quad \mathbb{P}\text{-fast sicher.}$$

Hierbei bezeichnet $\mathbb{E}[X \,|\, B_i]$ den Erwatungswert von X unter dem bedingten Wahrscheinlichkeitsmaß $\mathbb{P}^{B_i} = \mathbb{P}(\,\cdot \,|\, B_i)$.

Beweis. Übung. □

Lemma 1.8.6. *Es sei $S : \Omega \to E$ eine Statistik mit Werten in einem diskreten Raum E. Dann sind folgende Aussagen äquivalent:*

(i) S ist suffizient.

(ii) Für jede messbare, beschränkte Funktion $f : \Omega \to \mathbb{R}$ existiert eine Abbildung $g : E \to \mathbb{R}$, so dass

$$\mathbb{E}_\vartheta[f \,|\, S = s] = g(s)$$

für jedes $\vartheta \in \Theta$ und jedes $s \in E$ mit $\mathbb{P}_\vartheta(S = s) > 0$.

(iii) Es existieren Wahrscheinlichkeitsmaße $(\mu_s)_{s\in E}$ auf $(\Omega, \mathscr{F})$, so dass

$$\mathbb{P}_\vartheta^{\{S=s\}} = \mu_s$$

für jedes $\vartheta \in \Theta$ und jedes $s \in E$ mit $\mathbb{P}_\vartheta(S = s) > 0$.

Beweis. (i) $\Leftrightarrow$ (ii): $\{S = s\}_{s\in E}$ ist eine Zerlegung von Ω. Mit Lemma 1.8.5 folgt

$$\mathbb{E}_\vartheta[f \,|\, S] = \sum_{\substack{s\in E \\ \mathbb{P}_\vartheta(S=s)>0}} \mathbb{E}_\vartheta[f \,|\, S = s]\mathbb{1}_{\{S=s\}}.$$

(ii) $\Leftrightarrow$ (iii): Für jedes $B \subset E$ gilt

$$\mathbb{P}_\vartheta^{\{S=s\}}(B) = \mathbb{E}_\vartheta[\mathbb{1}_B \,|\, S = s]$$

Also folgt die Äquivalenz mit maßtheoretischer Induktion. □

Zur Erinnerung: Die diskrete Gleichverteilung auf einer endlichen Menge E ist die diskrete Verteilung mit stochastischem Vektor $\pi : E \to [0,1]$ gegeben durch

$$\pi(k) = \frac{1}{n}, \quad k \in E,$$

wobei $n = |E|$. Wir bezeichnen sie mit $\mathrm{UD}(E)$.

Beispiel 1.8.7. *Wir betrachten den mehrfachen Münzwurf aus Beispiel 1.1.11. Also gilt*

$$\mathbb{P}_\vartheta = \mathrm{Ber}(\vartheta)^{\otimes n}, \quad \vartheta \in \Theta = (0,1).$$

Es sei $S : \Omega = \{0,1\}^n \to \{0,1,\ldots,n\} = E$ *die Statistik*

$$S = \sum_{i=1}^{n} X_i.$$

Dann gilt (Übung)

$$\mathbb{P}_\vartheta^{\{S=k\}} = \mathrm{UD}(E_k)$$

für jedes $\vartheta \in \Theta$ *und jedes* $k \in E$*, wobei*

$$E_k = \{j \in \Omega : j_1 + \ldots + j_n = k\}.$$

Also ist S *nach Lemma 1.8.6 eine suffiziente Statistik.*

Definition 1.8.8. *Es seien* $(\Omega_1, \mathscr{F}_1)$ *und* $(\Omega_2, \mathscr{F}_2)$ *messbare Räume. Eine Abbildung* $\kappa : \Omega_1 \times \mathscr{F}_2 \to [0,1]$ *heißt ein stochastischer Kern von* $(\Omega_1, \mathscr{F}_1)$ *nach* $(\Omega_2, \mathscr{F}_2)$*, falls gilt:*

(a) $\omega_1 \mapsto \kappa(\omega_1, A_2)$ *ist* $\mathscr{F}_1$*-messbar für jedes* $A_2 \in \mathscr{F}_2$.

(b) $A_2 \mapsto \kappa(\omega_1, A_2)$ *ist ein Wahrscheinlichkeitsmaß auf* $(\Omega_2, \mathscr{F}_2)$ *für jedes* $\omega_1 \in \Omega_1$.

Definition 1.8.9. *Ein separabler topologischer Raum, dessen Topologie durch eine vollständige Metrik erzeugt wird, heißt ein polnischer Raum.*

Lemma 1.8.10. *Sofern* Ω *ein polnischer Raum ist, sind für eine Statistik* $S : \Omega \to E$ *folgende Aussagen sind äquivalent:*

(i) S *ist suffizient.*

(ii) Es existiert ein stochastischer Kern κ *von* $(\Omega, \sigma(S))$ *nach* $(\Omega, \mathscr{F})$*, so dass für jedes* $B \in \mathscr{F}$ *gilt*

$$\mathbb{P}_\vartheta(B \,|\, S) = \kappa(\cdot, B) \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für alle } \vartheta \in \Theta.$$

Lemma 1.8.11. *Es seien X, Y zwei Zufallsvariablen mit $X \in \mathscr{L}^2$. Dann gilt*

$$\mathbb{E}[X \,|\, Y]^2 \leq \mathbb{E}[X^2 \,|\, Y] \quad \mathbb{P}\text{-fast sicher,}$$

und die folgenden Aussagen sind äquivalent:

(i) Es gilt $\mathbb{E}[\mathbb{E}[X \,|\, Y]^2] = \mathbb{E}[\mathbb{E}[X^2 \,|\, Y]]$.

(ii) Es gilt $\mathbb{E}[X \,|\, Y]^2 = \mathbb{E}[X^2 \,|\, Y]$ $\mathbb{P}$-fast sicher.

(iii) Es gilt $X = \mathbb{E}[X \,|\, Y]$ $\mathbb{P}$-fast sicher.

Beweis. Die behauptete Ungleichung folgt aus der Jensen-Ungleichung für bedingte Erwartungen

$$\varphi(\mathbb{E}[X \,|\, \mathscr{G}]) \leq \mathbb{E}[\varphi(X) \,|\, \mathscr{G}]$$

mit der konvexen Funktion $\varphi(x) = x^2$.
(i) $\Rightarrow$ (ii): Es gilt

$$\mathbb{E}\big[\underbrace{\mathbb{E}[X^2 \,|\, Y] - \mathbb{E}[X \,|\, Y]^2}_{\geq 0}\big] = 0$$

Wegen $\mathbb{E}[X^2 \,|\, Y] - \mathbb{E}[X \,|\, Y]^2 \geq 0$ folgt

$$\mathbb{E}[X^2 \,|\, Y] - \mathbb{E}[X \,|\, Y]^2 = 0 \quad \mathbb{P}\text{-fast sicher.}$$

(ii) $\Rightarrow$ (iii): Es gilt

$$\begin{aligned}\mathbb{E}\big[\underbrace{(X - \mathbb{E}[X \,|\, Y])^2}_{\geq 0} \,|\, Y\big] &= \mathbb{E}\big[X^2 - 2X\mathbb{E}[X \,|\, Y] + \mathbb{E}[X \,|\, Y]^2 \,|\, Y\big] \\ &= \mathbb{E}[X^2 \,|\, Y] - 2\mathbb{E}[X \,|\, Y]^2 + \mathbb{E}[X \,|\, Y]^2 = 0\end{aligned}$$

Also folgt $X - \mathbb{E}[X \,|\, Y] = 0$ $\mathbb{P}$-fast sicher.
(iii) $\Rightarrow$ (i): ✓ □

Satz 1.8.12 (Satz von Rao-Blackwell)**.** *Es seien*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell, $\tau : \Theta \to \mathbb{R}$ eine Kenngröße, $T : \Omega \to \mathbb{R}$ ein erwartungstreuer Schätzer für τ mit $T \in \mathscr{L}^2(\mathbb{P}_\vartheta)$ für alle $\vartheta \in \Theta$, und $S : \Omega \to E$ eine suffiziente Statistik. Wir definieren den Schätzer $T' : \Omega \to \mathbb{R}$ durch

$$T' := \mathbb{E}[T \,|\, S].$$

(a) T' *ist ein erwartungstreuer Schätzer für* τ*.*

(b) *Es gilt*

$$\mathrm{Var}_\vartheta[T'] \leq \mathrm{Var}_\vartheta[T] \quad \textit{für alle } \vartheta \in \Theta.$$

(c) *Es gilt*

$$\mathrm{Var}_\vartheta[T'] = \mathrm{Var}_\vartheta[T] \quad \textit{für alle } \vartheta \in \Theta$$

genau dann, wenn

$$T = T' \quad \mathbb{P}_\vartheta\textit{-fast sicher} \quad \textit{für jedes } \vartheta \in \Theta.$$

Beweis.

(a) Für jedes $\vartheta \in \Theta$ gilt

$$\mathbb{E}_\vartheta[T'] = \mathbb{E}_\vartheta[\mathbb{E}_\vartheta[T \mid S]] = \mathbb{E}_\vartheta[T] = \tau(\vartheta).$$

(b) Für jedes $\vartheta \in \Theta$ gilt mit der Jensen-Ungleichung (Lemma 1.8.11)

$$\begin{aligned}
\mathrm{Var}_\vartheta[T'] &= \mathbb{E}_\vartheta[(T' - \tau(\vartheta))^2] \\
&= \mathbb{E}_\vartheta[(\mathbb{E}_\vartheta[T \mid S] - \tau(\vartheta))^2] \\
&= \mathbb{E}_\vartheta[\mathbb{E}_\vartheta[T - \tau(\vartheta) \mid S]^2] \\
&\leq \mathbb{E}_\vartheta[\mathbb{E}_\vartheta[(T - \tau(\vartheta))^2 \mid S]] \\
&= \mathbb{E}_\vartheta[(T - \tau(\vartheta))^2] = \mathrm{Var}_\vartheta[T].
\end{aligned}$$

(c) Es sei $\vartheta \in \Theta$ beliebig. Nach der Rechung aus Teil (b) gilt

$$\mathrm{Var}_\vartheta[T] = \mathrm{Var}_\vartheta[T']$$

genau dann, wenn

$$\mathbb{E}_\vartheta[\mathbb{E}_\vartheta[T - \tau(\vartheta) \mid S]^2] = \mathbb{E}_\vartheta[\mathbb{E}_\vartheta[(T - \tau(\vartheta))^2 \mid S]].$$

Nach Lemma 1.8.11 ist dies äquivalent zu

$$T - \tau(\vartheta) = \mathbb{E}_\vartheta[T - \tau(\vartheta) \mid S] \quad \mathbb{P}_\vartheta\text{-fast sicher},$$

und dies ist gleichbedeutend mit

$$T = \mathbb{E}_\vartheta[T \mid S] = T' \quad \mathbb{P}_\vartheta\text{-fast sicher}.$$

$\square$

Bemerkung 1.8.13. *In der Situation des Satzes von Rao-Blackwell (Satz 1.8.12) existiert nach dem Faktorisierungssatz (Satz 1.8.2) eine messbare Funktion $h : E \to \mathbb{R}$, so dass*

$$T' = h(S).$$

Lemma 1.8.14. *Es sei $X : \Omega \to G$ und $Y : \Omega \to E$ diskrete Zufallsvariablen, und $g : G \to \mathbb{R}$ eine Abbildung. Wir definieren*

$$h : E \to \mathbb{R}, \quad h(k) := \sum_{j \in G} g(j)\mathbb{P}(X = j \,|\, Y = k)$$

Dann gilt

$$\mathbb{E}[g(X) \,|\, Y] = h(Y).$$

Beispiel 1.8.15. *Wir betrachten den mehrfachen Münzwurf aus Beispiel 1.1.11. Also gilt*

$$\mathbb{P}_\vartheta = \mathrm{Ber}(\vartheta)^{\otimes n}, \quad \vartheta \in \Theta = (0,1).$$

Dann ist $T : \Omega = \{0,1\}^n \to \{0,1\} = G$ definiert durch

$$T = X_1$$

ein erwartungstreuer Schätzer für die Kenngröße $\tau : \Theta \to \mathbb{R}$, $\tau(\vartheta) = \vartheta$. Nach Beispiel 1.8.7 ist $S : \Omega \to \{0,1,\ldots,n\} = E$ gegeben durch

$$S = \sum_{i=1}^{n} X_i$$

eine suffiziente Statistik. Wir setzen

$$T' = \mathbb{E}[T \,|\, S].$$

Dann gilt

$$T' = \frac{S}{n} = \bar{X}.$$

In der Tat, es sei $\vartheta \in \Theta$ beliebig. Dann gilt $T \sim \mathrm{Ber}(\vartheta)$, $S \sim \mathrm{Bi}(n,\vartheta)$ und $X_2 + \ldots + X_n \sim \mathrm{Bi}(n-1,\vartheta)$ unter $\mathbb{P}_\vartheta$. Nun sei $k \in E$ beliebig. Dann gilt

$$\begin{aligned}\{T = 1\} \cap \{S = k\} &= \{X_1 = 1\} \cap \{X_1 + \ldots + X_n = k\} \\ &= \{X_1 = 1\} \cap \{X_2 + \ldots + X_n = k-1\},\end{aligned}$$

und daher

$$\begin{aligned}\mathbb{P}_\vartheta(T=1\,|\,S=k) &= \frac{\mathbb{P}_\vartheta(\{T=1\}\cap\{S=k\})}{\mathbb{P}_\vartheta(S=k)}\\ &= \frac{\mathbb{P}_\vartheta(X_1=1)\cdot\mathbb{P}_\vartheta(X_2+\ldots+X_n=k-1)}{\mathbb{P}_\vartheta(S=k)}\\ &= \frac{\vartheta\cdot\binom{n-1}{k-1}\vartheta^{k-1}(1-\vartheta)^{n-k}}{\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}} = \frac{\frac{(n-1)!}{(k-1)!(n-k)!}}{\frac{n!}{k!(n-k)!}} = \frac{k}{n}.\end{aligned}$$

Die Abbildung

$$h: E\to\mathbb{R},\quad h(k) := \sum_{j\in G} g(j)\mathbb{P}(T=j\,|\,S=k)$$

mit $g(j)=j$ ist also gegeben durch

$$h(k) = \sum_{j=0,1} j\mathbb{P}(T=j\,|\,S=k) = \mathbb{P}(T=1\,|\,S=k) = \frac{k}{n}.$$

Mit Lemma 1.8.14 folgt also

$$T' = \frac{S}{n} = \bar{X}.$$

In der Tat gilt

$$\mathrm{Var}_\vartheta[T'] = \frac{\vartheta(1-\vartheta)}{n} < \vartheta(1-\vartheta) = \mathrm{Var}_\vartheta[T] \quad \textit{für alle } \vartheta\in\Theta.$$

Satz 1.8.16 (Faktorisierungslemma von Neyman-Fischer)**.** *Es seien*

$$\mathscr{M} = \big(\Omega,\mathscr{F},(\mathbb{P}_\vartheta:\vartheta\in\Theta)\big)$$

ein Standardmodell und $S:\Omega\to E$ eine Statistik. Dann sind folgende Aussagen äquivalent:

(i) S ist suffizient.

(ii) Es existieren $g:\Theta\times E\to\mathbb{R}_+$ und eine messbare Funktion $h:\Omega\to\mathbb{R}_+$, so dass

$$L(\vartheta,x) = g(\vartheta,S(x))\cdot h(x) \quad \textit{für alle } (\vartheta,x)\in\Theta\times\Omega.$$

Beweis. Für den Fall, dass Ω und E diskret sind, und μ das Zählmaß ζ ist.
(i) $\Rightarrow$ (ii): Wir definieren $g : \Theta \times E \to \mathbb{R}_+$ durch

$$g(\vartheta, s) := \mathbb{P}_\vartheta(S = s).$$

Nach Lemma 1.8.6 existieren Wahrscheinlichkeitsmaße $(\mu_s)_{s\in E}$ auf $(\Omega, \mathscr{F})$, so dass

$$\mathbb{P}_\vartheta^{\{S=s\}} = \mu_s$$

für jedes $\vartheta \in \Theta$ und jedes $s \in E$ mit $\mathbb{P}_\vartheta(S = s) > 0$. Wir definieren $h_0 : \Omega \times E \to \mathbb{R}_+$ durch

$$h_0(x, s) := \mu_s(\{x\}),$$

und $h : \Omega \to \mathbb{R}_+$ durch

$$h(x) := h_0(x, S(x)).$$

Dann gilt

$$\begin{aligned} L(\vartheta, x) &= \mathbb{P}_\vartheta(X = x) = \mathbb{P}_\vartheta(X = x, S = S(x)) \\ &= \mathbb{P}_\vartheta(X = x \,|\, S = S(x)) \cdot \mathbb{P}_\vartheta(S = S(x)) \\ &= \mu_{S(x)}(\{x\}) \cdot g(\vartheta, S(x)) = h_0(x, S(x)) \cdot g(\vartheta, S(x)) = h(x) \cdot g(\vartheta, S(x)). \end{aligned}$$

(ii) $\Rightarrow$ (i): Wir definieren die Wahrscheinlichkeitsmaße $(\mu_s)_{s\in E}$ auf $(\Omega, \mathscr{F})$ durch

$$\mu_s(\{x\}) = \frac{h(x)}{\sum_{\substack{y\in E \\ S(y)=s}} h(y)}$$

für alle $x \in \Omega$ mit $S(x) = s$. Es seien $\vartheta \in \Theta$ und $s \in E$ mit $\mathbb{P}_\vartheta(S = s) > 0$ beliebig. Dann gilt

$$\begin{aligned} \mathbb{P}_\vartheta(S = s) &= \sum_{\substack{y\in E \\ S(y)=s}} \mathbb{P}_\vartheta(X = y) = \sum_{\substack{y\in E \\ S(y)=s}} L(\vartheta, y) \\ &= \sum_{\substack{y\in E \\ S(y)=s}} g(\vartheta, S(y)) \cdot h(y) = g(s, \vartheta) \sum_{\substack{y\in E \\ S(y)=s}} h(y), \end{aligned}$$

was auch die Wohldefiniertheit von μ_s zeigt. Nun sei $x \in \Omega$ mit $S(x) = s$ beliebig. Dann gilt

$$\mathbb{P}_\vartheta(X = x, S = s) = \mathbb{P}_\vartheta(X = x) = L(\vartheta, x) = g(\vartheta, S(x)) \cdot h(x) = g(\vartheta, s) \cdot h(x).$$

Insgesamt folgt

$$\mathbb{P}_\vartheta(X = x \,|\, S = s) = \frac{\mathbb{P}_\vartheta(X = x, S = s)}{\mathbb{P}_\vartheta(S = s)} = \frac{g(\vartheta, s) \cdot h(x)}{g(\vartheta, s) \sum_{\substack{y \in E \\ S(y)=s}} h(y)}$$
$$= \frac{h(x)}{\sum_{\substack{y \in E \\ S(y)=s}} h(y)} = \mu_s(\{x\}).$$

Nach Lemma 1.8.6 ist S eine suffiziente Statistik. □

Korollar 1.8.17. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

eine exponentielle Familie mit kanonischer Statistik T. Dann ist T suffizient.

Beweis. Dies folgt aus dem Faktorisierungslemma von Neyman-Fischer (Satz 1.8.16) mit

$$g : \Theta \times \mathbb{R} \to \mathbb{R}_+, \quad g(\vartheta, t) = \exp(a(\vartheta)t - b(\vartheta)).$$

□

1.9 Vollständige Statistiken

Definition 1.9.1. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell. Eine Statistik $S : \Omega \to E$ heißt vollständig, falls für jede messbare Funktion $h : E \to \mathbb{R}$ mit $h(S) \in \mathscr{L}^2(\mathbb{P}_\vartheta)$ für alle $\vartheta \in \Theta$ aus der Bedingung

$$\mathbb{E}_\vartheta[h(S)] = 0 \quad \textit{für alle } \vartheta \in \Theta$$

folgt

$$h(S) = 0 \quad \mathbb{P}_\vartheta\textit{-fast sicher} \quad \textit{für alle } \vartheta \in \Theta.$$

Beispiel 1.9.2. *Wir betrachten den mehrfachen Münzwurf aus Beispiel 1.1.11. Also gilt*

$$\mathbb{P}_\vartheta = \mathrm{Ber}(\vartheta)^{\otimes n}, \quad \vartheta \in \Theta = (0, 1).$$

Es sei $S : \Omega = \{0,1\}^n \to \{0,1,\ldots,n\} = E$ *die Statistik*

$$S = \sum_{i=1}^{n} X_i.$$

Dann ist S *vollständig. In der Tat, es sei* $h : E \to \mathbb{R}$ *eine beliebige Funktion. Weiterhin sei* $\vartheta \in \Theta$ *beliebig. Dann gilt* $S \sim \mathrm{Bi}(n,\vartheta)$ *unter* $\mathbb{P}_\vartheta$*, und es folgt*

$$\begin{aligned}\mathbb{E}_\vartheta[h(S)] &= \sum_{k=0}^{n} h(k)\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k} \\ &= (1-\vartheta)^n \sum_{k=0}^{n} h(k)\binom{n}{k}\left(\frac{\vartheta}{1-\vartheta}\right)^k.\end{aligned}$$

Also folgt aus

$$\mathbb{E}_\vartheta[h(S)] = 0 \quad \textit{für alle } \vartheta \in \Theta,$$

dass

$$\sum_{k=0}^{n} h(k)\binom{n}{k} y^k = 0 \quad \textit{für alle } y \in (0,\infty),$$

und damit

$$h(k) = 0 \quad \textit{für alle } k \in E.$$

Satz 1.9.3 (Satz von Lehmann-Scheffé)**.** *Es seien*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell, $\tau : \Theta \to \mathbb{R}$ *eine Kenngröße,* $T : \Omega \to \mathbb{R}$ *ein erwartungstreuer Schätzer für* τ *mit* $T \in \mathscr{L}^2(\mathbb{P}_\vartheta)$ *für alle* $\vartheta \in \Theta$*, und* $S : \Omega \to E$ *eine vollständige, suffiziente Statistik.*

(a) *Es gibt einen eindeutig bestimmten gleichmäßig besten Schätzer* T' *für die Kenngröße* τ*.*

(b) *Dieser ist gegeben durch*

$$T' := \mathbb{E}[T \,|\, S].$$

(c) *Die Definition von* T' *ist unabhängig von der Wahl von* T*; das heißt, für jeden weiteren erwartungstreuen Schätzer* $U : \Omega \to \mathbb{R}$ *für* τ *mit* $U \in \mathscr{L}^2(\mathbb{P}_\vartheta)$ *für alle* $\vartheta \in \Theta$ *gilt*

$$T' = \mathbb{E}[U \,|\, S].$$

(d) Es existiert ein messbare Funktion $h : E \to \mathbb{R}$, so dass

$$T' = h(S).$$

Beweis. Nach dem Satz von Rao-Blackwell (Satz 1.8.12(a), (b)) ist T' ein erwartungstreuer Schätzer für τ, und es gilt

$$\operatorname{Var}_\vartheta[T'] \le \operatorname{Var}_\vartheta[T] \quad \text{für alle } \vartheta \in \Theta.$$

Nun sei U ein weiterer erwartungstreuer Schätzer wie in Teil (c). Wir setzen

$$U' := \mathbb{E}[U \,|\, S].$$

Nach dem Faktorisierungssatz (Satz 1.8.2) existieren messbare Funktionen $h_i : E \to \mathbb{R}$ für $i = 1, 2$, so dass

$$T' = h_1(S) \quad \text{und} \quad U' = h_2(S).$$

Wir definieren $h : E \to \mathbb{R}$ durch $h := h_1 - h_2$. Dann gilt $h(S) \in \mathscr{L}^2(\mathbb{P}_\vartheta)$ für alle ϑ, da $T', U' \in \mathscr{L}^2(\mathbb{P}_\vartheta)$ für alle ϑ. Weiterhin gilt für alle $\vartheta \in \Theta$

$$\begin{aligned}\mathbb{E}_\vartheta[h(S)] &= \mathbb{E}_\vartheta[h_1(S)] - \mathbb{E}_\vartheta[h_2(S)] = \mathbb{E}_\vartheta[T'] - \mathbb{E}_\vartheta[U'] \\ &= \tau(\vartheta) - \tau(\vartheta) = 0.\end{aligned}$$

Wegen der Vollständigkeit von S folgt für jedes $\vartheta \in \Theta$

$$\mathbb{P}_\vartheta(T' = U') = \mathbb{P}_\vartheta(h_1(S) = h_2(S)) = \mathbb{P}_\vartheta(h(S) = 0) = 1.$$

Also hängt die Definition von T' nicht von der Wahl von T ab, und es gilt insbesondere $T' = \mathbb{E}[U \,|\, S]$ für jeden erwartungstreuen Schätzer U für τ. Folglich gilt Nach dem Satz von Rao-Blackwell (Satz 1.8.12(b))

$$\operatorname{Var}_\vartheta[T'] \le \operatorname{Var}_\vartheta[U] \quad \text{für alle } \vartheta \in \Theta$$

und jeden erwartungstreuen Schätzer U für τ. Also ist T' ein gleichmäßig bester Schätzer für die Kenngröße τ.

Für die Eindeutigkeit sei U ein weiterer gleichmäßig bester Schätzer für τ. Dann gilt

$$\operatorname{Var}_\vartheta[T'] = \operatorname{Var}_\vartheta[U] \quad \text{für alle } \vartheta \in \Theta.$$

Außerdem gilt $T' = \mathbb{E}[U \,|\, S]$, und nach dem Satz von Rao-Blackwell (Satz 1.8.12(c)) folgt

$$U = T' \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für jedes } \vartheta \in \Theta.$$

Dies beweist (a)–(c). Teil (d) folgt aus dem Faktorisierungssatz (Satz 1.8.2). $\square$

Korollar 1.9.4. *Es seien $S : \Omega \to E$ eine vollständige, suffiziente Statistik, und $h : E \to \mathbb{R}$ eine messbare Funktion. Dann ist $T : \Omega \to \mathbb{R}$ gegeben durch $T = h(S)$ der eindeutig bestimmte gleichmäßig beste Schätzer für die Kenngröße*

$$\tau : \Theta \to \mathbb{R}, \quad \tau(\vartheta) = \mathbb{E}_\vartheta[h(S)].$$

Beweis. Folgt aus dem Satz von Lehmann-Scheffé (Satz 1.9.3). □

Beispiel 1.9.5. *Wir betrachten den mehrfachen Münzwurf aus Beispiel 1.1.11. Also gilt*

$$\mathbb{P}_\vartheta = \mathrm{Ber}(\vartheta)^{\otimes n}, \quad \vartheta \in \Theta = (0,1).$$

Es sei $S : \Omega = \{0,1\}^n \to \{0,1,\dots,n\} = E$ die Statistik

$$S = \sum_{i=1}^n X_i.$$

Wie wir in den Beispielen 1.8.7 und 1.9.2 gesehen haben, ist S eine vollständige, suffiziente Statistik. Die Statistik $T : \Omega \to \{0,1\}$ definiert durch

$$T = X_1$$

ist ein erwartungstreuer Schätzer für die Kenngröße $\tau : \Theta \to \mathbb{R}$, $\tau(\vartheta) = \vartheta$, und nach Beispiel 1.8.15 ist

$$T' := \mathbb{E}[T \mid S]$$

gegeben durch

$$T' = \frac{S}{n} = \bar{X}.$$

Nach dem Satz von Lehmann-Scheffé (Satz 1.9.3) ist $\bar{X}$ der eindeutig bestimmte gleichmäßig beste Schätzer für die Kenngröße τ.

1.10 Mehrdimensionale exponentielle Familien

Definition 1.10.1. *Ein Standardmodell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

heißt ein mehrdimensionales reguläres Modell, falls gilt:

(a) $\Theta \subset \mathbb{R}^d$ *ist eine nichtleere, offene Menge.*

(b) $\mathbb{P}_\vartheta \approx \mu$ *für alle* $\vartheta \in \Theta$.

(c) Für jeden Schätzer $S : \Omega \to \mathbb{R}$ *und jedes* $\vartheta \in \Theta$ *mit* $S \in \mathscr{L}^1(\mathbb{P}_\vartheta) < \infty$ *gilt*

$$\frac{\partial}{\partial \vartheta_i} \int_\Omega S(x) L(\vartheta, x) \mu(dx) = \int_\Omega S(x) \frac{\partial}{\partial \vartheta_i} L(\vartheta, x) \mu(dx) \quad \text{für alle } i = 1, \ldots, d.$$

Definition 1.10.2. *Ein mehrdimensionales reguläres Modell*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

heißt eine mehrdimensionale exponentielle Familie, falls messbare Funktionen $a : \Theta \to \mathbb{R}^d$ *mit* $a \in C^1(\Theta; \mathbb{R}^d)$, $b : \Theta \to \mathbb{R}$, *eine Statistik* $T : \Omega \to \mathbb{R}^d$ *und eine messbare Funktion* $h : \Omega \to (0, \infty)$ *existieren, so dass*

$$L(\vartheta, x) = \exp\big(\langle a(\vartheta), T(x)\rangle_{\mathbb{R}^d} - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega.$$

In diesem Fall nennen wir T *eine kanonische Statistik.*

Lemma 1.10.3. *Die Statistik* T *ist suffizient.*

Beweis. Dies folgt aus dem Faktorisierungslemma von Neyman-Fischer (Satz 1.8.16) mit

$$g : \Theta \times \mathbb{R}^d \to \mathbb{R}_+, \quad g(\vartheta, t) = \exp\big(\langle a(\vartheta), t\rangle_{\mathbb{R}^d} - b(\vartheta)\big).$$

□

Lemma 1.10.4. *Die Funktion* $b : \Theta \to \mathbb{R}$ *ist gegeben durch*

$$b(\vartheta) = \ln\left(\int_\Omega \exp\big(\langle a(\vartheta), T(x)\rangle_{\mathbb{R}^d}\big) h(x) \mu(dx)\right).$$

Beweis. Wie der Beweis von Lemma 1.7.4. □

Bemerkung 1.10.5. *Es gelte* $\det J_a(\vartheta_0) \neq 0$ *für ein* $\vartheta_0 \in \Theta$. *Hierbei bezeichnet* $J_a(\vartheta_0) \in \mathbb{R}^{d \times d}$ *die Jakobi-Matrix*

$$J_a(\vartheta_0) = \begin{pmatrix} \frac{\partial a_1}{\partial \vartheta_1}(\vartheta_0) & \cdots & \frac{\partial a_1}{\partial \vartheta_d}(\vartheta_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial a_d}{\partial \vartheta_1}(\vartheta_0) & \cdots & \frac{\partial a_d}{\partial \vartheta_d}(\vartheta_0) \end{pmatrix}.$$

Nach dem lokalen Umkehrsatz existiert eine offene Umgebung $U \subset \mathbb{R}^d$ *von* ϑ_0, *so dass* a *auf* U *injektiv ist, und es gilt*

$$L(a^{-1}(\eta), x) = \exp\big(\langle \eta, T(x)\rangle_{\mathbb{R}^d} - \bar{b}(\eta)\big) \cdot h(x) \quad \text{für alle } \eta \in a(U),$$

wobei $\bar{b} = b \circ a^{-1}$.

Lemma 1.10.6. *Das n-fache Produktmodell*

$$\mathscr{M}^{\otimes n} = \big(\Omega^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n} : \vartheta \in \Theta)\big)$$

ist eine exponentielle Familie der Form

$$L^{\otimes n}(\vartheta, x) = \exp\big(\langle a_n(\vartheta), T_n(x)\rangle_{\mathbb{R}^d} - b_n(\vartheta)\big) \cdot h_n(x)$$

mit

$$a_n(\vartheta) = n \cdot a(\vartheta), \quad b_n(\vartheta) = n \cdot b(\vartheta), \quad h_n(x) = \prod_{i=1}^n h(x_i)$$

und kanonischer Statistik

$$T_n = \frac{1}{n}\sum_{i=1}^n T(X_i).$$

Beweis. Nach Lemma 1.4.13(a) gilt

$$\begin{aligned} L^{\otimes n}(\vartheta, x) &= \prod_{i=1}^n \exp\big(\langle a(\vartheta), T(x_i)\rangle_{\mathbb{R}^d} - b(\vartheta)\big) \cdot h(x_i) \\ &= \exp\left(\left\langle a(\vartheta), \sum_{i=1}^n T(x_i)\right\rangle_{\mathbb{R}^d} - n \cdot b(\vartheta)\right) \cdot \prod_{i=1}^n h(x_i) \\ &= \exp\big(\langle a_n(\vartheta), T_n(x)\rangle_{\mathbb{R}^d} - b_n(\vartheta)\big) \cdot h_n(x). \end{aligned}$$

□

Beispiel 1.10.7. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

das Gauß-Modell aus Beispiel 1.1.3, wobei $\Theta = \mathbb{R} \times (0, \infty)$. Es gilt also

$$\mathbb{P}_\vartheta = \mathrm{N}(\mu, \sigma^2), \quad (\mu, \sigma^2) = \vartheta \in \Theta.$$

Dann gilt

$$\begin{aligned} L(\vartheta, x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2}\right) \\ &= \exp\left(- \frac{1}{2}\ln(2\pi\sigma^2) - \frac{x^2}{2\sigma^2} + \frac{x\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2}\right). \end{aligned}$$

Also gilt

$$L(\vartheta, x) = \exp\big(\langle a(\vartheta), T(x)\rangle_{\mathbb{R}^2} - b(\vartheta)\big) \cdot h(x) \quad \textit{für alle } (\vartheta, x) \in \Theta \times \Omega,$$

wobei

$$\begin{aligned} a(\vartheta) &= \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right), \\ b(\vartheta) &= \frac{1}{2}\left(\frac{\mu^2}{\sigma^2} + \ln(2\pi\sigma^2)\right), \\ T(x) &= (x, x^2), \\ h(x) &= 1. \end{aligned}$$

Für alle $\vartheta \in \Theta$ gilt also

$$J_a(\vartheta) = \begin{pmatrix} \frac{1}{\sigma^2} & -\frac{\mu}{\sigma^4} \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix},$$

und daher

$$\det J_a(\vartheta) = \frac{1}{2\sigma^6} > 0.$$

Das n-fache Produktmodell hat nach Lemma 1.10.6 die kanonische Statistik

$$T_n = \left(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n X_i^2\right).$$

Satz 1.10.8 (Eindeutigkeitssatz für die Laplace-Transformation)**.** *Es seien μ und ν zwei endliche Maße auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, und es seien $\vartheta_0 \in \mathbb{R}^d$ und $U \subset \mathbb{R}^d$ eine offene Umgebung von ϑ_0, so dass*

$$\int_{\mathbb{R}^d} \exp(\langle \vartheta, x\rangle_{\mathbb{R}^d})\mu(dx) = \int_{\mathbb{R}^d} \exp(\langle \vartheta, x\rangle_{\mathbb{R}^d})\nu(dx) < \infty \quad \textit{für alle } \vartheta \in U.$$

Dann gilt $\mu = \nu$.

Lemma 1.10.9. *Es seien $(\Omega, \mathscr{F}, \mu)$ ein Maßraum, und $T : \Omega \to E$ eine messbare Abbildung mit Werten in einem messbaren Raum $(E, \mathscr{E})$. Wir setzen $\nu := \mu \circ T$. Weiterhin seien $f, g : E \to \mathbb{R}$ messbare Funktionen, so dass $f = g$ ν-fast sicher. Dann gilt $f(T) = g(T)$ μ-fast sicher.*

Beweis. Es gibt eine ν-Nullmenge $N \in \mathscr{E}$, so dass $f(t) = g(t)$ für alle $t \in N^c$. Es gilt

$$\mu(T^{-1}(N)) = \nu(N) = 0.$$

Also ist $T^{-1}(N)$ eine μ-Nullmenge. Außerdem gilt

$$f(T(x)) = g(T(x)) \quad \text{für alle } x \in T^{-1}(N^c) = T^{-1}(N)^c.$$

$\square$

Satz 1.10.10. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

eine mehrdimensionale exponentielle Familie, so dass eine nichtleere, offene Menge $V \subset \mathbb{R}^d$ mit $V \subset a(\Theta)$ existiert. Dann ist die kanonische Statistik T vollständig und suffizient.

Zwischenbemerkung: Es gelte $\det J_a(\vartheta_0) \neq 0$ für ein $\vartheta_0 \in \Theta$. Nach dem lokalen Umkehrsatz existiert eine offene Umgebung $U \subset \mathbb{R}^d$ von ϑ_0, so dass $a|_U : U \to a(U)$ ein Homöomorphismus ist. Also ist $V := a(U)$ eine nichtleere, offene Menge mit $V \subset a(\Theta)$.

Beweis. Die Suffizienz folgt aus Lemma 1.10.3. Es sei $f : \mathbb{R}^d \to \mathbb{R}$ eine messbare Funktion mit $f(T) \in \mathscr{L}^2(\mathbb{P}_\vartheta)$ für alle $\vartheta \in \Theta$, so dass

$$\mathbb{E}_\vartheta[f(T)] = 0 \quad \text{für alle } \vartheta \in \Theta.$$

Wie in Bemerkung 1.10.5 gilt

$$L(a^{-1}(\eta), x) = \exp\big(\langle \eta, T(x)\rangle_{\mathbb{R}^d} - \bar{b}(\eta)\big) \cdot h(x) \quad \text{für alle } (\eta, x) \in V \times \Omega,$$

wobei $\bar{b} = b \circ a^{-1}$. Da $a^{-1}(V)$ auch offen ist, dürfen wir also annehmen, dass

$$L(\vartheta, x) = \exp\big(\langle \vartheta, T(x)\rangle_{\mathbb{R}^d} - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in V \times \Omega.$$

Durch Verschiebung dürfen wir außerdem annehmen, dass $0 \in V$. Genauer gesagt, wählen wir $\vartheta_0 \in V$ und eine offene Umgebung $W \subset \mathbb{R}^d$ der Null, so dass $\vartheta_0 + W \subset V$. Dann gilt für alle $\vartheta \in W$

$$\begin{aligned} L(\vartheta_0 + \vartheta, x) &= \exp\big(\langle \vartheta_0 + \vartheta, T(x)\rangle_{\mathbb{R}^d} - b(\vartheta_0 + \vartheta)\big) \cdot h(x) \\ &= \exp\big(\langle \vartheta, T(x)\rangle_{\mathbb{R}^d} - \bar{b}(\vartheta)\big) \cdot \bar{h}(x), \end{aligned}$$

wobei

$$\begin{aligned} &\bar{b}(\vartheta) = b(\vartheta_0 + \vartheta), \\ &\bar{h}(x) = \exp\big(\langle \vartheta_0, T(x)\rangle_{\mathbb{R}^d}\big) \cdot h(x). \end{aligned}$$

Wir zerlegen $f = f^+ - f^-$ in Positiv- und Negativteil. Wir definieren die Maße κ durch $\frac{d\kappa}{d\mu} := h$ und $\nu := \kappa \circ T$. Weiterhin definieren wir die Maße ν_+ und ν_- durch

$\frac{d\nu_+}{d\nu} := f_+$ und $\frac{d\nu_-}{d\nu} := f_-$. Dann gilt für alle $\vartheta \in V$

$$\begin{aligned}\mathbb{E}_\vartheta[f_+(T)] &= \int_\Omega L(\vartheta, x) f_+(T(x))\mu(dx) \\ &= \int_\Omega \exp\big(\langle \vartheta, T(x)\rangle_{\mathbb{R}^d} - b(\vartheta)\big) h(x) f_+(T(x))\mu(dx) \\ &= \exp(-b(\vartheta)) \int_\Omega \exp\big(\langle \vartheta, T(x)\rangle_{\mathbb{R}^d}\big) f_+(T(x))\kappa(dx) \\ &= \exp(-b(\vartheta)) \int_{\mathbb{R}^d} \exp\big(\langle \vartheta, t\rangle_{\mathbb{R}^d}\big) f_+(t)\nu(dt) \\ &= \exp(-b(\vartheta)) \int_{\mathbb{R}^d} \exp\big(\langle \vartheta, t\rangle_{\mathbb{R}^d}\big) \nu_+(dt).\end{aligned}$$

Analog gilt

$$\mathbb{E}_\vartheta[f_-(T)] = \exp(-b(\vartheta)) \int_{\mathbb{R}^d} \exp\big(\langle \vartheta, t\rangle_{\mathbb{R}^d}\big) \nu_-(dt),$$

und folglich

$$\int_{\mathbb{R}^d} \exp\big(\langle \vartheta, t\rangle_{\mathbb{R}^d}\big) \nu_+(dt) = \int_{\mathbb{R}^d} \exp\big(\langle \vartheta, t\rangle_{\mathbb{R}^d}\big) \nu_-(dt) \quad \text{für alle } \vartheta \in V.$$

Die Maße ν_+ und ν_- sind endlich, denn mit $\vartheta = 0$ folgt

$$\nu_+(\mathbb{R}^d) = \exp(b(0))\mathbb{E}_\vartheta[f_+(T)] < \infty,$$

und analog $\nu_-(\mathbb{R}^d) < \infty$. Nach dem Eindeutigkeitssatz für die Laplace-Transformation (Satz 1.10.8) folgt $\nu_+ = \nu_-$. Also gilt $f_+ = f_-$ ν-fast sicher, und mit Lemma 1.10.9 folgt $f_+(T) = f_-(T)$ κ-fast sicher. Wegen $\kappa \approx \mu \approx \mathbb{P}_\vartheta$ folgt

$$f(T) = 0 \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für alle } \vartheta \in U.$$

□

Beispiel 1.10.11. *Wir betrachten das n-fache Gaußmodell. Nach Satz 1.10.10 und Beispiel 1.10.7 ist*

$$S = \Big(\frac{1}{n}\sum_{i=1}^n X_i, \frac{1}{n}\sum_{i=1}^n X_i^2\Big).$$

eine vollständige, suffiziente Statistik. Wir definieren $h : \mathbb{R}^2 \to \mathbb{R}^2$ durch

$$h(s) := \Big(s_1, \frac{n}{n-1}(s_2 - s_1^2)\Big).$$

Dann gilt

$$h(S) = \left(\widehat{m}_1(X), \frac{n}{n-1}\big(\widehat{m}_2(X) - \widehat{m}_1(X)^2\big)\right) = \left(\bar{X}, \frac{n}{n-1}\widehat{\sigma}^2(X)\right) = \big(\bar{X}, s^2(X)\big).$$

Nach Satz 1.4.17 gilt außderdem

$$\mathbb{E}_\vartheta[h(S)] = \vartheta \quad \textit{für alle } \vartheta \in \Theta.$$

Also ist die Statistik $(\bar{X}, s^2(X))$ *nach Korollar 1.9.4 der eindeutig bestimmte gleichmäßig beste Schätzer für die Kenngröße*

$$\tau : \Theta \to \mathbb{R}^2, \quad \tau(\vartheta) = \vartheta.$$

1.11 Die mehrdimensionale Informationsungleichung

Definition 1.11.1. *Es sei*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein mehrdimensionales reguläres statistisches Model mit $\Theta \subset \mathbb{R}^d$*.*

(a) *Wir nennen*

$$u : \Theta \times \Omega \to \mathbb{R}^d, \quad u(\vartheta, x) = \nabla_\vartheta \ell(\vartheta, x) = \frac{\nabla_\vartheta L(\vartheta, x)}{L(\vartheta, x)}$$

die Einflussfunktion (oder Score-Funktion).

(b) *Wir nennen*

$$I : \Theta \to \mathbb{R}^{d\times d}, \quad I(\vartheta) = \mathbb{E}_\vartheta[u(\vartheta, X)u(\vartheta, X)^\top]$$

die Informationsmatrix. Es gilt also

$$I(\vartheta)_{ij} = \mathbb{E}_\vartheta[u_i(\vartheta, X)u_j(\vartheta, X)] \quad \textit{für alle } i, j = 1, \ldots, d.$$

Für einen $\mathbb{R}^d$-wertigen Zufallsvektor $X = (X_1, \ldots, X_d) \in \mathscr{L}^2$ bezeichnen wir mit $\mathrm{Cov}(X)$ die Kovarianzmatrix gegeben durch

$$\mathrm{Cov}(X)_{ij} = \mathrm{Cov}(X_i, X_j), \quad i, j = 1, \ldots, d.$$

Die Kovarianzmatrix ist bekanntlich symmetrisch und positiv semidefinit; letzteres bedeutet

$$\langle \mathrm{Cov}(X)b, b\rangle_{\mathbb{R}^d} \geq 0 \quad \text{für alle } b \in \mathbb{R}^d.$$

Lemma 1.11.2. *Es sei $\vartheta \in \Theta$ beliebig.*

(a) Es gilt $\mathbb{E}_\vartheta[u(\vartheta, X)] = 0$.

(b) Es gilt $I(\vartheta) = \mathrm{Cov}_\vartheta(u(\vartheta, X))$, und $I(\vartheta)$ ist symmetrisch und positiv semidefinit.

(c) Für alle $\eta \in \mathbb{R}^d$ gilt

$$\mathrm{Var}_\vartheta[\langle \eta, u(\vartheta, X)\rangle_{\mathbb{R}^d}] = \langle I(\vartheta)\eta, \eta\rangle_{\mathbb{R}^d}.$$

Beweis.

(a) Es gilt

$$\begin{aligned} \mathbb{E}_\vartheta[u(\vartheta, X)] &= \int_\Omega u(\vartheta, x) L(\vartheta, x)\mu(dx) \\ &= \int_\Omega \nabla_\vartheta L(\vartheta, x)\mu(dx) = \nabla_\vartheta \underbrace{\int_\Omega L(\vartheta, x)\mu(dx)}_{=1} = 0. \end{aligned}$$

(b) Folgt aus Teil (a).

(c) ✓

□

Lemma 1.11.3. *Es sei $X \in \mathscr{L}^2$ eine Zufallsvariable. Dann gilt*

$$\sqrt{\mathrm{Var}[X]} = \max_{Z \in \mathscr{Z}} \frac{\mathbb{E}[XZ]}{\sqrt{\mathrm{Var}[Z]}},$$

wobei

$$\mathscr{Z} = \{Z \in \mathscr{L}^2 : \mathbb{E}[Z] = 0 \textit{ und } \mathrm{Var}[Z] > 0\}.$$

Beweis. Wir dürfen annehmen, dass $\mathrm{Var}[X] > 0$. Es sei $Z \in \mathscr{Z}$ beliebig. Nach der Ungleichung von Cauchy-Schwarz gilt

$$\mathbb{E}[XZ] \leq \sqrt{\mathrm{Var}[X]}\sqrt{\mathrm{Var}[Z]},$$

und daher

$$\sqrt{\mathrm{Var}[X]} \geq \frac{\mathbb{E}[XZ]}{\sqrt{\mathrm{Var}[Z]}},$$

Nun setzen wir

$$Z := \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}[X]}} \in \mathscr{Z}.$$

Dann gilt

$$\frac{\mathbb{E}[XZ]}{\sqrt{\mathrm{Var}[Z]}} = \frac{\mathrm{Cov}(X,Z)}{\sqrt{\mathrm{Var}[Z]}} = \frac{\sqrt{\mathrm{Var}[X]}}{\sqrt{\mathrm{Var}[X]}} \cdot \frac{\mathrm{Cov}(X,X)}{\sqrt{\mathrm{Var}[X]}} = \sqrt{\mathrm{Var}[X]}.$$

□

Eine Matrix $A \in \mathbb{R}^{d \times d}$ heißt bekanntlich positiv definit, falls

$$\langle Ab, b \rangle_{\mathbb{R}^d} > 0 \quad \text{für alle } b \in \mathbb{R}^d \setminus \{0\}.$$

Für eine symmetrische, positiv semidefinite Matrix $A \in \mathbb{R}^{d \times d}$ gilt $\det A \geq 0$. Es gilt $\det A > 0$ genau dann, wenn A positiv definit ist. In dem Fall ist A^{-1} ebenfalls symmetrisch und positiv definit.

Satz 1.11.4 (Mehrdimensionale Informationsungleichung von Cramer-Rao)**.** *Es seien*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein mehrdimensionales reguläres statistisches Modell mit $\Theta \subset \mathbb{R}^d$, so dass $I(\vartheta)$ für jedes $\vartheta \in \Theta$ positiv definit ist, $\tau \in C^1(\Theta; \mathbb{R})$ eine Kenngröße und $T : \Omega \to \mathbb{R}$ ein erwartungstreuer Schätzer für τ. Dann gilt

$$\mathrm{Var}_\vartheta[T] \geq \langle I(\vartheta)^{-1} \nabla \tau(\vartheta), \nabla \tau(\vartheta) \rangle_{\mathbb{R}^d} \quad \textit{für alle } \vartheta \in \Theta.$$

Beweis. Es sei $\vartheta \in \Theta$ beliebig. Wir dürfen annehmen, dass $\nabla \tau(\vartheta) \neq 0$. Für alle $j = 1, \ldots, d$ gilt

$$\begin{aligned}
\frac{\partial}{\partial \vartheta_j} \tau(\vartheta) &= \frac{\partial}{\partial \vartheta_j} \mathbb{E}_\vartheta[T] = \frac{\partial}{\partial \vartheta_j} \int_\Omega T(x) L(\vartheta, x) \mu(dx) \\
&= \int_\Omega T(x) \frac{\partial}{\partial \vartheta_j} L(\vartheta, x) \mu(dx) = \int_\Omega T(x) u_j(\vartheta, x) L(\vartheta, x) \mu(dx) \\
&= \mathbb{E}_\vartheta[T\, u_j(\vartheta, \cdot)].
\end{aligned}$$

Also gilt für alle $\eta \in \mathbb{R}^d$

$$\mathbb{E}_\vartheta[T \langle \eta, u(\vartheta, X) \rangle_{\mathbb{R}^d}] = \langle \eta, \nabla \tau(\vartheta) \rangle_{\mathbb{R}^d}.$$

Es sei

$$\mathscr{Z}_\vartheta = \{Z \in \mathscr{L}^2(\mathbb{P}_\vartheta) : \mathbb{E}_\vartheta[Z] = 0 \text{ und } \mathrm{Var}_\vartheta[Z] > 0\}.$$

Für alle $\eta \in \mathbb{R}^d \setminus \{0\}$ gilt nach Lemma 1.11.2(c), dass $Z := \langle \eta, u(\vartheta, X)\rangle_{\mathbb{R}^d} \in \mathscr{Z}_\vartheta$ mit

$$\mathrm{Var}_\vartheta[Z] = \langle I(\vartheta)\eta, \eta\rangle_{\mathbb{R}^d} > 0.$$

Also folgt mit Lemma 1.11.3

$$\begin{aligned}\sqrt{\mathrm{Var}_\vartheta[T]} &= \max_{Z \in \mathscr{Z}_\vartheta} \frac{\mathbb{E}_\vartheta[TZ]}{\sqrt{\mathrm{Var}_\vartheta[Z]}} \\ &\geq \max_{\eta \in \mathbb{R}^d \setminus \{0\}} \frac{\mathbb{E}_\vartheta[T\langle \eta, u(\vartheta, X)\rangle_{\mathbb{R}^d}]}{\sqrt{\mathrm{Var}_\vartheta[\langle \eta, u(\vartheta, X)\rangle_{\mathbb{R}^d}]}} \\ &= \max_{\eta \in \mathbb{R}^d \setminus \{0\}} \frac{\langle \nabla\tau(\vartheta), \eta\rangle_{\mathbb{R}^d}}{\sqrt{\langle I(\vartheta)\eta, \eta\rangle_{\mathbb{R}^d}}}.\end{aligned}$$

Für $\eta := I(\vartheta)^{-1}\nabla\tau(\vartheta) \in \mathbb{R}^d$ gilt $\eta \neq 0$, da $\nabla\tau(\vartheta) \neq 0$ und $I(\vartheta)^{-1} : \mathbb{R}^d \to \mathbb{R}^d$ ein Isomorphismus ist. Also folgt

$$\sqrt{\mathrm{Var}_\vartheta[T]} \geq \frac{\langle \nabla\tau(\vartheta), I(\vartheta)^{-1}\nabla\tau(\vartheta)\rangle_{\mathbb{R}^d}}{\sqrt{\langle \nabla\tau(\vartheta), I(\vartheta)^{-1}\nabla\tau(\vartheta)\rangle_{\mathbb{R}^d}}} = \sqrt{\langle I(\vartheta)^{-1}\nabla\tau(\vartheta), \nabla\tau(\vartheta)\rangle_{\mathbb{R}^d}}.$$

$\square$

Definition 1.11.5. *Es seien*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein mehrdimensionales reguläres statistisches Modell mit $\Theta \subset \mathbb{R}^d$, so dass $I(\vartheta)$ für jedes $\vartheta \in \Theta$ positiv definit ist, $\tau \in C^1(\Theta; \mathbb{R})$ eine Kenngröße und $T : \Omega \to \mathbb{R}$ ein erwartungstreuer Schätzer für τ. Dann heißt T Cramer-Rao-optimal, falls

$$\mathrm{Var}_\vartheta[T] = \langle I(\vartheta)^{-1}\nabla\tau(\vartheta), \nabla\tau(\vartheta)\rangle_{\mathbb{R}^d} \quad \text{für alle } \vartheta \in \Theta.$$

Bemerkung 1.11.6. *Ein Cramer-Rao-optimaler Schätzer ist ein gleichmäßig bester Schätzer, aber ein gleichmäßig bester Schätzer braucht nicht Cramer-Rao-optimal zu sein.*

Beispiel 1.11.7. *Wir betrachten das n-fache Gaußmodell. Dann gilt $\Theta = \mathbb{R} \times (0, \infty)$. Es sei $\vartheta = (\mu, \sigma^2) \in \Theta$ beliebig. Bei einer Beobachtung ist die Likelihood-Funktion*

$$\begin{aligned}L(\vartheta, x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\bigg(- \frac{(x-\mu)^2}{2\sigma^2}\bigg) \\ &= \exp\bigg(- \frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}\bigg),\end{aligned}$$

und daher erhalten wir die Log-Likelihood-Funktion

$$\ell(\vartheta, x) = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}.$$

Also gilt beim n-fachen Gaußmodell

$$\ell^{\otimes n}(\vartheta, x) = -\frac{n}{2}\ln(2\pi\sigma^2) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}.$$

Es folgt

$$\frac{\partial}{\partial\mu}\ell^{\otimes n}(\vartheta, x) = \sum_{i=1}^{n}\frac{x_i-\mu}{\sigma^2} = \frac{1}{\sigma}\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma}$$

und

$$\frac{\partial}{\partial\sigma^2}\ell^{\otimes n}(\vartheta, x) = -\frac{n\cdot 2\pi}{2\cdot 2\pi\sigma^2} + \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^4} = \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 - \frac{n}{2\sigma^2}.$$

Also gilt

$$u(\vartheta, x) = \left(\frac{1}{\sigma}\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma}, \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 - \frac{n}{2\sigma^2}\right).$$

Wir setzen

$$Y_i := \frac{X_i-\mu}{\sigma}, \quad i = 1, \ldots, n.$$

Dann sind $Y_1, \ldots, Y_n$ unabhängig mit

$$Y_i \sim \mathrm{N}(0,1), \quad i = 1, \ldots, n.$$

Daher gilt

$$\mathrm{Var}_\vartheta[u_1(\vartheta, X)] = \mathrm{Var}_\vartheta\left[\frac{1}{\sigma}\sum_{i=1}^{n}Y_i\right] = \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathrm{Var}_\vartheta[Y_i] = \frac{n}{\sigma^2}.$$

Es gilt

$$\sum_{i=1}^{n}Y_i^2 \sim \chi_n^2,$$

und daher

$$\mathrm{Var}\left[\sum_{i=1}^n Y_i^2\right] = 2n.$$

Also folgt

$$\begin{aligned}\mathrm{Var}_\vartheta[u_2(\vartheta, X)] &= \mathrm{Var}_\vartheta\left[\frac{1}{2\sigma^2}\sum_{i=1}^n Y_i^2 - \frac{n}{2\sigma^2}\right]\\ &= \frac{1}{4\sigma^4}\mathrm{Var}_\vartheta\left[\sum_{i=1}^n Y_i^2\right] = \frac{2n}{4\sigma^4} = \frac{n}{2\sigma^4}.\end{aligned}$$

Für $Z \sim \mathrm{N}(0,1)$ *gilt*

$$\mathrm{Cov}(Z, Z^2) = \mathbb{E}[Z^3] = 0.$$

Also folgt

$$\begin{aligned}\mathrm{Cov}_\vartheta(u_1(\vartheta, X), u_2(\vartheta, X)) &= \mathrm{Cov}\left(\frac{1}{\sigma}\sum_{i=1}^n Y_i, \frac{1}{2\sigma^2}\sum_{i=1}^n Y_i^2\right)\\ &= \frac{1}{2\sigma^3}\sum_{i,j=1}^n \mathrm{Cov}(Y_i, Y_j^2) = 0.\end{aligned}$$

Wir erhalten die Informationsmatrix

$$I(\vartheta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix}, \quad \vartheta \in \Theta.$$

Also ist $I(\vartheta)$ *positiv definit mit*

$$I(\vartheta)^{-1} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}, \quad \vartheta \in \Theta.$$

Nun betrachten wir die Kenngröße $\tau : \Theta \to \mathbb{R}$ *gegeben durch*

$$\tau(\vartheta) = \sigma^2, \quad \vartheta = (\mu, \sigma^2) \in \Theta.$$

Dann gilt

$$\nabla\tau(\vartheta) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Also folgt

$$\langle I(\vartheta)^{-1}\nabla\tau(\vartheta), \nabla\tau(\vartheta)\rangle_{\mathbb{R}^2} = \left\langle \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle_{\mathbb{R}^2}$$
$$= \left\langle \begin{pmatrix} 0 \\ \frac{2\sigma^4}{n} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle_{\mathbb{R}^2} = \frac{2\sigma^4}{n}.$$

Nach Beispiel 1.10.11 ist $s^2(X)$ der eindeutig bestimmte gleichmäßig beste Schätzer für σ^2. Nach Beispiel 1.5.7(c) gilt jedoch

$$\mathrm{Var}_\vartheta[s^2(X)] = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n} = \langle I(\vartheta)^{-1}\nabla\tau(\vartheta), \nabla\tau(\vartheta)\rangle_{\mathbb{R}^2} \quad \text{für alle } \vartheta \in \Theta.$$

Also ist $s^2(X)$ nicht Cramer-Rao-optimal.

1.12 Bayes'sche Schätzer

Wir erinnern nochmal an die Definition eines stochastischen Kerns (siehe Definition 1.8.8).

Definition 1.12.1. *Es seien $(\Omega_1, \mathscr{F}_1)$ und $(\Omega_2, \mathscr{F}_2)$ messbare Räume. Eine Abbildung $\kappa : \Omega_1 \times \mathscr{F}_2 \to [0,1]$ heißt ein stochastischer Kern von $(\Omega_1, \mathscr{F}_1)$ nach $(\Omega_2, \mathscr{F}_2)$, falls gilt:*

(a) $\omega_1 \mapsto \kappa(\omega_1, A_2)$ ist $\mathscr{F}_1$-messbar für jedes $A_2 \in \mathscr{F}_2$.

(b) $A_2 \mapsto \kappa(\omega_1, A_2)$ ist ein Wahrscheinlichkeitsmaß auf $(\Omega_2, \mathscr{F}_2)$ für jedes $\omega_1 \in \Omega_1$.

Satz 1.12.2. *Es seien $(\Omega_1, \mathscr{F}_1, \nu)$ ein Wahrscheinlichkeitsraum, $(\Omega_2, \mathscr{F}_2)$ ein messbarer Raum und κ ein stochastischer Kern von $(\Omega_1, \mathscr{F}_1)$ nach $(\Omega_2, \mathscr{F}_2)$. Dann existiert ein eindeutig bestimmtes Wahrscheinlichkeitsmaß $\nu \otimes \kappa$ auf $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$, so dass*

$$(\nu \otimes \kappa)(A_1 \times A_2) = \int_{A_1} \kappa(\omega_1, A_2)\nu(d\omega_1) \quad \text{für alle } A_1 \in \mathscr{F}_1 \text{ und } A_2 \in \mathscr{F}_2.$$

Satz 1.12.3 (Satz von Fubini für stochastische Kerne)**.** *Es seien $(\Omega_1, \mathscr{F}_1)$ und $(\Omega_2, \mathscr{F}_2)$ messbare Räume, es sei ν ein Wahrscheinlichkeitsmaß auf $(\Omega_1, \mathscr{F}_1)$, und es sei κ ein stochastischer Kern von $(\Omega_1, \mathscr{F}_1)$ nach $(\Omega_2, \mathscr{F}_2)$. Weiterhin sei*

$$f \in \mathscr{L}^1(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \nu \otimes \kappa).$$

Dann gilt

$$\int_{\Omega_1 \times \Omega_2} f \, d(\nu \otimes \kappa) = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\kappa(\omega_1, d\omega_2) \right) \nu(d\omega_1).$$

Definition 1.12.4. *Ein Paar $(\mathscr{M}, \mathscr{T})$ heißt ein Bayes'sches Modell, falls gilt:*

(a) $\mathscr{M} = (\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta))$ *ist ein Standardmodell mit Likelihood-Funktion* $L : \Theta \times \Omega \to \mathbb{R}_+$.

(b) $\mathscr{T}$ *ist eine σ-Algebra über Θ, und die Likelihood-Funktion L ist $\mathscr{T} \otimes \mathscr{F}$-messbar.*

Bemerkung 1.12.5. *Die Abbildung $\kappa : \Theta \times \mathscr{F} \to [0,1]$ gegeben durch*

$$\kappa(\vartheta, A) = \mathbb{P}_\vartheta(A)$$

ist ein stochastischer Kern von $(\Theta, \mathscr{T})$ nach $(\Omega, \mathscr{F})$. In der Tat, für jedes $A \in \mathscr{F}$ ist

$$\Theta \to \mathbb{R}, \quad \vartheta \mapsto \mathbb{P}_\vartheta(A) = \int_A L(\vartheta, x)\mu(dx)$$

nach dem Satz von Fubini $\mathscr{T}$-messbar.

Definition 1.12.6. *Es sei $(\mathscr{M}, \mathscr{T})$ ein Bayes'sches Modell. Ein Wahrscheinlichkeitsmaß ν auf $(\Theta, \mathscr{T})$ mit $L(\cdot, x) \in \mathscr{L}^1(\Theta, \mathscr{T}, \nu)$ für jedes $x \in \Omega$ heißt eine a-priori-Verteilung für den Parameter $\vartheta \in \Theta$.*

Definition 1.12.7. *Es seien $(\mathscr{M}, \mathscr{T})$ ein Bayes'sches Modell, und ν eine a-priori-Verteilung für den Parameter $\vartheta \in \Theta$. Wir bezeichnen mit $\nu \otimes \mathbb{P}$ das gemäß Satz 1.12.2 eindeutig bestimmte Wahrscheinlichkeitsmaß auf $(\Theta \times \Omega, \mathscr{T} \otimes \mathscr{F})$, und nennen es die Mischung der Familie $(\mathbb{P}_\vartheta)_{\vartheta \in \Theta}$ mit der Verteilung ν.*

Korollar 1.12.8. *Es seien $(\mathscr{M}, \mathscr{T})$ ein Bayes'sches Modell, und ν eine a-priori-Verteilung für den Parameter $\vartheta \in \Theta$. Für jede Funktion $f \in \mathscr{L}^1(\Theta \times \Omega, \mathscr{T} \otimes \mathscr{F}, \nu \otimes \mathbb{P})$ gilt*

$$\int_{\Theta \times \Omega} f d(\nu \otimes \mathbb{P}) = \int_\Theta \left(\int_\Omega f(\vartheta, x)\mathbb{P}_\vartheta(dx) \right) \nu(d\vartheta) = \int_\Theta \mathbb{E}_\vartheta[f(\vartheta, X)]\nu(d\vartheta).$$

Beweis. Foglt aus dem Satz von Fubini für stochastische Kerne (Satz 1.12.3). □

Definition 1.12.9. *Es seien $(\mathscr{M}, \mathscr{T})$ ein Bayes'sches Modell, und ν eine a-priori-Verteilung für den Parameter $\vartheta \in \Theta$.*

(a) Wir definieren $L_\nu : \Omega \to \mathbb{R}$ durch

$$L_\nu(x) := \int_\Theta L(\vartheta, x)\nu(d\vartheta).$$

(b) Wir definieren $\pi : \Theta \times \Omega \to \mathbb{R}_+$ durch

$$\pi(\vartheta, x) := \frac{L(\vartheta, x)}{L_\nu(x)}.$$

(c) *Für $x \in \Omega$ heißt $\pi(\cdot, x)$ die a-posteriori-Dichte für den Parameter $\vartheta \in \Theta$ gegeben die Beobachtung x.*

(d) *Für $x \in \Omega$ heißt das Wahrscheinlichkeitsmaß ν_x auf $(\Theta, \mathscr{T})$ mit*

$$\frac{d\nu_x}{d\nu} = \pi(\cdot, x)$$

die a-posteriori-Verteilung für den Parameter $\vartheta \in \Theta$ gegeben die Beobachtung x.

Bemerkung 1.12.10. *Es gilt*

$$\int_\Theta \pi(\vartheta, x)\nu(d\vartheta) = 1 \quad \text{für alle } x \in \Omega.$$

Definition 1.12.11. *Die Betaverteilung mit Parametern $r, s \in (0, \infty)$ ist die absolutstetige Verteilung auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Dichte $f : \mathbb{R} \to \mathbb{R}_+$ gegeben durch*

$$f(x) = \frac{x^{r-1}(1-x)^{s-1}}{B(r, s)} \mathbb{1}_{(0,1)}(x),$$

Wir bezeichnen sie mit $\mathrm{Beta}(r, s)$. Hierbei bezeichnet

$$B : (0, \infty)^2 \to \mathbb{R}, \quad B(r, s) = \int_0^1 y^{r-1}(1-y)^{s-1} dy$$

die Betafunktion. Es gilt auch

$$B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}.$$

Beispiel 1.12.12. *Wir betrachten das Binomialmodell aus Beispiel 1.1.10. Dann ist $(\mathscr{M}, \mathscr{T})$ wie folgt gegeben:*

- $\mathscr{M} = (\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta))$ *ist gegeben durch*

  $$\Omega = \{0, 1, \ldots, n\}, \quad \mathscr{F} = \mathfrak{P}(\Omega)$$

  und $\mathbb{P}_\vartheta = \mathrm{Bi}(n, \vartheta)$ für $\vartheta \in \Theta = (0, 1)$.

- $\mathscr{T} = \mathcal{B}(\Theta)$.

Dann ist $(\mathscr{M}, \mathscr{T})$ ein Bayes'sches Modell. Wir wählen die a-priori-Verteilung

$$\nu = \mathrm{UC}(\Theta).$$

Es sei $k \in \Omega$ beliebig. Wir erhalten als a-posteriori-Verteilung ν_k die absolutstetige Verteilung mit a-posteriori-Dichte

$$\pi(\vartheta, k) = \frac{\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}}{\int_0^1 \binom{n}{k} t^k(1-t)^{n-k}dt} = \frac{\vartheta^k(1-\vartheta)^{n-k}}{B(k+1, n-k+1)}, \quad \vartheta \in \Theta.$$

Also gilt

$$\nu_k = \mathrm{Beta}(k+1, n-k+1) \quad \textit{für jedes } k \in \Omega.$$

Zur Erinnerung: Der mittlere quadratische Fehler eines Schätzers T ist gegeben durch

$$\mathbb{F}_\vartheta(T) = \mathbb{E}_\vartheta[(T - \tau(\vartheta))^2], \quad \vartheta \in \Theta.$$

Definition 1.12.13. *Es seien $(\mathscr{M}, \mathscr{T})$ ein Bayes'sches Modell, und ν eine a-priori-Verteilung für den Parameter $\vartheta \in \Theta$. Weiterhin seien $\tau : \Theta \to \mathbb{R}$ eine messbare Kenngröße und $T : \Omega \to \mathbb{R}$ ein Schätzer.*

(a) Wir definieren den (über x und ϑ gemittelten) quadratischen Fehler

$$\mathbb{F}_\nu(T) := \int_\Theta \mathbb{F}_\vartheta(T)\nu(d\vartheta).$$

(b) Der Schätzer T heißt ein Bayes-Schätzer zur Kenngröße τ, falls

$$\mathbb{F}_\nu(T) \le \mathbb{F}_\nu(S)$$

für jeden Schätzer $S : \Omega \to \mathbb{R}$.

Lemma 1.12.14.

(a) Es gilt

$$\mathbb{F}_\nu(T) = \int_{\Theta\times\Omega} (T(x) - \tau(\vartheta))^2(\nu \otimes \mathbb{P})(d\vartheta, dx).$$

(b) Es gilt

$$\mathbb{F}_\nu(T) = \int_\Omega \left(\int_\Theta (T(x) - \tau(\vartheta))^2 L(\vartheta, x)\nu(d\vartheta) \right) \mu(dx).$$

Beweis.

(a) Nach dem Satz von Fubini für stochastische Kerne (Satz 1.12.3) gilt

$$\begin{aligned}
\mathbb{F}_\nu(T) &= \int_\Theta \mathbb{F}_\vartheta(T)\nu(d\vartheta) = \int_\Theta \mathbb{E}_\vartheta[(T-\tau(\vartheta))^2]\nu(d\vartheta) \\
&= \int_\Theta \bigg(\int_\Omega (T(x)-\tau(\vartheta))^2 \mathbb{P}_\vartheta(dx) \bigg) \nu(d\vartheta) \\
&= \int_{\Theta\times\Omega} (T(x)-\tau(\vartheta))^2 (\nu\otimes\mathbb{P})(d\vartheta, dx).
\end{aligned}$$

(b) Mit dem gewöhnlichen Satz von Fubini erhalten wir

$$\begin{aligned}
\mathbb{F}_\nu(T) &= \int_\Theta \bigg(\int_\Omega (T(x)-\tau(\vartheta))^2 L(\vartheta,x)\mu(dx) \bigg) \nu(d\vartheta) \\
&= \int_\Omega \bigg(\int_\Theta (T(x)-\tau(\vartheta))^2 L(\vartheta,x)\nu(d\vartheta) \bigg) \mu(dx).
\end{aligned}$$

□

Satz 1.12.15. *Es seien* $(\mathscr{M}, \mathscr{T})$ *ein Bayes'sches Modell, und* ν *eine a-priori-Verteilung für den Parameter* $\vartheta \in \Theta$. *Weiterhin seien* $\tau : \Theta \to \mathbb{R}$ *eine messbare Kenngröße mit* $\tau \in \mathscr{L}^1(\nu_x)$ *für alle* $x \in \Omega$. *Dann ist*

$$T : \Omega \to \mathbb{R}, \quad T(x) := \int_\Theta \tau(\vartheta)\nu_x(d\vartheta) = \int_\Theta \tau(\vartheta)\pi(\vartheta,x)\nu(d\vartheta)$$

ein Bayes-Schätzer zur Kenngröße τ.

Beweis. Es sei $S : \Omega \to \mathbb{R}$ ein Schätzer. Mit Lemma 1.12.14(b) und der Gleichung $\int_\Theta \pi(\vartheta,x)\nu(d\vartheta) = 1$ (siehe Bemerkung 1.12.10) für alle $x \in \Omega$ folgt

$$\begin{aligned}
\mathbb{F}_\nu(S) - \mathbb{F}_\nu(T) &= \int_\Omega \bigg(\int_\Theta \big[(S(x)-\tau(\vartheta))^2 - (T(x)-\tau(\vartheta))^2\big] L(\vartheta,x)\nu(d\vartheta) \bigg) \mu(dx) \\
&= \int_\Omega \bigg(\int_\Theta \big[S(x)^2 - 2S(x)\tau(\vartheta) - T(x)^2 + 2T(x)\tau(\vartheta)\big] L(\vartheta,x)\nu(d\vartheta) \bigg) \mu(dx) \\
&= \int_\Omega L_\nu(x) \bigg(\int_\Theta \big[S(x)^2 - 2S(x)\tau(\vartheta) - T(x)^2 + 2T(x)\tau(\vartheta)\big] \pi(\vartheta,x)\nu(d\vartheta) \bigg) \mu(dx) \\
&= \int_\Omega L_\nu(x) \big[S(x)^2 - 2S(x)T(x) - T(x)^2 + 2T(x)^2\big] \mu(dx) \\
&= \int_\Omega L_\nu(x) (S(x)-T(x))^2 \mu(dx) \geq 0.
\end{aligned}$$

□

Beispiel 1.12.16. *Beim Binomialmodell aus Beispiel 1.12.12 betrachten wir die Kenngröße* $\tau : \Theta \to \mathbb{R}$, $\tau(\vartheta) = \vartheta$. *Wir erhalten den Bayes-Schätzer*

$$\begin{aligned} T(k) &= \int_\Theta \vartheta \pi(\vartheta, k) \nu(d\vartheta) = \frac{1}{B(k+1, n-k+1)} \int_0^1 \vartheta^{k+1}(1-\vartheta)^{n-k} d\vartheta \\ &= \frac{B(k+2, n-k+1)}{B(k+1, n-k+1)} = \frac{\Gamma(k+2)\Gamma(n-k+1)}{\Gamma(n+3)} \cdot \frac{\Gamma(n+2)}{\Gamma(k+1)\Gamma(n-k+1)} \\ &= \frac{\Gamma(k+2) \cdot \Gamma(n+2)}{\Gamma(n+3) \cdot \Gamma(k+1)} = \frac{(k+1)! \cdot (n+1)!}{(n+2)! \cdot k!} = \frac{k+1}{n+2}. \end{aligned}$$

Hierbei haben wir benutzt, dass $\Gamma(n+1) = n!$ *für alle* $n \in \mathbb{N}_0$. *Also gilt*

$$T = \frac{X+1}{n+2};$$

vergleiche Beispiel 1.5.4.

Kapitel 2

Bereichsschätzer

2.1 Definitionen und Eigenschaften

Definition 2.1.1. *Es seien*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell, und $\tau : \Theta \to E$ eine Kenngröße mit Werten in einem messbaren Raum $(E, \mathscr{E})$.

(a) Ein Abbildung $C : \Omega \to \mathscr{E}$ heißt ein <u>Bereichsschätzer</u>.

(b) Für $\alpha \in (0,1)$ heißt ein Bereichsschätzer C eine <u>Konfidenzmenge</u> (oder ein <u>Konfidenzbereich</u>) für τ zum Niveau α, falls $\{\tau(\vartheta) \in C\} \in \mathscr{F}$ und

$$\mathbb{P}_\vartheta(\tau(\vartheta) \in C) \geq 1 - \alpha \quad \textit{für alle } \vartheta \in \Theta.$$

(c) Gilt $(E, \mathscr{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, und ist C von der Form $C = [\underline{T}, \overline{T}]$ mit zwei Statistiken $\underline{T} \leq \overline{T}$, so nennen wir C auch ein <u>Konfidenzintervall</u>. In diesem Fall gilt automatisch

$$\{\tau(\vartheta) \in C\} = \{\underline{T} \leq \tau(\vartheta) \leq \overline{T}\} = \{\underline{T} \leq \tau(\vartheta)\} \cap \{\tau(\vartheta) \leq \overline{T}\} \in \mathscr{F}$$

für jedes $\vartheta \in \Theta$.

Lemma 2.1.2. *Für beliebige Ereignisse $A_1, \ldots, A_m \in \mathscr{F}$ gilt*

$$\mathbb{P}\Big(\bigcap_{j=1}^{m} A_j\Big) \geq 1 - \sum_{j=1}^{m} \mathbb{P}(A_j^c).$$

Beweis. Nach den De Morgan'schen Gesetzen und der σ-Subadditivität von $\mathbb{P}$ gilt

$$\mathbb{P}\Big(\bigcap_{j=1}^{m} A_j\Big) = \mathbb{P}\Big(\Big(\bigcup_{j=1}^{m} A_j^c\Big)^c\Big) = 1 - \mathbb{P}\Big(\bigcup_{j=1}^{m} A_j^c\Big) \geq 1 - \sum_{j=1}^{m} \mathbb{P}(A_j^c).$$

□

Satz 2.1.3. *Es seien $\tau_i : \Theta \to E_i$ Kenngrößen mit Werten in einem messbaren Raum $(E_i, \mathscr{E}_i)$ für $i = 1, \ldots, m$, und es seien $\alpha \in (0,1)$ und $\alpha_i \in (0,1)$, $i = 1, \ldots, m$, so dass $\alpha = \alpha_1 + \ldots + \alpha_m$. Weiterhin seien $C_i : \Omega \to \mathscr{E}_i$ Konfidenzmengen für τ_i zum Niveau α_i für $i = 1, \ldots, m$. Dann ist $C : \Omega \to \mathscr{E}_1 \otimes \ldots \otimes \mathscr{E}_m$ gegeben durch $C := (C_1, \ldots, C_m)$ eine Konfidenzmenge für $\tau : \Theta \to E_1 \times \ldots \times E_m$ gegeben durch $\tau := (\tau_1, \ldots, \tau_m)$ zum Niveau α.*

Beweis. Für jedes $\vartheta \in \Theta$ gilt

$$\{\tau(\vartheta) \in C\} = \bigcap_{j=1}^{m} \underbrace{\{\tau_j(\vartheta) \in C_j\}}_{\in \mathscr{F}} \in \mathscr{F},$$

und nach Lemma 2.1.2 gilt

$$\begin{aligned}\mathbb{P}_\vartheta(\tau(\vartheta) \in C) &= \mathbb{P}_\vartheta\Big(\bigcap_{j=1}^{m} \{\tau_j(\vartheta) \in C_j\}\Big) \\ &\geq 1 - \sum_{j=1}^{m} \mathbb{P}_\vartheta(\tau_j(\vartheta) \notin C_j) \geq 1 - \sum_{j=1}^{m} \alpha_j = 1 - \alpha.\end{aligned}$$

□

Definition 2.1.4. *Eine Abbildung $\pi : \Theta \times \Omega \to E$ heißt eine Pivot-Statistik (oder ein Pivot), falls gilt:*

(a) Für jedes $\vartheta \in \Theta$ ist $\pi(\vartheta, \cdot) : (\Omega, \mathscr{F}) \to (E, \mathscr{E})$ messbar.

(b) Es existiert ein Wahrscheinlichleitsmaß η auf $(E, \mathscr{E})$, so dass

$$\mathbb{P}_\vartheta \circ \pi(\vartheta, \cdot) = \eta \quad \text{für alle } \vartheta \in \Theta.$$

In diesem Fall nennen wir η die zugehörige Pivot-Verteilung.

Satz 2.1.5. *Es sei π eine Pivot-Statistik mit Pivot-Verteilung η. Es seien $\alpha \in (0,1)$ und $B \in \mathscr{E}$, so dass*

$$\eta(B) \geq 1 - \alpha.$$

Weiterhin sei $C : \Omega \to \mathscr{E}$ eine Abbildung, so dass

$$\{\tau(\vartheta) \in C\} = \{\pi(\vartheta, \cdot) \in B\} \quad \text{für jedes } \vartheta \in \Theta.$$

Dann ist C eine Konfidenzmenge für τ zum Niveau α.

Beweis. Für jedes $\vartheta \in \Theta$ gilt

$$\{\tau(\vartheta) \in C\} = \{\pi(\vartheta, \cdot) \in B\} \in \mathscr{F}$$

und

$$\mathbb{P}_\vartheta(\tau(\vartheta) \in C) = \mathbb{P}_\vartheta(\pi(\vartheta, \cdot) \in B) = \eta(B) \geq 1 - \alpha.$$

□

2.2 Bereichsschätzer in Gaußmodellen

Definition 2.2.1. *Es sei $n \in \mathbb{N}$ beliebig.*

(a) Wir nennen $\chi_n^2 = \Gamma(\frac{n}{2}, \frac{1}{2})$ eine Chi-Quadrat-Verteilung mit n Freiheitsgraden. Ihre Dichte $f : \mathbb{R} \to \mathbb{R}_+$ ist gegeben durch

$$f(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{(0,\infty)}(x), \quad x \in \mathbb{R}.$$

(b) Wir nennen die absolutstetige Verteilung t_n auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Dichte $f : \mathbb{R} \to \mathbb{R}_+$ gegeben durch

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}$$

eine t-Verteilung mit n Freiheitsgraden.

Satz 2.2.2. *Es sei $n \in \mathbb{N}$ beliebig.*

(a) Es seien $X_1, \ldots, X_n \sim \mathrm{N}(0,1)$ unabhängige, standardnormalverteilte Zufallsvariablen. Dann gilt

$$\sum_{i=1}^n X_i^2 \sim \chi_n^2.$$

(b) Es seien $X \sim \mathrm{N}(0,1)$ und $Y \sim \chi_n^2$ unabhängige Zufallsvariablen. Dann gilt

$$\frac{X}{\sqrt{Y/n}} \sim t_n.$$

Beweis. Übung. □

Satz 2.2.3. *Es seien $X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma^2)$ unabhängige, normalverteilte Zufallsvariablen. Dann sind $\bar{X}$ und $s^2(X)$ unabhängig mit*

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim \mathrm{N}(0,1), \qquad \frac{n-1}{\sigma^2} s^2(X) \sim \chi^2_{n-1} \quad \textit{und} \quad \frac{\bar{X} - \mu}{\sqrt{s^2(X)/n}} \sim t_{n-1}.$$

Beweis. Nach Satz 1.4.17 sind $\bar{X}$ und $s^2(X)$ unabhängig, und es gilt

$$\bar{X} \sim \mathrm{N}\Big(\mu, \frac{\sigma^2}{n}\Big) \quad \text{und} \quad Z := \frac{n-1}{\sigma^2} s^2(X) \sim \chi^2_{n-1}.$$

Also gilt

$$Y := \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim \mathrm{N}(0,1).$$

Die Zufallsvariablen Y und Z sind ebenfalls unabhängig, und mit Satz 2.2.2(b) folgt

$$\frac{\bar{X} - \mu}{\sqrt{s^2(X)/n}} = \frac{Y}{\sqrt{Z/(n-1)}} \sim t_{n-1}.$$

□

Definition 2.2.4. *Für $a \in (0,1)$ bezeichnen wir mit $z_a := \Phi^{-1}(a)$ das a-Quantil der Standardnormalverteilung. Hierbei bezeichnet $\Phi : \mathbb{R} \to [0,1]$ die Verteilungsfunktion von $\mathrm{N}(0,1)$.*

Im Folgenden benutzen wir für $a \in \mathbb{R}$ und $\epsilon > 0$ die Schreibweise

$$a \pm \epsilon := [a - \epsilon, a + \epsilon].$$

Satz 2.2.5. *Wir betrachten das n-fache Gaußmodell mit bekannter Varianz $\sigma^2 > 0$.*

(a) Die Zufallsvariable

$$\pi(\mu, \cdot) = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$$

ist eine Pivot-Statistik mit Pivot-Verteilung $\eta = \mathrm{N}(0,1)$.

(b) Für jedes $\alpha \in (0,1)$ ist das Zufallsintervall $C : \Omega \to \mathcal{B}(\mathbb{R})$ gegeben durch

$$C = \bar{X} \pm z_{1-\alpha/2} \sqrt{\sigma^2/n}$$

ein Konfidenzintervall für μ zum Niveau α.

Beweis.

(a) Folgt aus Satz 2.2.3.

(b) Mit $c := z_{1-\alpha/2}$ und $B = [-c, c]$ gilt wegen der Symmetrie der Dichte

$$\begin{aligned}\eta(B) &= \eta((-\infty, c]) - \eta((-\infty, -c)) = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) \\ &= 2\Phi(c) - 1 = 2\Big(1 - \frac{\alpha}{2}\Big) - 1 = 1 - \alpha,\end{aligned}$$

und für jedes $\mu \in \mathbb{R}$ gilt

$$\begin{aligned}&\{\mu \in C\} = \Big\{\mu \in \bar{X} \pm c\sqrt{\sigma^2/n}\Big\} = \Big\{\bar{X} - c\sqrt{\sigma^2/n} \le \mu \le \bar{X} + c\sqrt{\sigma^2/n}\Big\} \\ &= \Big\{\bar{X} - \mu - c\sqrt{\sigma^2/n} \le 0 \le \bar{X} - \mu + c\sqrt{\sigma^2/n}\Big\} \\ &= \{\pi(\mu, \cdot) - c \le 0 \le \pi(\mu, \cdot) + c\} = \{|\pi(\mu, \cdot)| \le c\} = \{\pi(\mu, \cdot) \in B\}.\end{aligned}$$

Also folgt dieser Teil mit Satz 2.1.5.

□

Satz 2.2.6. *Wir betrachten das n-fache Gaußmodell mit bekanntem Erwartungswert $\mu \in \mathbb{R}$, und setzen*

$$\tilde{\sigma}^2(X) := \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

(a) Die Zufallsvariable

$$\pi(\sigma^2, \cdot) = \frac{n\tilde{\sigma}^2(X)}{\sigma^2}$$

ist eine Pivot-Statistik mit Pivot-Verteilung $\eta = \chi^2_n$.

(b) Für jedes $\alpha \in (0, 1)$ ist das Zufallsintervall $C : \Omega \to \mathcal{B}(\mathbb{R})$ gegeben durch

$$C = \left[\frac{n\tilde{\sigma}^2(X)}{\chi^2_{n,1-\alpha/2}}, \frac{n\tilde{\sigma}^2(X)}{\chi^2_{n,\alpha/2}}\right]$$

ein Konfidenzintervall für σ^2 zum Niveau α.

Beweis.

(a) Mit Satz 2.2.2(a) folgt

$$\pi(\sigma^2, \cdot) = \frac{n\tilde{\sigma}^2(X)}{\sigma^2} = \sum_{i=1}^{n} \Bigg(\underbrace{\frac{X_i - \mu}{\sigma}}_{\sim \mathrm{N}(0,1)} \Bigg)^2 \sim \chi^2_n.$$

(b) Es sei F die Verteilungsfunktion von η. Mit $B = [\chi^2_{n,\alpha/2}, \chi^2_{n,1-\alpha/2}]$ gilt

$$\begin{aligned}\eta(B) &= \eta\big((-\infty, \chi^2_{n-d,1-\alpha/2}]\big) - \eta\big((-\infty, \chi^2_{n-d,\alpha/2})\big)\\ &= F(\chi^2_{n-d,1-\alpha/2}) - F(\chi^2_{n-d,\alpha/2}) = \Big(1 - \frac{\alpha}{2}\Big) - \frac{\alpha}{2} = 1 - \alpha,\end{aligned}$$

und für jedes $\sigma^2 > 0$ gilt

$$\begin{aligned}\{\sigma^2 \in C\} &= \left\{ \sigma^2 \in \left[\frac{n\tilde{\sigma}^2(X)}{\chi^2_{n,1-\alpha/2}}, \frac{n\tilde{\sigma}^2(X)}{\chi^2_{n,\alpha/2}} \right] \right\} = \left\{ \frac{n\tilde{\sigma}^2(X)}{\chi^2_{n,1-\alpha/2}} \le \sigma^2 \le \frac{n\tilde{\sigma}^2(X)}{\chi^2_{n,\alpha/2}} \right\}\\ &= \left\{ \frac{1}{\chi^2_{n,1-\alpha/2}} \le \frac{1}{\pi(\sigma^2, \cdot)} \le \frac{1}{\chi^2_{n,\alpha/2}} \right\} = \{\chi^2_{n,\alpha/2} \le \pi(\sigma^2, \cdot) \le \chi^2_{n,1-\alpha/2}\}\\ &= \{\pi(\sigma^2, \cdot) \in B\}.\end{aligned}$$

Also folgt dieser Teil mit Satz 2.1.5.

□

Satz 2.2.7. *Wir betrachten das n-fache Gaußmodell mit $\Theta = \mathbb{R} \times (0, \infty)$.*

(a) Die Zufallsvariable

$$\pi_1(\vartheta, \cdot) = \frac{\bar{X} - \mu}{\sqrt{s^2(X)/n}}, \quad \vartheta = (\mu, \sigma^2)$$

ist eine Pivot-Statistik mit Pivot-Verteilung $\eta_1 = t_{n-1}$.

(b) Die Zufallsvariable

$$\pi_2(\vartheta, \cdot) = \frac{n-1}{\sigma^2} s^2(X), \quad \vartheta = (\mu, \sigma^2)$$

ist eine Pivot-Statistik mit Pivot-Verteilung $\eta_2 = \chi^2_{n-1}$.

(c) Für jedes $\alpha \in (0, 1)$ ist das Zufallsintervall $C_1 : \Omega \to \mathcal{B}(\mathbb{R})$ gegeben durch

$$C_1 = \bar{X} \pm t_{n-1,1-\alpha/2} \sqrt{s^2(X)/n}$$

ein Konfidenzintervall für μ zum Niveau α.

(d) Für jedes $\alpha \in (0,1)$ ist das Zufallsintervall $C_2 : \Omega \to \mathcal{B}(\mathbb{R})$ gegeben durch

$$C_2 = \left[\frac{n-1}{\chi^2_{n-1,1-\alpha/2}} s^2(X), \frac{n-1}{\chi^2_{n-1,\alpha/2}} s^2(X)\right]$$

ein Konfidenzintervall für σ^2 zum Niveau α.

(e) Für jedes $\alpha \in (0,1)$ ist das Zufallsrechteck $C : \Omega \to \mathcal{B}(\mathbb{R}^2)$ gegeben durch

$$\left(\bar{X} \pm t_{n-1,1-\alpha/4}\sqrt{s^2(X)/n}\right) \times \left[\frac{n-1}{\chi^2_{n-1,1-\alpha/4}} s^2(X), \frac{n-1}{\chi^2_{n-1,\alpha/4}} s^2(X)\right]$$

ist ein Konfidenzbereich für $\vartheta = (\mu, \sigma^2)$ zum Niveau α.

Beweis.

(a) Folgt aus Satz 2.2.3.

(b) Folgt aus Satz 2.2.3.

(c) Mit $c := t_{n-1,1-\alpha/2}$ und $B = [-c, c]$ gilt wegen der Symmetrie der Dichte

$$\begin{aligned}\eta_1(B) &= \eta((-\infty, c]) - \eta((-\infty, -c)) = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) \\ &= 2\Phi(c) - 1 = 2\Big(1 - \frac{\alpha}{2}\Big) - 1 = 1 - \alpha,\end{aligned}$$

und für alle $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ gilt

$$\begin{aligned}\{\mu \in C_1\} &= \Big\{\mu \in \bar{X} \pm c\sqrt{s^2(X)/n}\Big\} \\ &= \Big\{\bar{X} - c\sqrt{s^2(X)/n} \le \mu \le \bar{X} + c\sqrt{s^2(X)/n}\Big\} \\ &= \Big\{\bar{X} - \mu - c\sqrt{s^2(X)/n} \le 0 \le \bar{X} - \mu + c\sqrt{s^2(X)/n}\Big\} \\ &= \{\pi_1(\vartheta, \cdot) - c \le 0 \le \pi_1(\vartheta, \cdot) + c\} \\ &= \{|\pi_1(\vartheta, \cdot)| \le c\} = \{\pi_1(\vartheta, \cdot) \in B\}.\end{aligned}$$

Also folgt dieser Teil mit Satz 2.1.5.

(d) Es sei F die Verteilungsfunktion von η_2. Mit $B = [\chi^2_{n-1,\alpha/2}, \chi^2_{n-1,1-\alpha/2}]$ gilt

$$\begin{aligned}\eta_2(B) &= \eta_2\big((-\infty, \chi^2_{n-d,1-\alpha/2}]\big) - \eta_2\big((-\infty, \chi^2_{n-d,\alpha/2})\big) \\ &= F(\chi^2_{n-d,1-\alpha/2}) - F(\chi^2_{n-d,\alpha/2}) = \Big(1 - \frac{\alpha}{2}\Big) - \frac{\alpha}{2} = 1 - \alpha,\end{aligned}$$

und für alle $\vartheta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ gilt

$$\begin{aligned}
\{\sigma^2 \in C_2\} &= \left\{\sigma^2 \in \left[\frac{n-1}{\chi^2_{n-1,1-\alpha/2}} s^2(X), \frac{n-1}{\chi^2_{n-1,\alpha/2}} s^2(X)\right]\right\} \\
&= \left\{\frac{n-1}{\chi^2_{n-1,1-\alpha/2}} s^2(X) \leq \sigma^2 \leq \frac{n-1}{\chi^2_{n-1,\alpha/2}} s^2(X)\right\} \\
&= \left\{\frac{1}{\chi^2_{n-1,1-\alpha/2}} \leq \frac{1}{\pi_2(\vartheta, \cdot)} \leq \frac{1}{\chi^2_{n-1,\alpha/2}}\right\} \\
&= \{\chi^2_{n-1,\alpha/2} \leq \pi_2(\vartheta, \cdot) \leq \chi^2_{n-1,1-\alpha/2}\} \\
&= \{\pi_2(\vartheta, \cdot) \in B\}.
\end{aligned}$$

Also folgt dieser Teil mit Satz 2.1.5.

(e) Folgt aus den Teilen (c) und (d) sowie Satz 2.1.3.

$\square$

Kapitel 3

Hypothesentests

3.1 Definitionen und grundlegende Eigenschaften

Es sei

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell. Es seien $\Theta_0, \Theta_1 \subset \Theta$ zwei Teilmengen, so dass $\Theta_0 \cap \Theta_1 = \emptyset$ und $\Theta = \Theta_0 \cup \Theta_1$.

Definition 3.1.1.

(a) Wir nennen $H_0 = \{\vartheta \in \Theta_0\}$ die Null-Hypothese.

(b) Wir nennen $H_1 = \{\vartheta \in \Theta_1\}$ die Alternative.

(c) Ist $\Theta_0 = \{\vartheta_0\}$ für ein $\vartheta_0 \in \Theta$, so sprechen wir von einer einfachen Hypothese, ansonsten von einer zusammengesetzten Hypothese.

(d) Sind $\Theta \subset \mathbb{R}$ und $\Theta_1 = \{\vartheta \in \Theta : \vartheta \neq \vartheta_0\}$ für ein $\vartheta_0 \in \Theta$, so nennen wir die Alternative zweiseitig.

(e) Sind $\Theta \subset \mathbb{R}$ und $\Theta_1 = \{\vartheta \in \Theta : \vartheta > \vartheta_0\}$ für ein $\vartheta_0 \in \Theta$, so nennen wir die Alternative einseitig.

Definition 3.1.2. *Eine Statistik $\varphi : \Omega \to \{0, 1\}$ heißt ein Test (von H_0 gegen H_1).*

- *Falls $\varphi(x) = 0$ für eine Beobachtung $x \in \Omega$, so wird die Null-Hypothese H_0 akzeptiert.*

- *Falls $\varphi(x) = 1$ für eine Beobachtung $x \in \Omega$, so wird die Null-Hypothese H_0 verworfen; das heißt, die Alternative H_1 wird akzeptiert.*

Bemerkung 3.1.3. *Jeder Test ist also von der Form* $\varphi = \mathbb{1}_A$ *für eine Menge* $A \in \mathscr{F}$.

Definition 3.1.4.

(a) Eine Entscheidung für H_1, *obwohl* H_0 *richtig ist, heißt ein* *Fehler 1. Art.*

(b) Eine Entscheidung für H_0, *obwohl* H_1 *richtig ist, heißt ein* *Fehler 2. Art.*

Bemerkung 3.1.5. *Es können also folgende Arten von Fehlern entstehen:*

	H_0 *wahr*	H_1 *wahr*
H_0 *wird akzeptiert*	*kein Fehler*	*Fehler 2. Art*
H_0 *wird verworfen*	*Fehler 1. Art*	*kein Fehler*

Beispiel 3.1.6. *Ein Unternehmen hat ein Medikament entwickelt, das angeblich in weniger als* 5% *der Anwendungen schädliche Nebenwirkungen hat. Wir betrachten die Null-Hypothese* $H_0 = \{\vartheta \geq 0.05\}$ *und die Alternative* $H_1 = \{\vartheta < 0.05\}$.

	$H_0 = \{\vartheta \geq 0.05\}$	$H_1 = \{\vartheta < 0.05\}$
H_0 *wird akzeptiert*	*kein Fehler*	*Fehler 2. Art: Das Medikament ist unschädlich, man hält es jedoch für schädlich.*
H_0 *wird verworfen*	*Fehler 1. Art!! Das Medikament ist schädlich, man hält es jedoch für unschädlich.*	*kein Fehler*

Definition 3.1.7. *Für einen Test* φ *ist die* *Gütefunktion* $G_\varphi : \Theta \to [0,1]$ *definiert durch*

$$G_\varphi(\vartheta) := \mathbb{E}_\vartheta[\varphi], \quad \vartheta \in \Theta.$$

Bemerkung 3.1.8. *Bekanntlich ist jeder Test ist also von der Form* $\varphi = \mathbb{1}_A$ *für eine Menge* $A \in \mathscr{F}$. *Für jedes* $\vartheta \in \Theta$ *ist* $G_\varphi(\vartheta) = \mathbb{P}_\vartheta(A)$ *die Wahrscheinlichkeit unter* $\mathbb{P}_\vartheta$, *sich für die Alternative* H_1 *zu entscheiden. Insbesondere gilt also:*

- *Falls* $\vartheta \in \Theta_0$, *so ist* $G_\varphi(\vartheta)$ *die Wahrscheinlichkeit unter* $\mathbb{P}_\vartheta$, *einen Fehler 1. Art zu begehen.*

- *Falls* $\vartheta \in \Theta_1$, *so ist* $1 - G_\varphi(\vartheta)$ *die Wahrscheinlichkeit unter* $\mathbb{P}_\vartheta$, *einen Fehler 2. Art zu begehen.*

Wir werden einen Test also als gut ansehen, wenn mit primärer Priorität $G_\varphi|_{\Theta_0}$ *klein (siehe Definition 3.1.9) und mit sekundärer Priorität* $G_\varphi|_{\Theta_1}$ *groß (siehe Definitionen 3.1.11 und 3.1.13) ist.*

Definition 3.1.9. *Es seien* φ *ein Test und* $\alpha \in (0,1)$.

(a) Der Test φ heißt <u>zulässig zum Irrtumsniveau α</u>, falls

$$\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta) \le \alpha.$$

In diesem Fall sagen wir auch, dass der Test φ das <u>Signifikanzniveau</u> α hat.

(b) Der Test φ heißt ein <u>Level-α-Test</u>, falls

$$\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta) = \alpha.$$

Satz 3.1.10 (Zweiseitiger Gauß-Test für μ)**.** *Wir betrachten das n-fache Gaußmodell mit bekannter Varianz $\sigma^2 > 0$. Wir zerlegen $\Theta = \mathbb{R}$ in*

$$\Theta_0 = \{\mu_0\} \quad \textit{und} \quad \Theta_1 = \mathbb{R} \setminus \{\mu_0\}$$

für ein $\mu_0 \in \mathbb{R}$. Dann erhalten wir die Null-Hypothese $H_0 = \{\mu = \mu_0\}$ und die Alternative $H_1 = \{\mu \neq \mu_0\}$. Weiterhin sei $\alpha \in (0,1)$ gegeben. Für $c \in \mathbb{R}$ definieren wir den Test

$$\varphi_c := \mathbb{1}_{\{|\bar{X} - \mu_0| \ge c\}}.$$

Dann gelten folgende Aussagen:

(a) Für jedes $c \ge \sqrt{\sigma^2/n} \cdot z_{1-\alpha/2}$ ist der Test φ_c zulässig zum Irrtumsniveau α.

(b) Für $c = \sqrt{\sigma^2/n} \cdot z_{1-\alpha/2}$ ist φ_c ein Level-α-Test.

Beweis. Es gilt nach Satz 2.2.3

$$\begin{aligned}
G_{\varphi_c}(\mu_0) &= \mathbb{E}_{\mu_0}[\varphi_c] = \mathbb{E}_{\mu_0}[\mathbb{1}_{\{|\bar{X}-\mu_0| \ge c\}}] = \mathbb{P}_{\mu_0}(|\bar{X} - \mu_0| \ge c) \\
&= 1 - \mathbb{P}_{\mu_0}(-c < \bar{X} - \mu_0 < c) \\
&= 1 - \mathbb{P}_{\mu_0}\Bigg(- \frac{c}{\sqrt{\sigma^2/n}} < \underbrace{\frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}}_{\sim \mathrm{N}(0,1)} < \frac{c}{\sqrt{\sigma^2/n}} \Bigg) \\
&= 1 - \Bigg[\Phi\Bigg(\frac{c}{\sqrt{\sigma^2/n}} \Bigg) - \Phi\Bigg(- \frac{c}{\sqrt{\sigma^2/n}} \Bigg) \Bigg] \\
&= 2 \Bigg[1 - \Phi\Bigg(\frac{c}{\sqrt{\sigma^2/n}} \Bigg) \Bigg].
\end{aligned}$$

(a) Also erhalten wir

$$\begin{aligned}
& G_{\varphi_c}(\mu_0) \leq \alpha \\
\Leftrightarrow\quad & 1 - \Phi\left(\frac{c}{\sqrt{\sigma^2/n}}\right) \leq \frac{\alpha}{2} \\
\Leftrightarrow\quad & \Phi\left(\frac{c}{\sqrt{\sigma^2/n}}\right) \geq 1 - \frac{\alpha}{2} \\
\Leftrightarrow\quad & \frac{c}{\sqrt{\sigma^2/n}} \geq z_{1-\alpha/2} \\
\Leftrightarrow\quad & c \geq \sqrt{\sigma^2/n} \cdot z_{1-\alpha/2}.
\end{aligned}$$

(b) Analog erhalten wir

$$G_{\varphi_c}(\mu_0) = \alpha \quad \Leftrightarrow \quad c = \sqrt{\sigma^2/n} \cdot z_{1-\alpha/2}.$$

□

Definition 3.1.11. *Es sei $\alpha \in (0,1)$ beliebig. Es sei φ ein Test, der zulässig zum Irrtumsniveau α ist. Weiterhin sei $\beta \in (0,1)$ beliebig. Eine Teilmenge $\Gamma_1 \subset \Theta_1$ heißt eine Indifferenzzone von φ zu den Irrtumsniveaus α und β, falls*

$$\sup_{\vartheta \in \Delta_1} (1 - G_\varphi(\vartheta)) \leq \beta,$$

wobei $\Delta_1 \subset \Theta_1$ gegeben ist durch $\Delta_1 := \Theta_1 \setminus \Gamma_1$.

Satz 3.1.12 (Einseitiger Gauß-Test für μ)**.** *Wir betrachten das n-fache Gaußmodell mit bekannter Varianz $\sigma^2 > 0$ mit einer nichtleeren, beliebigen Teilmenge $\Theta \subset \mathbb{R}$. Wir wählen $\mu_0 \in \Theta$ und zerlegen Θ in*

$$\Theta_0 = \{\mu \in \Theta : \mu \leq \mu_0\} \quad \textit{und} \quad \Theta_1 = \{\mu \in \Theta : \mu > \mu_0\}.$$

Dann erhalten wir die Null-Hypothese $H_0 = \{\mu \leq \mu_0\}$ und die Alternative $H_1 = \{\mu > \mu_0\}$. Weiterhin sei $\alpha \in (0,1)$ gegeben. Für $c \in \mathbb{R}$ definieren wir den Test

$$\varphi_c := \mathbb{1}_{\{\bar{X} - \mu_0 \geq c\}}.$$

Dann gelten folgende Aussagen:

(a) Die Gütefunktion $G_{\varphi_c} : \Theta \to \mathbb{R}$ ist gegeben durch

$$G_{\varphi_c}(\mu) = 1 - \Phi\left(\frac{c + \mu_0 - \mu}{\sqrt{\sigma^2/n}}\right), \quad \mu \in \Theta.$$

(b) Insbesondere ist die Gütefunktion $G_{\varphi_c} : \Theta \to \mathbb{R}$ *monoton wachsend.*

(c) Für jedes $c \geq \sqrt{\sigma^2/n} \cdot z_{1-\alpha}$ *ist der Test* φ_c *zulässig zum Irrtumsniveau* α*.*

(d) Für $c := \sqrt{\sigma^2/n} \cdot z_{1-\alpha}$ *ist* φ_c *ein Level-*α*-Test.*

(e) Es sei $c := \sqrt{\sigma^2/n} \cdot z_{1-\alpha}$*. Für alle* $\beta \in (0,1)$ *und* $\nu_0 > \mu_0$ *ist* $(\mu_0, \nu_0) \cap \Theta_1$ *genau dann eine Indifferenzzone von* φ_c *zu den Irrtumsniveaus* α *und* β*, wenn*

$$n \geq \frac{\sigma^2 (z_{1-\alpha} - z_\beta)^2}{(\nu_0 - \mu_0)^2}.$$

Beweis.

(a) Für jedes $\mu \in \Theta_0$ gilt nach Satz 2.2.3

$$\begin{aligned} G_{\varphi_c}(\mu) &= \mathbb{E}_\mu[\varphi_c(X)] = \mathbb{E}_\mu[\mathbb{1}_{\{\bar{X} - \mu_0 \geq c\}}] = \mathbb{P}_\mu(\bar{X} \geq c + \mu_0) \\ &= \mathbb{P}_\mu\Bigg(\underbrace{\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}}_{\sim \mathrm{N}(0,1)} \geq \frac{c + \mu_0 - \mu}{\sqrt{\sigma^2/n}} \Bigg) = 1 - \Phi\Bigg(\frac{c + \mu_0 - \mu}{\sqrt{\sigma^2/n}} \Bigg). \end{aligned}$$

(b) ✓

(c) Es folgt

$$\sup_{\mu \in \Theta_0} G_{\varphi_c}(\mu) = 1 - \Phi\Bigg(\frac{c}{\sqrt{\sigma^2/n}} \Bigg).$$

Somit erhalten wir

$$\begin{aligned} & \sup_{\mu \in \Theta_0} G_{\varphi_c}(\mu) \leq \alpha \\ \Leftrightarrow \quad & 1 - \Phi\Bigg(\frac{c}{\sqrt{\sigma^2/n}} \Bigg) \leq \alpha \\ \Leftrightarrow \quad & \Phi\Bigg(\frac{c}{\sqrt{\sigma^2/n}} \Bigg) \geq 1 - \alpha \\ \Leftrightarrow \quad & \frac{c}{\sqrt{\sigma^2/n}} \geq z_{1-\alpha} \\ \Leftrightarrow \quad & c \geq \sqrt{\sigma^2/n} \cdot z_{1-\alpha}. \end{aligned}$$

(d) Analog erhalten wir

$$\sup_{\mu \in \Theta_0} G_{\varphi_c}(\mu) = \alpha \quad \Leftrightarrow \quad c = \sqrt{\sigma^2/n} \cdot z_{1-\alpha}.$$

(e) Es sei $\mu \geq \nu_0$ beliebig. Nach Teil (a) gilt

$$G_\varphi(\mu) = 1 - \Phi\left(\frac{c + \mu_0 - \mu}{\sqrt{\sigma^2/n}}\right) = 1 - \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}}\right).$$

Also gilt

$$1 - G_\varphi(\mu) = \Phi\left(z_{1-\alpha} - \frac{\mu - \mu_0}{\sqrt{\sigma^2/n}}\right) \quad \text{für alle } \mu \geq \nu_0,$$

und somit

$$\sup_{\mu \geq \nu_0}(1 - G_\varphi(\mu)) = \Phi\left(z_{1-\alpha} - \frac{\nu_0 - \mu_0}{\sqrt{\sigma^2/n}}\right).$$

Daraus folgt

$$\begin{aligned}
& \sup_{\mu \geq \Delta}(1 - G_\varphi(\mu)) \leq \beta \\
\Leftrightarrow \quad & \Phi\left(z_{1-\alpha} - \frac{\nu_0 - \mu_0}{\sqrt{\sigma^2/n}}\right) \leq \beta \\
\Leftrightarrow \quad & z_{1-\alpha} - \frac{\nu_0 - \mu_0}{\sqrt{\sigma^2/n}} \leq z_\beta \\
\Leftrightarrow \quad & z_{1-\alpha} - z_\beta \leq \frac{\nu_0 - \mu_0}{\sqrt{\sigma^2/n}} \\
\Leftrightarrow \quad & \sqrt{n} \geq \frac{\sigma(z_{1-\alpha} - z_\beta)}{\nu_0 - \mu_0} \\
\Leftrightarrow \quad & n \geq \frac{\sigma^2(z_{1-\alpha} - z_\beta)^2}{(\nu_0 - \mu_0)^2}.
\end{aligned}$$

□

Definition 3.1.13. *Es sei $\alpha \in (0,1)$ beliebig. Ein Test φ heißt ein gleichmäßig bester Test zum Irrtumsniveau α (von H_0 gegen H_1), falls gilt:*

(a) Der Test φ ist zulässig zum Irrtumsniveau α.

(b) Es gilt

$$G_\varphi(\vartheta) \geq G_\psi(\vartheta) \quad \textit{für alle } \vartheta \in \Theta_1$$

und für jeden zulässigen Test ψ (von H_0 gegen H_1) zum Irrtumsniveau α.

3.2 Das Neymann-Pearson-Lemma

Es sei

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein Standardmodell mit $\Theta = \{\vartheta_0, \vartheta_1\}$ für $\vartheta_0 \neq \vartheta_1$. Wir zerlegen Θ in $\Theta_0 = \{\vartheta_0\}$ und $\Theta_1 = \{\vartheta_1\}$. Wir betrachten das Testproblem $H_0 = \{\vartheta = \vartheta_0\}$ gegen $H_1 = \{\vartheta = \vartheta_1\}$. Wir nehmen an, dass $\mathbb{P}_{\vartheta_0} \approx \mu \approx \mathbb{P}_{\vartheta_1}$, wobei μ das dominierende Maß bezeichnet. Weiterhin sei $L : \Theta \times \Omega \to (0, \infty)$ die Likelihood-Funktion.

Definition 3.2.1.

(a) Wir definieren den <u>Likelihood-Quotienten</u> durch

$$L(\vartheta_0, \vartheta_1, x) := \frac{L(\vartheta_1, x)}{L(\vartheta_0, x)}.$$

(b) Für $k \in \mathbb{R}_+$ definieren den <u>Likelihood-Quotienten-Test</u> durch

$$\varphi_k := \mathbb{1}_{\{L(\vartheta_0, \vartheta_1, X) \geq k\}}.$$

Satz 3.2.2 (Neyman-Pearson-Lemma)**.** *Es seien $\alpha \in (0,1)$ und $k \in \mathbb{R}_+$, so dass der Likelihood-Quotienten-Test φ_k ein Level-α-Test ist. Dann ist φ_k ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α.*

Beweis. Es sei ψ ein beliebiger Test von H_0 gegen H_1 zum Irrtumsniveau α. Weiterhin sei $x \in \Omega$ beliebig. Wir unterscheiden zwei Fälle:

(1) Es gelte $L(\vartheta_1, x) - kL(\vartheta_0, x) \geq 0$. Dann gilt

$$L(\vartheta_0, \vartheta_1, x) - k = \frac{L(\vartheta_1, x)}{L(\vartheta_0, x)} - k = \frac{L(\vartheta_1, x) - kL(\vartheta_0, x)}{L(\vartheta_0, x)} \geq 0.$$

Also gilt $L(\vartheta_0, \vartheta_1, x) \geq k$, und daher $\varphi_k(x) = 1$. Also folgt

$$\varphi_k(x) - \psi(x) \geq 0.$$

(2) Es gelte $L(\vartheta_1, x) - kL(\vartheta_0, x) < 0$. Dann gilt

$$L(\vartheta_0, \vartheta_1, x) - k = \frac{L(\vartheta_1, x)}{L(\vartheta_0, x)} - k = \frac{L(\vartheta_1, x) - kL(\vartheta_0, x)}{L(\vartheta_0, x)} < 0.$$

Also gilt $L(\vartheta_0, \vartheta_1, x) < k$, und daher $\varphi_k(x) = 0$. Also folgt

$$\varphi_k(x) - \psi(x) \leq 0.$$

Insgesamt folgt

$$(\varphi_k(x) - \psi(x))(L(\vartheta_1, x) - kL(\vartheta_0, x)) \geq 0 \quad \text{für alle } x \in \Omega.$$

Also gilt

$$\int_\Omega (\varphi_k(x) - \psi(x))(L(\vartheta_1, x) - kL(\vartheta_0, x))\mu(dx) \geq 0.$$

Hieraus folgt

$$\mathbb{E}_{\vartheta_1}[\varphi_k - \psi] - k\,\mathbb{E}_{\vartheta_0}[\varphi_k - \psi] \geq 0.$$

Da φ_k ein Level-α-Test ist, gilt $\mathbb{E}_{\vartheta_0}[\varphi_k] = \alpha$. Außerdem gilt $\mathbb{E}_{\vartheta_0}[\psi] \leq \alpha$, und es folgt

$$\mathbb{E}_{\vartheta_1}[\varphi_k] - \mathbb{E}_{\vartheta_1}[\psi] \geq k\,\mathbb{E}_{\vartheta_0}[\varphi_k] - k\,\mathbb{E}_{\vartheta_0}[\psi] = k(\alpha - \mathbb{E}_{\vartheta_0}[\psi]) \geq 0.$$

Also gilt

$$\mathbb{E}_{\vartheta_1}[\varphi_k] \geq \mathbb{E}_{\vartheta_1}[\psi].$$

$\square$

Satz 3.2.3. *Wir betrachten das n-fache Gaußmodell mit bekannter Varianz $\sigma^2 > 0$ und Parameterraum $\Theta = \{\mu_0, \mu_1\}$ für $\mu_0 < \mu_1$. Wir zerlegen Θ in*

$$\Theta_0 = \{\mu_0\} \quad \textit{und} \quad \Theta_1 = \{\mu_1\}.$$

Dann erhalten wir die Null-Hypothese $H_0 = \{\mu = \mu_0\}$ und die Alternative $H_1 = \{\mu = \mu_1\}$. Weiterhin sei $\alpha \in (0,1)$ gegeben. Wir definieren den Test

$$\varphi_c = \mathbb{1}_{\{\bar{X} - \mu_0 \geq c\}},$$

wobei $c = \sqrt{\sigma^2/n} \cdot z_{1-\alpha}$. Dann ist φ_c ein gleichmäßig bester Test zum Irrtumsniveau α.

Beweis. Nach Satz 3.1.12(d) ist der Test φ_c ein Level-α-Test. Weiterhin gilt

$$L(\mu_0, x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(- \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu_0)^2 \right)$$

und

$$\begin{aligned} L(\mu_1, x) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(- \frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu_1)^2 \right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(- \frac{1}{2\sigma^2} \sum_{i=1}^n ((X_i - \mu_0) - (\mu_1 - \mu_0))^2 \right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(- \frac{1}{2\sigma^2} \sum_{i=1}^n \left((X_i - \mu_0)^2 - 2(\mu_1 - \mu_0)(X_i - \mu_0) + (\mu_1 - \mu_0)^2\right) \right) \end{aligned}$$

Also gilt

$$
\begin{aligned}
L(\mu_0,\mu_1,x) = \frac{L(\mu_1,x)}{L(\mu_0,x)} &= \exp\Bigg(-\frac{1}{2\sigma^2}\sum_{i=1}^n \big(-2(\mu_1-\mu_0)(X_i-\mu_0) + (\mu_1-\mu_0)^2\big)\Bigg) \\
&= \exp\left(\frac{\mu_1-\mu_0}{\sigma^2}\sum_{i=1}^n (X_i-\mu_0) - \frac{n(\mu_1-\mu_0)^2}{2\sigma^2}\right) \\
&= \exp\left(\frac{n(\mu_1-\mu_0)}{\sigma^2}(\bar{X}-\mu_0) - \frac{n(\mu_1-\mu_0)^2}{2\sigma^2}\right) \\
&= \exp\left(\frac{\mu_1-\mu_0}{(\sigma^2/n)}(\bar{X}-\mu_0) - \frac{(\mu_1-\mu_0)^2}{2(\sigma^2/n)}\right)
\end{aligned}
$$

Wir definieren die streng monoton wachsende, bijektive Funktion

$$
\Phi : (0,\infty) \to \mathbb{R}, \quad \Phi(y) := \frac{\sqrt{\sigma^2/n}}{\mu_1-\mu_0}\left(\ln y + \frac{(\mu_1-\mu_0)^2}{2(\sigma^2/n)}\right).
$$

Dann gilt

$$
\begin{aligned}
\Phi(L(\mu_0,\mu_1,x)) &= \frac{\sqrt{\sigma^2/n}}{\mu_1-\mu_0}\left(\ln L(\mu_0,\mu_1,x) + \frac{(\mu_1-\mu_0)^2}{2(\sigma^2/n)}\right) \\
&= \frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}.
\end{aligned}
$$

Mit

$$
k := \Phi^{-1}\left(\frac{c}{\sqrt{\sigma^2/n}}\right)
$$

folgt

$$
\{\bar{X}-\mu_0 \geq c\} = \left\{\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \geq \frac{c}{\sqrt{\sigma^2/n}}\right\} = \{L(\mu_0,\mu_1,X) \geq k\}.
$$

Nach dem Neyman-Pearson-Lemma (Satz 3.2.2) ist φ_c also ein gleichmäßig bester Test. $\square$

3.3 Erweiterung auf allgemeinere Bereiche

Satz 3.3.1. *Es seien $\Theta \subset \mathbb{R}$ und $\vartheta_0 \in \Theta$ beliebig. Wir zerlegen Θ in*

$$
\Theta_0 = (-\infty,\vartheta_0] \cap \Theta \quad \text{und} \quad \Theta_1 = (\vartheta_0,\infty) \cap \Theta.
$$

Dann erhalten wir die Null-Hypothese $H_0 = \{\vartheta \leq \vartheta_0\}$ und die Alternative $H_1 = \{\vartheta > \vartheta_0\}$. Es seien $\alpha \in (0,1)$ und $\varphi : \Omega \to \{0,1\}$ eine Statistik, so dass gilt:

(1) Die Gütefunktion $\vartheta \mapsto G_\varphi(\vartheta)$ ist monoton wachsend auf Θ_0.

(2) Für alle $\vartheta_1 \in \Theta_1$ ist φ ein gleichmäßig bester Test von $\{\vartheta = \vartheta_0\}$ gegen $\{\vartheta = \vartheta_1\}$ zum Irrtumsniveau α.

Dann ist φ ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α.

Beweis. Wegen der Monotonie der Gütefunktion gilt

$$\sup_{\vartheta \in \Theta_0} G_\varphi(\vartheta) = G_\varphi(\vartheta_0) \leq \alpha.$$

Also ist φ ein zulässiger Test von H_0 gegen H_1 zum Irrtumsniveau α. Nun sei ψ ein zulässiger Test von H_0 gegen H_1 zum Irrtumsniveau α. Es gilt also

$$\sup_{\vartheta \in \Theta_0} G_\psi(\vartheta) \leq \alpha,$$

und damit insbesondere

$$G_\psi(\vartheta_0) \leq \alpha.$$

Also ist ψ für jedes $\vartheta_1 \in \Theta_1$ ein zulässiger Test von $\{\vartheta = \vartheta_0\}$ gegen $\{\vartheta = \vartheta_1\}$ zum Irrtumsniveau α. Es folgt

$$G_\varphi(\vartheta_1) \geq G_\psi(\vartheta_1) \quad \text{für alle } \vartheta_1 \in \Theta_1.$$

Also ist φ ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α. □

Satz 3.3.2 (Einseitiger Gauß-Test für μ)**.** *Wir betrachten das n-fache Gaußmodell mit bekannter Varianz $\sigma^2 > 0$ mit $\Theta = \mathbb{R}$. Wir wählen $\mu_0 \in \mathbb{R}$ und zerlegen Θ in*

$$\Theta_0 = (-\infty, \mu_0] \quad \textit{und} \quad \Theta_1 = (\mu_0, \infty).$$

Dann erhalten wir die Null-Hypothese $H_0 = \{\mu \leq \mu_0\}$ und die Alternative $H_1 = \{\mu > \mu_0\}$. Weiterhin sei $\alpha \in (0,1)$ gegeben. Wir definieren den Test

$$\varphi_c := \mathbb{1}_{\{\bar{X} - \mu_0 \geq c\}},$$

wobei $c := \sqrt{\sigma^2/n} \cdot z_{1-\alpha}$. Dann ist φ_c ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α.

Beweis. Nach Satz 3.1.12(b) ist die Gütefunktion G_{φ_c} monoton wachsend, und nach Satz 3.2.3 ist φ_c für jedes $\mu_1 \in \Theta_1$ ein gleichmäßig bester Test von $\{\mu = \mu_0\}$ gegen $\{\mu = \mu_1\}$ zum Irrtumsniveau α. Mit Satz 3.3.1 folgt, dass φ_c ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α ist. □

3.4 Exponentielle Familien

Es sei $\mathscr{M}$ eine exponentielle Familie mit Likelihood-Funktion

$$L(\vartheta, x) = \exp\big(a(\vartheta)T(x) - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in \Theta \times \Omega.$$

Definition 3.4.1. *Die exponentielle Familie $\mathscr{M}$ heißt monoton wachsend, falls $a : \Theta \to \mathbb{R}$ streng monoton wachsend ist.*

Lemma 3.4.2. *Die exponentielle Familie $\mathscr{M}$ sei monoton wachsend. Dann existiert für alle $\vartheta_0, \vartheta_1 \in \Theta$ mit $\vartheta_0 < \vartheta_1$ eine streng monoton wachsende, bijektive Funktion $\Phi = \Phi_{\vartheta_0,\vartheta_1} : (0, \infty) \to \mathbb{R}$, so dass*

$$\Phi(L(\vartheta_0, \vartheta_1, x)) = T(x) \quad \textit{für alle } x \in \Omega.$$

Beweis. Wir definieren $\Phi : (0, \infty) \to \mathbb{R}$ durch

$$\Phi(y) := \frac{1}{a(\vartheta_1) - a(\vartheta_0)} \Big(\ln y + (b(\vartheta_1) - b(\vartheta_0)) \Big).$$

Dann ist Φ streng monoton wachsend und bijektiv, und wegen

$$L(\vartheta_0, \vartheta_1, x) = \frac{L(\vartheta_1, x)}{L(\vartheta_0, x)} = \exp\big((a(\vartheta_1) - a(\vartheta_0))T(x) - (b(\vartheta_1) - b(\vartheta_0))\big), \quad x \in \Omega$$

folgt

$$\Phi(L(\vartheta_0, \vartheta_1, x)) = T(x) \quad \text{für alle } x \in \Omega.$$

□

Satz 3.4.3. *Die exponentielle Familie $\mathscr{M}$ sei monoton wachsend. Es sei $\Theta = \{\vartheta_0, \vartheta_1\}$ mit $\vartheta_0 < \vartheta_1$. Wir zerlegen Θ in*

$$\Theta_0 = \{\vartheta_0\} \quad \textit{und} \quad \Theta_1 = \{\vartheta_1\}.$$

Dann erhalten wir die Null-Hypothese $H_0 = \{\vartheta = \vartheta_0\}$ und die Alternative $H_1 = \{\vartheta = \vartheta_1\}$. Es seien $\alpha \in (0, 1)$ und $\gamma \in \mathbb{R}$, so dass

$$\varphi_\gamma := \mathbb{1}_{\{T(X) \geq \gamma\}}$$

ein Level-α-Test ist. Dann ist φ_γ ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α.

Beweis. Nach Lemma 3.4.2 gilt

$$\varphi_\gamma = \mathbb{1}_{\{T(X)\geq\gamma\}} = \mathbb{1}_{\{L(\vartheta_0,\vartheta_1,X)\geq k\}},$$

wobei $k = \Phi^{-1}(\gamma)$. Nach dem Neyman-Pearson-Lemma (Satz 3.2.2) ist φ_γ ein gleichmäßig bester Test von $H_0 = \{\vartheta = \vartheta_0\}$ gegen $H_1 = \{\vartheta = \vartheta_1\}$ zum Irrtumsniveau α. □

Lemma 3.4.4. *Die exponentielle Familie $\mathscr{M}$ sei monoton wachsend. Dann ist für jedes $\gamma \in \mathbb{R}$ die Gütefunktion $G_{\varphi_\gamma} : \Theta \to [0,1]$ des Tests $\varphi_\gamma = \mathbb{1}_{\{T(X)\geq\gamma\}}$ monoton wachsend.*

Beweis. Es seien $\vartheta_1, \vartheta_2 \in \Theta$ mit $\vartheta_1 < \vartheta_2$ beliebig. Weiterhin sei $x \in \Omega$ beliebig. Nach Lemma 3.4.2 gilt $T(x) \geq \gamma$ genau dann, wenn

$$\frac{L(\vartheta_1, x)}{L(\vartheta_0, x)} = L(\vartheta_0, \vartheta_1, x) \geq k,$$

wobei $k = \Phi^{-1}(\gamma)$. Also folgt für alle $x \in \Omega$ mit $T(x) \geq \gamma$, dass $\varphi_\gamma(x) = 1$ und

$$\begin{aligned} &(L(\vartheta_1, x) - kL(\vartheta_0, x))(\varphi_\gamma(x) - b) \geq 0 \quad \text{für alle } b \in [0,1] \\ \Leftrightarrow\ &(L(\vartheta_1, x) - kL(\vartheta_0, x))\varphi_\gamma(x) \geq (L(\vartheta_1, x) - kL(\vartheta_0, x))b \quad \text{für alle } b \in [0,1]. \end{aligned}$$

Daraus folgt

$$\mathbb{E}_{\vartheta_1}[\varphi] - k\mathbb{E}_{\vartheta_0}[\varphi] \geq (1-k)b \quad \text{für alle } b \in [0,1].$$

Mit $b := \mathbb{E}_{\vartheta_0}[\varphi] \in [0,1]$ folgt

$$\mathbb{E}_{\vartheta_1}[\varphi] \geq \mathbb{E}_{\vartheta_0}[\varphi].$$

□

Satz 3.4.5. *Die exponentielle Familie $\mathscr{M}$ sei monoton wachsend. Es seien $\Theta \subset \mathbb{R}$ und $\vartheta_0 \in \Theta$ beliebig. Wir zerlegen Θ in*

$$\Theta_0 = (-\infty, \vartheta_0] \cap \Theta \quad \textit{und} \quad \Theta_1 = (\vartheta_0, \infty) \cap \Theta.$$

Dann erhalten wir die Null-Hypothese $H_0 = \{\vartheta \leq \vartheta_0\}$ und die Alternative $H_1 = \{\vartheta > \vartheta_0\}$. Es seien $\alpha \in (0,1)$ und $\gamma \in \mathbb{R}$, so dass für

$$\varphi_\gamma := \mathbb{1}_{\{T(X)\geq\gamma\}}$$

gilt $G_{\varphi_\gamma}(\vartheta_0) = \alpha$. Dann ist φ_γ ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α.

Beweis. Nach Lemma 3.4.4 ist die Gütefunktion $G_{\varphi_\gamma} : \Theta \to [0,1]$ monoton wachsend. Nach Satz 3.4.3 ist für alle $\vartheta_1 \in \Theta_1$ der Test φ_γ ein gleichmäßig bester Test von $\{\vartheta = \vartheta_0\}$ gegen $\{\vartheta = \vartheta_1\}$ zum Irrtumsniveau α. Also folgt mit Satz 3.3.1, dass φ_γ ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α ist. □

Beispiel 3.4.6. *Für das n-fache Gaußmodell mit bekannter Varianz $\sigma^2 > 0$ können wir die Aussage von Satz 3.3.2 nun leicht alternativ beweisen. In der Tat, es handelt sich um eine exponentielle Familie mit*

$$a(\mu) = \frac{n\mu}{\sigma^2} \quad \textit{und} \quad T(X) = \bar{X}.$$

Die Funktion a ist streng monoton wachsend. Also ist die exponentielle Familie $\mathscr{M}$ monoton. Wir setzen $\gamma := \mu_0 + c$ mit $c := \sqrt{\sigma^2/n} \cdot z_{1-\alpha}$ und betrachten den Test

$$\varphi_\gamma = \mathbb{1}_{\{T(X) \geq \gamma\}}.$$

Nach Satz 3.1.12(d) gilt $G_{\varphi_\gamma}(\mu_0) = \alpha$. Also folgt mit Satz 3.4.5, dass φ_γ ein gleichmäßig bester Test von H_0 gegen H_1 zum Irrtumsniveau α ist.

3.5 Dualität zwischen Konfidenzmengen und Hypothesentests

In diesem Abschnitt sei

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein statistisches Modell.

Satz 3.5.1. *Es sei $\alpha \in (0,1)$ beliebig, und es sei $C : \Omega \to \mathscr{E}$ eine Konfidenzmenge für ϑ zum Niveau α. Weiterhin sei $\vartheta_0 \in \Theta$ beliebig. Wir zerlegen Θ in*

$$\Theta_0 = \{\vartheta_0\} \quad \textit{und} \quad \Theta_1 = \Theta \setminus \{\vartheta_0\}.$$

Das heißt, wird erhalten die Null-Hypothese $H_0 = \{\vartheta = \vartheta_0\}$ und die Alternative $H_1 = \{\vartheta \neq \vartheta_0\}$. Dann ist

$$\varphi : \Omega \to \{0,1\}, \quad \varphi := \mathbb{1}_{\{\vartheta_0 \notin C\}}$$

ein Test, der zulässig zum Irrtumsniveau α ist.

Beweis. Nach Voraussetzung gilt $\{\vartheta \in C\} \in \mathscr{F}$ und

$$\mathbb{P}_\vartheta(\vartheta \in C) \geq 1 - \alpha \quad \text{für alle } \vartheta \in \Theta.$$

Also gilt insbesondere $\{\vartheta_0 \in C\} \in \mathscr{F}$, und folglich ist φ ein Test. Weiterhin gilt

$$\begin{aligned} G_\varphi(\vartheta_0) &= \mathbb{E}_{\vartheta_0}[\varphi] = \mathbb{E}_{\vartheta_0}[\mathbb{1}_{\{\vartheta_0 \notin C\}}] = \mathbb{P}_{\vartheta_0}(\vartheta_0 \notin C) \\ &= 1 - \mathbb{P}_{\vartheta_0}(\vartheta_0 \in C) \leq 1 - (1-\alpha) = \alpha. \end{aligned}$$

□

Nun erhalten wir leicht einen alternativen Beweis von Satz 3.1.10.

Korollar 3.5.2. *Wir betrachten das n-fache Gaußmodell mit bekannter Varianz $\sigma^2 > 0$. Wir zerlegen $\Theta = \mathbb{R}$ in*

$$\Theta_0 = \{\mu_0\} \quad \textit{und} \quad \Theta_1 = \mathbb{R} \setminus \{\mu_0\}$$

für ein $\mu_0 \in \mathbb{R}$. Das heißt, wir erhalten die Null-Hypothese $H_0 = \{\mu = \mu_0\}$ und die Alternative $H_1 = \{\mu \neq \mu_0\}$. Weiterhin sei $\alpha \in (0,1)$ gegeben. Dann ist

$$\varphi := \mathbb{1}_{\left\{|\bar{X} - \mu_0| > \sqrt{\sigma^2/n} \cdot z_{1-\alpha/2}\right\}}$$

ein Test, der zulässig zum Irrtumsniveau α ist.

Beweis. Nach Satz 2.2.5(b) ist das Zufallsintervall $C : \Omega \to \mathcal{B}(\mathbb{R})$ gegeben durch

$$C = \bar{X} \pm z_{1-\alpha/2}\sqrt{\sigma^2/n}$$

ein Konfidenzintervall für μ zum Niveau α. Außerdem gilt

$$\{\mu_0 \notin C\} = \left\{\mu_0 \notin \bar{X} \pm z_{1-\alpha/2}\sqrt{\sigma^2/n}\right\} = \left\{|\bar{X} - \mu_0| > \sqrt{\sigma^2/n} \cdot z_{1-\alpha/2}\right\}.$$

Nach Satz 3.5.1 folgt, dass φ ein Test ist, der zulässig zum Irrtumsniveau α ist. □

Satz 3.5.3. *Es sei $\alpha \in (0,1)$ beliebig, und es sei $\mathscr{T}$ eine σ-Algebra über Θ. Weiterhin sei $(\varphi_\vartheta)_{\vartheta \in \Theta}$ eine Familie von Tests, so dass gilt:*

- *Für jedes $\vartheta_0 \in \Theta$ ist φ_{ϑ_0} ein Test, der für $H_0 = \{\vartheta = \vartheta_0\}$ gegen $H_1 = \{\vartheta \neq \vartheta_0\}$ zulässig zum Irrtumsniveau α ist.*
- *Die Abbildung $\Theta \times \Omega \to \{0,1\}$, $(\vartheta, x) \mapsto \varphi_\vartheta(x)$ ist $\mathscr{T} \otimes \mathscr{F}$-messbar.*

Dann ist

$$C : \Omega \to \mathscr{T}, \quad C(x) = \{\vartheta \in \Theta : \varphi_\vartheta(x) = 0\}$$

eine Konfidenzmenge für ϑ zum Niveau α.

Beweis. Es sei $\vartheta \in \Theta$ beliebig. Nach Voraussetzung gilt

$$\mathbb{P}_\vartheta(\varphi_\vartheta = 1) = \mathbb{E}_\vartheta[\varphi_\vartheta] = G_{\varphi_\vartheta}(\vartheta) \leq \alpha.$$

Nach Voraussetzung gilt außerdem

$$\{\vartheta \in C\} = \{\varphi_\vartheta = 0\} \in \mathscr{F}$$

sowie

$$\mathbb{P}_\vartheta(\vartheta \in C) = \mathbb{P}_\vartheta(\varphi_\vartheta = 0) = 1 - \mathbb{P}_\vartheta(\varphi_\vartheta = 1) \geq 1 - \alpha.$$

□

Nun erhalten wir leicht einen alternativen Beweis von Satz 2.2.5.

Korollar 3.5.4. *Wir betrachten das n-fache Gaußmodell mit bekannter Varianz $\sigma^2 > 0$. Für jedes $\alpha \in (0,1)$ ist das Zufallsintervall $C : \Omega \to \mathcal{B}(\mathbb{R})$ gegeben durch*

$$C = \bar{X} \pm z_{1-\alpha/2}\sqrt{\sigma^2/n}$$

ein Konfidenzintervall für μ zum Niveau α.

Beweis. Hier sind $\Theta = \mathbb{R}$ und $\mathscr{T} = \mathcal{B}(\mathbb{R})$. Nach Satz 3.1.10 ist

$$\varphi_{\mu_0} = \mathbb{1}_{\left\{|\bar{X}-\mu_0| > \sqrt{\sigma^2/n}\cdot z_{1-\alpha/2}\right\}}$$

für jedes $\mu_0 \in \Theta$ ein Test, der für $H_0 = \{\mu = \mu_0\}$ gegen $H_1 = \{\mu \neq \mu_0\}$ zulässig zum Irrtumsniveau α ist. Außerdem ist $(\mu, x) \mapsto \varphi_\mu(x)$ bezüglich $\mathscr{T} \otimes \mathscr{F}$ messbar. Weiterhin gilt

$$\{\mu \in \mathbb{R} : \varphi_\mu = 0\} = \left\{\mu \in \mathbb{R} : |\bar{X} - \mu| \leq \sqrt{\sigma^2/n} \cdot z_{1-\alpha/2}\right\} = \bar{X} \pm z_{1-\alpha/2}\sqrt{\sigma^2/n}.$$

Also ist C nach Satz 3.5.3 ein Konfidenzintervall für μ zum Niveau α. □

Kapitel 4

Nichtparametrische Modelle

4.1 Der Satz von Glivenko-Cantelli

Ein statistisches Modell

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

heißt bekanntlich ein parametrisches Modell, falls $\Theta \subset \mathbb{R}^d$ für ein $d \in \mathbb{N}$. Andernfalls sprechen wir von einem nichtparametrischen Modell.

Beispiel 4.1.1. *Es sei Θ die Menge aller Wahrscheinlichkeitsmaße auf $(\Omega, \mathscr{F})$. Für jedes $\vartheta \in \Theta$ sei $\mathbb{P}_\vartheta = \vartheta$. Dann ist $\mathscr{M}$ ein nichtparametrisches Modell.*

Beispiel 4.1.2. *Es seien Ω eine abzählbare Menge, und $\mathscr{F} = \mathfrak{P}(\Omega)$ die Potenzmenge. Weiterhin sei Θ die Menge aller stochastischen Vektoren; also aller Abbildungen $\vartheta : \Omega \to [0,1]$, so dass*

$$\sum_{k \in \Omega} \vartheta(k) = 1.$$

Für jedes $\vartheta \in \Theta$ sei $\mathbb{P}_\vartheta$ das Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{F})$ gegeben durch

$$\mathbb{P}_\vartheta(B) = \sum_{k \in B} \vartheta(k), \quad B \in \mathscr{F}.$$

Dann ist $\mathscr{M}$ ein nichtparametrisches Modell.

Beispiel 4.1.3. *Es sei $(\Omega, \mathscr{F}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Weiterhin sei Θ die Menge aller Äquivalenzklassen von Wahrscheinlichkeitsdichten; also aller messbaren Funktionen $\vartheta : \mathbb{R}^d \to \mathbb{R}_+$, so dass*

$$\int_{\mathbb{R}^d} \vartheta(x)dx = 1.$$

Für jedes $\vartheta \in \Theta$ sei $\mathbb{P}_\vartheta$ das Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{F})$ gegeben durch

$$\mathbb{P}_\vartheta(B) = \int_B \vartheta(x)dx, \quad B \in \mathscr{F}.$$

Dann ist $\mathscr{M}$ ein nichtparametrisches Modell.

Beispiel 4.1.4. *Es sei $(\Omega, \mathscr{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Weiterhin sei Θ die Menge aller Verteilungsfunktionen $\vartheta : \mathbb{R} \to [0,1]$. Für jedes $\vartheta \in \Theta$ sei $\mathbb{P}_\vartheta$ das Wahrscheinlichkeitsmaß auf $(\Omega, \mathscr{F})$ gegeben durch*

$$\mathbb{P}_\vartheta((-\infty, t]) = \vartheta(t), \quad t \in \mathbb{R}_+.$$

Dann ist $\mathscr{M}$ ein nichtparametrisches Modell.

Nun sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Verteilungsfunktion F.

Definition 4.1.5. *Für jedes $n \in \mathbb{N}$ definieren wir die empirische Verteilungsfunktion $F_n : \mathbb{R} \times \Omega \to [0,1]$ durch*

$$F_n(x) := \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_j \le x\}}, \quad x \in \mathbb{R}.$$

Satz 4.1.6.

(a) Es gilt für $n \to \infty$

$$F_n(x) \overset{f.s.}{\to} F(x) \quad \textit{für alle } x \in \mathbb{R}.$$

(b) Es gilt für $n \to \infty$

$$\mathbb{P} \circ \big(\sqrt{n}(F_n(x) - F(x))\big) \overset{w}{\to} \mathrm{N}(0, \sigma^2(x)) \quad \textit{für alle } x \in \mathbb{R},$$

wobei $\sigma^2(x) = F(x)(1 - F(x))$.

Beweis. Es sei $x \in \mathbb{R}$ fest. Wir definieren die Folge $(Y_n(x))_{n\in\mathbb{N}}$ durch

$$Y_n(x) := \mathbb{1}_{\{X_n \le x\}}.$$

Dann sind die Zufallsvariablen $(Y_n(x))_{n\in\mathbb{N}} \subset \mathscr{L}^2$ unabhängig und identisch verteilt mit

$$\begin{aligned} \mathbb{E}[Y_n(x)] &= \mathbb{P}(X_n \le x) = F(x), \\ \mathrm{Var}[Y_n(x)] &= \mathbb{E}[Y_n(x)^2] - \mathbb{E}[Y_n(x)]^2 = F(x) - F(x)^2 = F(x)(1 - F(x)). \end{aligned}$$

Außerdem erhalten wir die Darstellung

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} Y_j(x) \quad \text{für jedes } n \in \mathbb{N}.$$

(a) Nach dem Gesetz der großen Zahlen (Satz 1.2.6) folgt für $n \to \infty$

$$F_n(x) \stackrel{f.s.}{\to} F(x).$$

(b) Wir setzen

$$S_n(x) := \sum_{j=1}^{n} Y_j(x) \quad \text{für jedes } n \in \mathbb{N}.$$

Dann gilt für jedes $n \in \mathbb{N}$

$$\sqrt{n}(F_n(x) - F(x)) = \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} Y_j(x) - F(x)\right) = \frac{S_n(x) - nF(x)}{\sqrt{n}}.$$

Nach dem zentralen Grenzwertsatz (Satz 1.2.9) folgt für $n \to \infty$

$$\mathbb{P} \circ \left(\sqrt{n}(F_n(x) - F(x))\right) \stackrel{w}{\to} \mathrm{N}(0, \sigma^2(x)).$$

$\square$

Definition 4.1.7. *Es sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Die Quantilfunktion $F^{-1} : (0,1) \to \mathbb{R}$ ist gegeben durch*

$$F^{-1}(y) := \inf\{x \in \mathbb{R} : F(x) \geq y\}.$$

Bemerkung 4.1.8. *Nur wenn F stetig und streng monoton wachsend ist, dann ist F^{-1} die Umkehrfunktion von F im üblichen Sinne.*

Bemerkung 4.1.9. *Die Quantilfunktion F^{-1} ist stets monoton wachsend und linksstetig.*

Lemma 4.1.10. *Es sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Für alle $x \in \mathbb{R}$ und $y \in (0,1)$ sind folgende Aussagen äquivalent:*

(i) Es gilt $y \leq F(x)$.

(ii) Es gilt $F^{-1}(y) \leq x$.

Beweis. (i) $\Rightarrow$ (ii): Folgt aus Definition 4.1.7.
(ii) $\Rightarrow$ (i): Es gelte $F(x) < y$. Wegen der Rechtsstetigkeit von F exsitiert ein $\epsilon > 0$, so dass $F(x + \epsilon) < y$. Mit Definition 4.1.7 folgt $x + \epsilon \leq F^{-1}(y)$, und folglich $x < F^{-1}(y)$. $\square$

Lemma 4.1.11. *Es sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion. Dann gilt*

$$F\big(F^{-1}(y)-\big) \le y \le F\big(F^{-1}(y)\big) \quad \textit{für alle } y \in (0,1).$$

Beweis. Wir definieren die streng monoton wachsende Folge $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ durch

$$x_n := F^{-1}(y) - \frac{1}{n} \quad \text{für alle } n \in \mathbb{N}.$$

Dann gilt $x_n \uparrow F^{-1}(y)$, und nach Definition 4.1.7 gilt

$$F(x_n) < y \quad \text{für alle } n \in \mathbb{N}.$$

Da F linksseitige Grenzwerte besitzt, folgt

$$F\big(F^{-1}(y)-\big) = \lim_{n\to\infty} F(x_n) \le y.$$

Es sei $x := F^{-1}(y)$. Dann gilt $F^{-1}(y) \le x$, und daher nach Lemma 4.1.10

$$F\big(F^{-1}(y)\big) = F(x) \ge y.$$

□

Satz 4.1.12 (Satz von Glivenko-Cantelli, Hauptsatz der Mathematischen Statistik)**.** *Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit Verteilungsfunktion F. Dann gilt*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \overset{f.s.}{\to} 0 \quad \textit{für } n \to \infty,$$

wobei die $(F_n)_{n\in\mathbb{N}}$ die empirischen Verteilungsfunktionen sind.

Beweis. Es sei $x \in \mathbb{R}$ beliebig. Dann gilt nach Satz 4.1.6(a) für $n \to \infty$

$$F_n(x) \overset{f.s.}{\to} F(x).$$

Wir definieren die Folge $(Z_n(x))_{n\in\mathbb{N}}$ durch

$$Z_n(x) := \mathbb{1}_{\{X_n < x\}}.$$

Dann sind die Zufallsvariablen $(Z_n(x))_{n\in\mathbb{N}} \subset \mathscr{L}^2$ unabhängig und identisch verteilt mit

$$\mathbb{E}[Z_n(x)] = \mathbb{P}(X_n < x) = F(x-).$$

Außerdem erhalten wir die Darstellung

$$F_n(x-) = \frac{1}{n}\sum_{j=1}^{n} Z_j(x) \quad \text{für jedes } n \in \mathbb{N}.$$

Nach dem Gesetz der großen Zahlen (Satz 1.2.6) folgt für $n \to \infty$

$$F_n(x-) \stackrel{f.s.}{\to} F(x-).$$

Also existiert ein $N \in \mathscr{F}$ mit $\mathbb{P}(N) = 0$, so dass für $n \to \infty$

$$F_n(x, \omega) \to F(x) \quad \text{und} \quad F_n(x-, \omega) \to F(x-)$$

für alle $x \in F^{-1}((0,1) \cap \mathbb{Q})$ und alle $\omega \in N^c$.

Es sei $\omega \in N^c$ beliebig. Weiterhin sei $\epsilon > 0$ beliebig. Dann existiert ein $k \in \mathbb{N}$ mit $\frac{1}{k} < \frac{\epsilon}{2}$. Wir definieren $(x_i)_{i=0,\dots,k} \subset [-\infty, \infty]$ durch

$$\begin{aligned} x_0 &:= -\infty, \\ x_i &:= F^{-1}\Big(\frac{i}{k}\Big), \quad i = 1, \dots, k-1, \\ x_k &:= \infty. \end{aligned}$$

Dann existiert ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$ und alle $i = 0, \dots, k$ gilt

$$|F_n(x_i, \omega) - F(x_i)| \leq \frac{\epsilon}{2} \quad \text{und} \quad |F_n(x_i-, \omega) - F(x_i-)| \leq \frac{\epsilon}{2},$$

wobei wir etwa die Konvention $F(\infty-) = 1$ verwenden. Nun sei $x \in \mathbb{R}$ beliebig. Dann existiert ein $i \in \{1, \dots, k\}$, so dass $x \in [x_{i-1}, x_i)$. Weiterhin sei $n \geq n_0$ beliebig. Wegen der Monotonie von F und F_n folgt mit Lemma 4.1.11

$$\begin{aligned} F_n(x, \omega) - F(x) &\leq F_n(x_i-, \omega) - F(x_{i-1}) \\ &\leq |F_n(x_i-, \omega) - F(x_i-)| + F(x_i-) - F(x_{i-1}) \\ &\leq \frac{\epsilon}{2} + \left(\frac{i}{k} - \frac{i-1}{k}\right) = \frac{\epsilon}{2} + \frac{1}{k} \leq \epsilon \end{aligned}$$

und

$$\begin{aligned} F(x) - F_n(x, \omega) &\leq F(x_i-) - F_n(x_{i-1}, \omega) \\ &\leq F(x_i-) - F(x_{i-1}) + |F(x_{i-1}) - F_n(x_{i-1}, \omega)| \\ &\leq \left(\frac{i}{k} - \frac{i-1}{k}\right) + \frac{\epsilon}{2} = \frac{1}{k} + \frac{\epsilon}{2} \leq \epsilon. \end{aligned}$$

Also gilt für alle $n \geq n_0$

$$\sup_{x \in \mathbb{R}} |F_n(x, \omega) - F(x)| \leq \epsilon.$$

□

Die empirische Verteilungsfunktion war gegeben durch

$$F_n : \mathbb{R} \times \Omega \to [0,1], \quad F_n(x,\omega) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j(\omega)\leq x\}}.$$

Bezeichnet Θ die Menge aller Verteilungsfunktionen, so können wir die empirische Verteilungsfunktion als Abbildung

$$F_n : \Omega \to \Theta$$

ansehen; und zwar ist $F_n(\omega)$ für festes $\omega \in \Omega$ dann die Verteilungsfunktion

$$x \mapsto \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j(\omega)\leq x\}}.$$

Wir betrachten nun das statistische Modell

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

aus Beispiel 4.1.4. Wie in Abschnitt 1.2 betrachten wir die Folge $(\mathscr{M}_n)_{n\in\mathbb{N}}$ gegeben durch

$$\mathscr{M}_n = \big(\Omega^{\mathbb{N}}, \mathscr{F}^{\otimes\mathbb{N}}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big).$$

Es sei $\tau : \Theta \to \Theta$ die Kenngröße $\tau(F) = F$. Wir betrachten die Folge von Schätzern $(T_n)_{n\in\mathbb{N}}$ gegeben durch

$$T_n : \Omega^{\mathbb{N}} \to \Theta, \quad T_n(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{X_j\leq x\}} \quad \text{für } x \in \mathbb{R}.$$

Korollar 4.1.13. *Es gilt für $n \to \infty$*

$$\sup_{x\in\mathbb{R}} |T_n(x) - F(x)| \stackrel{f.s.}{\to} 0 \textit{ bezüglich } \mathbb{P}_F \quad \textit{für alle } F \in \Theta.$$

Beweis. Folgt aus dem Satz von Glivenko-Cantelli (Satz 4.1.12). □

Bemerkung 4.1.14. *In diesem Sinne ist die Folge von Schätzern $(T_n)_{n\in\mathbb{N}}$ stark konsistent für τ.*

4.2 Der Kolmogorov-Smirnov-Anpassungstest

Definition 4.2.1. *Es sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktion F.*

(a) Für $a \in (0,1)$ nennen wir $\mu_a := F^{-1}(a)$ das a-Quantil von μ, wobei $F^{-1} : (0,1) \to \mathbb{R}$ die Quantilfunktion aus Definition 4.1.7 bezeichnet.

(b) Für $\alpha \in (0,1)$ nennen wir $\mu_{1-\alpha}$ auch das α-Fraktil von μ.

Lemma 4.2.2. *Es sei μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktion F. Weiterhin sei $X \sim \mu$ eine Zufallsvariable. Dann gilt für jedes $\alpha \in (0,1)$*

$$\mathbb{P}(X > \mu_{1-\alpha}) \leq \alpha.$$

Beweis. Nach Lemma 4.1.11 gilt

$$\begin{aligned} \mathbb{P}(X > \mu_{1-\alpha}) &= \mathbb{P}(X > F^{-1}(1-\alpha)) \\ &= 1 - F\big(F^{-1}(1-\alpha)\big) \leq 1 - (1-\alpha) = \alpha. \end{aligned}$$

□

Definition 4.2.3. *Eine Funktion $F : \mathbb{R} \to \mathbb{R}$ heißt càdlàg, falls sie rechtsstetig ist und linksseitige Grenzwerte besitzt.*

Definition 4.2.4. *Es sei $F : \mathbb{R} \to \mathbb{R}$ eine càdlàg-Funktion.*

(a) Für jedes $x \in \mathbb{R}$ definieren wir den linksseitigen Grenzwert

$$F(x-) := \lim_{y \uparrow x} F(y).$$

(b) Für jedes $x \in \mathbb{R}$ definieren wir den Sprung

$$\Delta F(x) := F(x) - F(x-).$$

(c) Wir nennen $\{\Delta F \neq 0\}$ die Menge aller Sprungstellen von F.

(d) Wir nennen $\{\Delta F = 0\}$ die Menge aller Stetigkeitsstellen von F.

Lemma 4.2.5. *Es seien I eine Indexmenge und $(A_i)_{i \in I} \subset \mathscr{F}$ paarweise disjunkte Ereignisse mit $\mathbb{P}(A_i) > 0$ für alle $i \in I$. Dann ist I höchstens abzählbar.*

Beweis. Es genügt, zu zeigen, dass für jedes $n \in \mathbb{N}$ die Indexmenge

$$I_n := \left\{ i \in I : \mathbb{P}(A_i) \geq \frac{1}{n} \right\}$$

höchstens n Elemente hat. Andernfalls exsitiert eine Teilmenge $J_n \subset I_n$ mit $n+1$ Elementen, und wir erhalten den Widerspruch

$$1 < \frac{n+1}{n} \leq \sum_{j \in J_n} \mathbb{P}(A_j) = \mathbb{P}\bigg(\bigcup_{j \in J_n} A_j \bigg) \leq 1.$$

□

Satz 4.2.6. *Es sei $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion.*

(a) F ist càdlàg.

(b) Die Menge $\{\Delta F \neq 0\}$ ist höchstens abzählbar.

(c) Zu jedem $x \in \mathbb{R}$ existiert eine Folge $(x_k)_{k \in \mathbb{N}} \subset \{\Delta F = 0\}$ von Stetigkeitspunkten von F, so dass $x_k \downarrow x$.

Beweis.

(a) Da F monoton wachsend ist, existierten die linksseitigen Grenzwerte. Es sei μ das zu F gehörige Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktion F. Es sei $x \in \mathbb{R}$ beliebig, und es sei $(x_n)_{n \in \mathbb{N}} \subset \mathbb{R}$ eine Folge mit $x_n \downarrow x$. Wegen $(-\infty, x_n] \downarrow (-\infty, x]$ folgt

$$F(x_n) = \mu((-\infty, x_n]) \downarrow \mu((-\infty, x]) = F(x) \quad \text{für } n \to \infty.$$

(b) Es sei μ das zu F gehörige Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktion F. Dann gilt

$$\{\Delta F \neq 0\} = \{x \in \mathbb{R} : \mu(\{x\}) > 0\}.$$

Da die Mengen $(\{x\})_{x \in \mathbb{R}}$ paarweise disjunkt sind, ist $\{\Delta F \neq 0\}$ nach Lemma 4.2.5 höchstens abzählbar.

(c) Es sei $x \in \mathbb{R}$ beliebig. Es genügt, zu zeigen, dass zu jedem $\epsilon > 0$ ein Stetigkeitspunkt $y \in [x, x+\epsilon)$ existiert. Dies ist der Fall, da die Menge der Sprungstellen von F nach Teil (b) ist höchstens abzählbar ist.

□

Lemma 4.2.7.

(a) Es sei X eine Zufallsvariable mit stetiger Verteilungsfunktion F_X. Wir setzen $U := F_X(X)$. Dann gilt $U \sim \mathrm{UC}(0,1)$.

(b) Es seien $F : \mathbb{R} \to [0,1]$ eine stetige Verteilungsfunktion und $U \sim \mathrm{UC}(0,1)$ eine gleichverteilte Zufallsvariable. Dann hat die Zufallsvariable $X := F^{-1}(U)$ die Verteilungsfunktion F.

Beweis.

(a) Es sei $u \in (0,1)$ beliebig. Da F_X stetig ist mit $F_X(\mathbb{R}) \subset (0,1)$, existiert nach dem Zwischenwertsatz der Topologie ein $x \in \mathbb{R}$ mit $F_X(x) = u$. Wegen

$$\{F_X(X) < F_X(x)\} \subset \{X \leq x\}$$

gilt

$$F_U(u-) = \mathbb{P}(U < u) = \mathbb{P}(F_X(X) < F_X(x)) \leq \mathbb{P}(X \leq x) = F_X(x) = u,$$

und wegen

$$\{X \leq x\} \subset \{F_X(X) \leq F_X(x)\}$$

gilt

$$u = F_X(x) = \mathbb{P}(X \leq x) \leq \mathbb{P}(F_X(X) \leq F_X(x)) = \mathbb{P}(U \leq u) = F_U(u).$$

Insgesamt folgt

$$F_U(u-) \leq u \leq F_U(u) \quad \text{für alle } u \in (0,1).$$

Also gilt für jeden Stetigkeitspunkt $u \in (0,1)$ von F_U, dass

$$F_U(u) = u.$$

Nun sei $u \in (0,1)$ beliebig. Nach Satz 4.2.6(c) existiert eine Folge $(u_k)_{k \in \mathbb{N}} \subset (0,1)$ von Stetigkeitspunkten von F_U mit $u_k \downarrow u$. Wegen der Rechtsstetigkeit von F_U folgt

$$F_U(u) = F_U\Big(\lim_{k\to\infty} u_k\Big) = \lim_{k\to\infty} F_U(u_k) = \lim_{k\to\infty} u_k = u.$$

Insgesamt folgt

$$F_U(u) = u \mathbb{1}_{(0,1)}(u) + \mathbb{1}_{[1,\infty)}(u), \quad u \in \mathbb{R}.$$

(b) Es gilt $F_U(u) = u$ für alle $u \in [0,1]$. Nach Lemma 4.1.10 gilt also für jedes $x \in \mathbb{R}$

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(F^{-1}(U) \le x) = \mathbb{P}(U \le F(x)) = F_U(F(x)) = F(x).$$

$\square$

Bemerkung 4.2.8.

(a) Es sei X eine Zufallsvariable mit stetiger Verteilungsfunktion F, und es sei $T : \mathbb{R} \to \mathbb{R}$ eine stetige Transformation. Dann gilt für alle $t \in \mathbb{R}$

$$\mathbb{P}(T(X) = t) = \mathbb{P}(X \in T^{-1}(\{t\})).$$

Also braucht die Verteilungsfunktion von $T(X)$ nicht stetig zu sein; es sei denn T ist injektiv.

(b) Eine Verteilung mit stetiger Verteilungsfunktion F braucht nicht absolutstetig zu sein. Ein Gegenbeispiel ist die Cantor-Verteilung.

Definition 4.2.9. *Es seien $U_1, \ldots, U_n \sim \mathrm{UC}(0,1)$ unabhängige, gleichverteilte Zufallsvariablen. Es sei*

$$\widehat{F}_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{U_i \le u\}}$$

die empirische Verteilungsfunktion. Dann nennen wir die Verteilung auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ gegeben durch $\mathbb{P} \circ T_n$ mit

$$T_n = \sup_{u \in [0,1]} |\widehat{F}_n(u) - u|$$

die diskrete Kolmogorov-Verteilung mit n Freiheitsgraden, und bezeichnen diese mit K_n.

Bemerkung 4.2.10. *Nach dem Satz von Glivenko-Cantelli (Satz 4.1.12) gilt $T_n \overset{f.s.}{\to} 0$ für $n \to \infty$.*

Definition 4.2.11. *Wir nennen die Verteilung auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit der Verteilungsfunktion $F : \mathbb{R} \to [0,1]$ gegeben durch*

$$F(x) = \left(1 - 2\sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2x^2)\right) \mathbb{1}_{(0,\infty)}(x), \quad x \in \mathbb{R}$$

die Kolmogorov-Verteilung, und bezeichnen sie mit K.

Satz 4.2.12 (Satz von Kolmogorov-Smirnov). *Es gilt für $n \to \infty$*

$$\mathbb{P} \circ (\sqrt{n} T_n) \xrightarrow{w} K.$$

Beweis. Siehe [Löw, Satz 7.7]. □

Nun betrachten wir folgendes nichtparametrische statistische Modell

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big).$$

Hierbei ist $(\Omega, \mathscr{F}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Weiterhin ist Θ die Menge aller stetigen Verteilungsfunktionen $F : \mathbb{R} \to [0,1]$. Für jedes $F \in \Theta$ ist $\mathbb{P}_F$ das Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mit Verteilungsfunktion F; also

$$\mathbb{P}_F((-\infty, t]) = F(t), \quad t \in \mathbb{R}_+.$$

Wir betrachten im Folgenden das Produktmodell $\mathscr{M}^{\otimes n}$ für ein $n \in \mathbb{N}$.

Definition 4.2.13. *Wir definieren $D_n : \Theta \times \Omega^n \to \mathbb{R}$ durch*

$$D_n(F, \omega) := \sup_{x \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i(\omega) \leq x\}} - F(x) \right|,$$

oder in Kurzform

$$D_n(F) := \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|.$$

Bemerkung 4.2.14. *Nach dem Satz von Glivenko-Cantelli (Satz 4.1.12) gilt für $n \to \infty$*

$$D_n(F) \xrightarrow{f.s.} 0 \quad \text{bezüglich } \mathbb{P}_F \quad \text{für jedes } F \in \Theta.$$

Lemma 4.2.15. *Es gilt $\mathbb{P}_F \circ D_n(F) = K_n$ für alle $F \in \Theta$. Mit anderen Worten $D_n(F)$ ist eine Pivot-Statistik mit Pivot-Verteilung K_n.*

Beweis. Es sei $F \in \Theta$ beliebig. Wir setzen $U_i := F(X_i)$ für $i = 1, \dots, n$. Da $X_1, \dots, X_n$ unabhängig unter $\mathbb{P}_F$ sind, sind auch $U_1, \dots, U_n$ unabhängig unter $\mathbb{P}_F$. Nach Lemma 4.2.7(a) gilt $U_i \sim \mathrm{UC}(0,1)$ für alle $i = 1, \dots, n$. Wir setzen $\widetilde{X}_i := F^{-1}(U_i)$ für $i = 1, \dots, n$. Da $U_1, \dots, U_n$ unabhängig unter $\mathbb{P}_F$ sind, sind auch $\widetilde{X}_1, \dots, \widetilde{X}_n$ unabhängig unter $\mathbb{P}_F$. Nach Lemma 4.2.7(b) gilt $\widetilde{X}_i \sim F$ für alle $i = 1, \dots, n$. Nun sei $x \in \mathbb{R}$ beliebig. Mit Lemma 4.1.10 folgt

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}} \stackrel{d}{=} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\widetilde{X}_i \leq x\}} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{F^{-1}(U_i) \leq x\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{U_i \leq F(x)\}} = \widehat{F}_n(F(x)). \end{aligned}$$

Hierbei bedeutet $Y \stackrel{d}{=} Z$ für zwei Zufallsvariablen Y und Z, dass $\mathbb{P}_F \circ Y = \mathbb{P}_F \circ Z$. Es folgt

$$\begin{aligned} D_n(F) &= \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \stackrel{d}{=} \sup_{x\in\mathbb{R}} |\widehat{F}_n(F(x)) - F(x)| \\ &= \sup_{u\in[0,1]} |\widehat{F}_n(u) - u| = T_n. \end{aligned}$$

□

Im Folgenden bezeichnen wir mit $X_{(1)} \le \ldots \le X_{(n)}$ die Ordnungsstatistiken der Stichprobe $(X_1, \ldots, X_n)$.

Lemma 4.2.16. *Für jedes $F \in \Theta$ gilt*

$$D_n(F) = \max_{i=1,\ldots,n} \max\left\{\left|F(X_{(i)}) - \frac{i-1}{n}\right|, \left|\frac{i}{n} - F(X_{(i)})\right|\right\}.$$

Beweis. Die empirische Verteilungsfunktion ist gegeben durch

$$F_n(x) = \sum_{i=1}^{n-1} \frac{i}{n} \mathbb{1}_{[X_{(i)}, X_{(i+1)})} + \mathbb{1}_{[X_{(n)},\infty)}, \quad x \in \mathbb{R}.$$

Da F stetig und monoton wachsend ist, folgt

$$\begin{aligned} D_n(F) &= \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = \max_{i=1,\ldots,n} \left\{|F(X_{(i)}) - F_n(X_{(i)}-)|, |F_n(X_{(i)}) - F(X_{(i)})|\right\} \\ &= \max_{i=1,\ldots,n} \max\left\{\left|F(X_{(i)}) - \frac{i-1}{n}\right|, \left|\frac{i}{n} - F(X_{(i)})\right|\right\}. \end{aligned}$$

□

Korollar 4.2.17. *Für jedes $F \in \Theta$ gilt*

$$D_n(F) \stackrel{d}{=} \max_{i=1,\ldots,n} \max\left\{\left|U_{(i)} - \frac{i-1}{n}\right|, \left|\frac{i}{n} - U_{(i)}\right|\right\},$$

wobei $U_{(1)} \le \ldots \le U_{(n)}$ die Ordnungsstatistiken von unabhängigen, gleichverteilten Zufallsvariablen $U_1, \ldots, U_n \sim \mathrm{UC}(0,1)$ sind.

Wir fixieren nun eine stetige Verteilungsfunktion F_0 und setzen $D_n = D_n(F_0)$.

Satz 4.2.18 (Kolmogorov-Smirnov-Anpassungstest)**.** *Wir zerlegen Θ in*

$$\Theta_0 = \{F_0\} \quad \textit{und} \quad \Theta_1 = \Theta \setminus \{F_0\}.$$

Wir betrachten also die Null-Hypothese $H_0 = \{F = F_0\}$ und die Alternative $H_1 = \{F \ne F_0\}$. Es sei $\alpha \in (0,1)$ beliebig. Dann ist

$$\varphi := \mathbb{1}_{\{D_n > K_{n,1-\alpha}\}}$$

ein zulässiger Test zum Irrtumsniveau α.

Beweis. Nach Lemma 4.2.2 gilt

$$G_\varphi(F_0) = \mathbb{E}_{F_0}[\varphi] = \mathbb{E}_{F_0}[\mathbb{1}_{\{D_n > K_{n,1-\alpha}\}}] = \mathbb{P}_{F_0}(D_n > K_{n,1-\alpha}) \leq \alpha.$$

□

Bemerkung 4.2.19. *Alternativ können wir den Test*

$$\psi := \mathbb{1}_{\{\sqrt{n}D_n > K_{1-\alpha}\}}$$

betrachten. Dies ist die asymptotische Variante des Kolmogorov-Smirnov-Anpassungstests. Nach dem Satz von Kolmogorov-Smirnov (Satz 4.2.12) gilt für große $n \in \mathbb{N}$

$$G_\psi(F_0) = \mathbb{E}_{F_0}[\psi] = \mathbb{E}_{F_0}[\mathbb{1}_{\{\sqrt{n}D_n > K_{1-\alpha}\}}] = \mathbb{P}_{F_0}(\sqrt{n}D_n > K_{1-\alpha}) \approx \mathbb{F}_{F_0}(D > K_{1-\alpha}) \leq \alpha,$$

wobei $\mathbb{P}_{F_0} \circ D = K$. In diesem Sinne ist φ asymptotisch zulässig zum Irrtumsniveau α.

Bemerkung 4.2.20. *Es seien $n \in \mathbb{N}$ und eine konkrete Stichprobe $(X_1, \ldots, X_n)$ gegeben. Weiterhin sei eine vermutete Verteilung F_0 gegeben; beispielsweise $F_0 = \mathrm{N}(\mu, \sigma^2)$ mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$, etwa*

$$\mu = \bar{X} \quad \text{und} \quad \sigma^2 = s^2(X).$$

Gemäß Lemma 4.2.16 berechnen wir D_n als

$$D_n = \max_{i=1,\ldots,n} \max \left\{ \left| F_0(X_{(i)}) - \frac{i-1}{n} \right|, \left| \frac{i}{n} - F_0(X_{(i)}) \right| \right\}.$$

Dann stehen uns zwei Möglichkeiten zur Verfügung:

- *Wir verwerfen H_0, falls $D_n > K_{n,1-\alpha}$; und akzeptieren H_0 andernfalls. Ist n nicht zu groß, so sind die Fraktile $K_{n,1-\alpha}$ bekannt. Bei großen n ist es sehr schwierig, die Fraktile zu ermitteln.*
- *Wir verwerfen H_0, falls $\sqrt{n}D_n > K_{1-\alpha}$; und akzeptieren H_0 andernfalls. Dieses Verfahren bietet sich für große n an.*

Beispiel 4.2.21. *Wir betrachten den Kurs eines Wertpapiers $(S_0, S_1, \ldots, S_n)$ für ein $n \in \mathbb{N}$, und definieren $(X_1, \ldots, X_n)$ durch $X_i := \ln(S_i/S_{i-1})$ für $i = 1, \ldots, n$. Dann ist die Nullhypothese $H_0 = \{F = \mathrm{N}(\mu, \sigma^2)\}$ typischerweise zu verwerfen.*

4.3 Mehrdimensionale Normalverteilungen

Weitere Details zu diesem Thema können beispielsweise in [JP04, Chapter 16] oder [Tap13, Kapitel 12] nachgelesen werden.

Definition 4.3.1. *Für jedes $\mu \in \mathbb{R}$ setzen wir $\mathrm{N}(\mu, 0) := \delta_\mu$, und sprechen von einer entarteten Normalverteilung.*

Definition 4.3.2. *Ein $\mathbb{R}^d$-wertiger Zufallsvektor X heißt ein Gauß'scher Zufallsvektor, falls für jedes lineare Funktional $\ell : \mathbb{R}^d \to \mathbb{R}$ die Zufallsvariable $\ell(X)$ (möglicherweise entartet) normalverteilt ist.*

Bemerkung 4.3.3. *Jedes lineare Funktional $\ell : \mathbb{R}^d \to \mathbb{R}$ ist von der Form $\ell(x) = \langle a, x \rangle_{\mathbb{R}^d}$ für ein $a \in \mathbb{R}^d$.*

Satz 4.3.4. *Es seien X und Y Gauß'sche Zufallsvektoren mit $\mathbb{E}[X] = \mathbb{E}[Y]$ und $\mathrm{Cov}(X) = \mathrm{Cov}(Y)$. Dann gilt $\mathbb{P} \circ X = \mathbb{P} \circ Y$.*

Definition 4.3.5. *Es sei X ein Gauß'scher Zufallsvektor. Wir setzen $\mu := \mathbb{E}[X] \in \mathbb{R}^d$ und $\Sigma^2 := \mathrm{Cov}(X) \in \mathbb{R}^{d \times d}$, und nennen die Verteilung $\mathrm{N}(\mu, \Sigma^2) := \mathbb{P} \circ X$ auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ die mehrdimensionale Normalverteilung mit Parametern μ und Σ^2.*

Bemerkung 4.3.6. *Die Kovarianzmatrix Σ^2 ist stets symmetrisch und positiv semidefinit.*

Satz 4.3.7. *Es seien $\mu \in \mathbb{R}^d$ und $\Sigma^2 \in \mathbb{R}^{d \times d}$ eine symmetrische, positiv semidefinite Matrix. Dann existiert die mehrdimensionale Normalverteilung $\mathrm{N}(\mu, \Sigma^2)$ auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

Satz 4.3.8. *Es sei $X \sim \mathrm{N}(\mu, \Sigma^2)$ ein $\mathbb{R}^d$-wertiger Gauß'scher Zufallsvektor. Dann sind folgende Aussagen äquivalent:*

(i) X ist absolutstetig.

(ii) Σ^2 ist positiv definit.

Satz 4.3.9. *Für einen $\mathbb{R}^d$-wertigen Gauß'schen Zufallsvektor $X \sim \mathrm{N}(\mu, \Sigma^2)$ sind folgende Aussagen äquivalent:*

(i) $X_1, \ldots, X_n$ sind unabhängig.

(ii) Σ^2 ist eine Diagonalmatrix.

Satz 4.3.10. *Es sei $X \sim \mathrm{N}(\mu, \Sigma^2)$ ein $\mathbb{R}^d$-wertiger Gauß'scher Zufallsvektor. Weiterhin seien $A \in \mathbb{R}^{d \times e}$ eine Matrix und $b \in \mathbb{R}^e$ ein Vektor. Dann ist $Y := AX + b$ ein $\mathbb{R}^e$-wertiger Gauß'scher Zufallsvektor mit*

$$Y \sim \mathrm{N}(A\mu + b, A\Sigma^2 A^\top).$$

Korollar 4.3.11. *Es seien $X_1, \ldots, X_d \sim \mathrm{N}(0, \sigma^2)$ unabhängige, normalverteilte Zufallsvariablen. Weiterhin sei $A \in \mathbb{R}^{d \times d}$ eine orthogonale Matrix. Wir setzen $Y := AX$. Dann sind $Y_1, \ldots, Y_d$ unabhängig mit $Y_i \sim \mathrm{N}(0, \sigma^2)$ für $i = 1, \ldots, d$.*

Satz 4.3.12 (Mehrdimensionaler zentraler Grenzwertsatz)**.** *Es sei $(X_j)_{j \in \mathbb{N}} \subset \mathscr{L}^2$ eine Folge unabhängiger, identisch verteilter $\mathbb{R}^d$-wertiger Zufallsvektoren mit Erwartungswert $\mu \in \mathbb{R}^d$ und Kovarianzmatrix $\Sigma^2 \in \mathbb{R}^{d \times d}$. Wir setzen*

$$S_n := \sum_{j=1}^{n} X_j \quad \text{und} \quad Y_n := \frac{S_n - n\mu}{\sqrt{n}} \quad \text{für } n \in \mathbb{N}.$$

Dann gilt $\mathbb{P} \circ Y_n \xrightarrow{w} \mathrm{N}(0, \Sigma^2)$.

4.4 Multinomialverteilungen

Es sei $E = \{1, \ldots, d\}$ für ein $d \in \mathbb{N}$.

Definition 4.4.1. *Es sei $\pi : E \to [0,1]$ ein stochastischer Vektor. Die Verteilung auf $(\mathbb{N}_0^d, \mathfrak{P}(\mathbb{N}_0^d))$ mit stochastischem Vektor $p : \mathbb{N}_0^d \to [0,1]$ gegeben durch*

$$p(e_k) = \pi_k, \quad k = 1, \ldots, d$$

nennen wir die verallgemeinerte Bernoulli-Verteilung mit Parameter π, und bezeichnen sie mit $\mathrm{VBer}(\pi)$.

Bemerkung 4.4.2. *Für alle $p \in [0,1]$ gilt*

$$\mathrm{Ber}(p)(1) = \mathrm{VBer}((p, 1-p))(1,0) \quad \text{und} \quad \mathrm{Ber}(p)(0) = \mathrm{VBer}((p, 1-p))(0,1).$$

Definition 4.4.3.

(a) Für $k \in \mathbb{N}_0^d$ setzen wir

$$|k| := k_1 + \ldots + k_d.$$

(b) Für $n \in \mathbb{N}$ und $k \in \mathbb{N}_0^d$ mit $|k| = n$ setzen wir

$$\binom{n}{k} := \frac{n!}{k_1! \cdot \ldots \cdot k_d!}.$$

(c) Für $\pi \in \mathbb{R}^d$ und $k \in \mathbb{N}_0^d$ setzen wir

$$\pi^k := \pi_1^{k_1} \cdot \ldots \cdot \pi_d^{k_d}.$$

Definition 4.4.4. *Es seien $n \in \mathbb{N}$ und $\pi : E \to [0,1]$ ein stochastischer Vektor. Dann nennen wir die Verteilung auf $(\mathbb{N}_0^d, \mathfrak{P}(\mathbb{N}_0^d))$ mit stochastischem Vektor $p : \mathbb{N}_0^d \to [0,1]$ gegeben durch*

$$p(k) = \binom{n}{k} \pi^k \mathbb{1}_{\{|k|=n\}}$$

die Multinomialverteilung mit Parametern n und π, und bezeichnen sie mit $\mathrm{Mult}(n, \pi)$.

Bemerkung 4.4.5. *Für alle $p \in [0,1]$ gilt*

$$\mathrm{Bi}(n,p)(k) = \mathrm{Mult}(n,(p,1-p))(k, n-k), \quad k = 0, \dots, n.$$

Bemerkung 4.4.6. *Für jeden stochastischer Vektor $\pi : E \to [0,1]$ gilt*

$$\mathrm{VBer}(\pi) = \mathrm{Mult}(1, \pi).$$

Satz 4.4.7. *Es seien $X_1, \dots, X_n \sim \mathrm{VBer}(\pi)$ unabhängige, identisch verteilte Zufallsvariablen. Dann gilt*

$$X_1 + \ldots + X_n \sim \mathrm{Mult}(n, \pi).$$

Satz 4.4.8 (Verallgemeinerter Satz von Moivre-Laplace)**.** *Es sei $(X_j)_{j\in\mathbb{N}}$ eine Folge unabhängiger, identisch verteilter $\mathbb{R}^d$-wertiger Zufallsvariablen mit $X_1 \sim \mathrm{VBer}(\pi)$. Wir setzen*

$$S_n := \sum_{j=1}^n X_j \quad \text{und} \quad Y_n := \frac{S_n - n\pi}{\sqrt{n}} \quad \text{für } n \in \mathbb{N}.$$

Dann gilt $\mathbb{P} \circ Y_n \xrightarrow{w} \mathrm{N}(0, \Sigma^2)$, wobei die Kovarianzmatrix $\Sigma^2 \in \mathbb{R}^{d\times d}$ gegeben ist durch $\Sigma^2 = \mathrm{diag}(\pi) - \pi\pi^\top$; das heißt

$$\Sigma^2_{ij} = \begin{cases} \pi_i(1-\pi_i), & \text{falls } i = j, \\ -\pi_i\pi_j, & \text{falls } i \neq j. \end{cases}$$

Beweis. Wir setzen $X := X_1$. Dann gilt $X^i \sim \mathrm{Ber}(\pi_i)$ für alle $i = 1, \dots, d$. Folglich gilt für alle $i = 1, \dots, d$

$$\mathbb{E}[X^i] = \pi_i \quad \text{und} \quad \mathrm{Var}[X^i] = \pi_i(1-\pi_i).$$

Nun seien $i, j \in \{1, \dots, d\}$ mit $i \neq j$ beliebig. Wegen $|X| = 1$ gilt $X^iX^j = 0$, und es folgt

$$\mathrm{Cov}(X^i, X^j) = \mathbb{E}[X^iX^j] - \mathbb{E}[X^i]\mathbb{E}[X^j] = -\pi_i\pi_j.$$

Nun folgt die Aussage mit dem mehrdimensionalen zentralen Grenzwertsatz (Satz 4.3.12). □

Nun sei $\pi : E \to (0,1)$ eine stochastischer Vektor. Weiterhin seien $(X_k)_{k\in\mathbb{N}}$ unanhängige, identisch verteilte E-wertige Zufallsvariablen mit $X_1 \sim \pi$. Wir definieren die Folge $(H_n)_{n\in\mathbb{N}}$ wie folgt. Für $n \in \mathbb{N}$ sei $H_n = (H_n^1, \dots, H_n^d) : \Omega \to \mathbb{N}_0^d$ der Zufallsvektor mit den Häufigkeiten

$$H_n^i = \sum_{k=1}^{n} \mathbb{1}_{\{X_k=i\}}, \quad i \in E.$$

Lemma 4.4.9. *Es gilt $H_n \sim \mathrm{Mult}(n,\pi)$ für jedes $n \in \mathbb{N}$.*

Satz 4.4.10 (Satz von Pearson)**.** *Wir definieren die Folge $(T_n)_{n\in\mathbb{N}}$ durch*

$$T_n := \sum_{i=1}^{d} \frac{1}{n\pi_i}(H_n^i - n\pi_i)^2.$$

Dann gilt für $n \to \infty$

$$\mathbb{P} \circ T_n \xrightarrow{w} \chi^2_{d-1}.$$

Beweis. Wir definieren die Folge $(Y_n)_{n\in\mathbb{N}}$ von $\mathbb{R}^d$-wertigen Zufallsvariablen durch

$$Y_n := \frac{H_n - n\pi}{\sqrt{n}} \quad \text{für } n \in \mathbb{N}.$$

Nach Lemma 4.4.9 und Satz 4.4.7 gilt $H_n \sim \mathrm{Mult}(n,\pi) = \mathrm{VBer}(\pi)^{\otimes n}$. Also gilt nach dem Verallgemeinerten Satz von Moivre-Laplace (Satz 4.4.8) $\mathbb{P}\circ Y_n \xrightarrow{w} \mathrm{N}(0,\Sigma^2)$, wobei $\Sigma^2 = \mathrm{diag}(\pi) - \pi\pi^\top$. Nun definieren wir die stetige Transformation

$$h : \mathbb{R}^d \to \mathbb{R}, \quad h(x) := \sum_{i=1}^{d} \frac{x_i^2}{\pi_i}.$$

Dann gilt $T_n = h(Y_n)$ für alle $n \in \mathbb{N}$, und folglich $T_n \xrightarrow{\mathscr{D}} h(Z)$ mit $Z \sim \mathrm{N}(0,\Sigma^2)$. Wir setzen

$$D := \mathrm{diag}(\pi^{-1/2}) := \mathrm{diag}(\pi_1^{-1/2}, \dots, \pi_d^{-1/2}) \in \mathbb{R}^{d\times d}$$

und $\widetilde{Z} := DZ$. Nach Satz 4.3.10 ist $\widetilde{Z}$ ein Gauß'scher Zufallsvektor mit

$$\begin{aligned}\mathrm{Cov}(\widetilde{Z}) &= D\Sigma^2 D^\top = \mathrm{diag}(\pi^{-1/2})(\mathrm{diag}(\pi) - \pi\pi^\top)\mathrm{diag}(\pi^{-1/2}) \\ &= \mathrm{Id} - (\pi^{1/2})^\top \pi^{1/2}.\end{aligned}$$

Also gilt

$$\widetilde{Z} \sim \mathrm{N}(0, \mathrm{Id} - (\pi^{1/2})^\top \pi^{1/2}).$$

Für den Vektor $v := (\pi^{1/2})^\top$ gilt $\|v\|_{\mathbb{R}^d} = 1$, wobei

$$\|v\|_{\mathbb{R}^d} = \langle v, v\rangle_{\mathbb{R}^d} = \sqrt{v_1 + \ldots + v_d}$$

die euklidische Norm bezeichnet. Also existiert eine orthogonale Matrix $A \in \mathbb{R}^{d\times d}$ mit $A_{i1} = v_i$ für alle $i = 1, \ldots, d$. Wir setzen $\bar{Z} := A^\top \widetilde{Z}$. Nach Satz 4.3.10 ist $\widetilde{Z}$ ein Gauß'scher Zufallsvektor. Es gilt $v^\top A = e_1$, und folglich

$$\begin{aligned}
\operatorname{Cov}(\bar{Z}) &= A^\top \operatorname{Cov}(\widetilde{Z}) A = A^\top (\mathrm{Id} - (\pi^{1/2})^\top \pi^{1/2}) A \\
&= A^\top A - A^\top v v^\top A = \mathrm{Id} - e_1^\top e_1 \\
&= \mathrm{Id} - \operatorname{diag}(e_1) = \operatorname{diag}(0, 1, \ldots, 1).
\end{aligned}$$

Also gilt

$$\bar{Z} \sim \mathrm{N}(0, \operatorname{diag}(0, 1, \ldots, 1)).$$

Nach Satz 4.3.9 sind $\bar{Z}_2, \ldots, \bar{Z}_d \sim \mathrm{N}(0,1)$ unabhängig. Da $AA^\top = \mathrm{Id}$, folgt mit Lemma 1.4.16

$$\begin{aligned}
h(Z) &= \sum_{i=1}^d \frac{Z_i^2}{\pi_i} = Z^\top \operatorname{diag}(\pi^{-1}) Z = Z^\top D^2 Z = (DZ)^\top DZ \\
&= \widetilde{Z}^\top \widetilde{Z} = \widetilde{Z}^\top A A^\top \widetilde{Z} = (A^\top \widetilde{Z})^\top A^\top \widetilde{Z} = \bar{Z}^\top \bar{Z} = \sum_{j=2}^d \bar{Z}_j^2 \sim \chi^2_{d-1}.
\end{aligned}$$

□

4.5 Der χ^2-Anpassungstest

Beispiel 4.5.1. *Bei insgesamt 189 Würfen mit einem Würfel zweifelhafter Qualität erschienen die Augenzahlen 1 bis 6 jeweils 30, 37, 26, 29, 29 und 38 mal. Geben diese Daten Anlass, die Hypothese, dass der Würfel fair ist, zu verwerfen?*

Es sei $E = \{1, \ldots, d\}$. Wir betrachten das statistische Modell

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $(\Omega, \mathscr{F}) = (E, \mathfrak{P}(E))$, und Θ bestehend aus der Menge aller stochastischen Vektoren $\pi : E \to (0,1)$. Für jedes $\pi \in \Theta$ ist $\mathbb{P}_\pi$ das Wahrscheinlichkeitsmaß mit stochastischem Vektor π. Wir gehen zum Produktmodell $\mathscr{M}^{\otimes n}$ für ein $n \in \mathbb{N}$ über, und fixieren einen stochastischen Vektor $\pi_0 : E \to (0,1)$. Wir zerlegen Θ in

$$\Theta_0 = \{\pi_0\} \quad \text{und} \quad \Theta_1 = \Theta \setminus \{\pi_0\}.$$

Wir betrachten also die Null-Hypothese $H_0 = \{\pi = \pi_0\}$ und die Alternative $H_1 = \{\pi \neq \pi_0\}$. Es sei $H_n = (H_n^1, \dots, H_n^d) : \Omega^n \to \mathbb{N}_0^d$ der Zufallsvektor mit den Häufigkeiten

$$H_n^i = \sum_{k=1}^n \mathbb{1}_{\{X_k = i\}}, \quad i \in E.$$

Satz 4.5.2 (Multinomial-Anpassungstest)**.** *Es sei $\alpha \in (0,1)$ beliebig. Weiterhin sei $B \subset \mathbb{N}_0^d$, so dass*

$$\mathrm{Mult}(n, \pi_0)(B) \geq 1 - \alpha.$$

Dann ist

$$\varphi := \mathbb{1}_{\{H_n \notin B\}}$$

ein zulässiger Test zum Irrtumsniveau α.

Beweis. Nach Lemma 4.4.9 gilt $\mathbb{P}_{\pi_0} \circ H_n = \mathrm{Mult}(n, \pi_0)$, und daher

$$G_\varphi(\pi_0) = \mathbb{E}_{\pi_0}[\varphi] = \mathbb{E}_{\pi_0}[\mathbb{1}_{\{H_n \notin B\}}] = \mathbb{P}_{\pi_0}(H_n \notin B) = \mathrm{Mult}(n, \pi_0)(B^c) \leq \alpha.$$

□

Nun betrachten wir die Statistik

$$T_n := \sum_{i=1}^d \frac{1}{n\pi_0^i}(H_n^i - n\pi_0^i)^2.$$

Definition 4.5.3. *Für $\alpha \in (0,1)$ nennen wir*

$$\varphi = \mathbb{1}_{\{T_n > \chi^2_{d-1,1-\alpha}\}}$$

den χ^2-Anpassungstest zum Irrtumsniveau α.

Bemerkung 4.5.4. *Der χ^2-Anpassungstest ist ein asymptotischer Test. Für große $n \in \mathbb{N}$ gilt nach dem Satz von Pearson (Satz 4.4.10) und Lemma 4.2.2*

$$\begin{aligned} G_\varphi(\pi_0) = \mathbb{E}_{\pi_0}[\varphi] = \mathbb{E}_{\pi_0}[\mathbb{1}_{\{T_n > \chi^2_{d-1,1-\alpha}\}}] &= \mathbb{P}_{\pi_0}(T_n > \chi^2_{d-1,1-\alpha}) \\ &\approx \mathbb{P}_{\pi_0}(T > \chi^2_{d-1,1-\alpha}) \leq \alpha, \end{aligned}$$

wobei $\mathbb{P}_{\pi_0} \circ T = \chi^2_{d-1}$. In diesem Sinne ist φ asymptotisch zulässig zum Irrtumsniveau α.

Beispiel 4.5.5. *In Beispiel 4.5.1 sind $d = 6$ und $n = 189$. Weiterhin ist $\pi_0 = \mathrm{UD}(E)$ mit $E = \{1, \dots, 6\}$. Wir wählen das Irrtumsniveau $\alpha = 0.05$. Es gilt*

$$T_{189} = 3.730 < 11.1 = \chi^2_{5,0.95},$$

so dass wir die Null-Hypothese H_0 annehmen. Es besteht also kein Anlass, die Hypothese, dass der Würfel fair ist, zu verwerfen.

4.6 Der χ^2-Test auf Unabhängigkeit

Es seien $E = \{1, \ldots, d\}$ und $F = \{1, \ldots, e\}$ für $d, e \in \mathbb{N}$. Es sei $(X_k)_{k\in\mathbb{N}}$ eine Folge unabhängiger, identisch verteilter $E \times F$-wertiger Zufallsvariablen. Für $n \in \mathbb{N}$ seien $H^n : \Omega \to \mathbb{N}_0^{d\times e}$, $K^n : \Omega \to \mathbb{N}_0^d$ und $L^n : \Omega \to \mathbb{N}_0^e$ gegeben durch

$$\begin{aligned} H_n^{ij} &= \sum_{k=1}^n \mathbb{1}_{\{X_k=(i,j)\}}, \\ K_n^i &= \sum_{k=1}^n \mathbb{1}_{\{Y_k=i\}}, \\ L_n^j &= \sum_{k=1}^n \mathbb{1}_{\{Z_k=j\}}, \end{aligned}$$

wobei $X_k = (Y_k, Z_k)$. Weiterhin setzen wir

$$D_n = \sum_{i=1}^d \sum_{j=1}^e \frac{(H_n^{ij} - \frac{K_n^i L_n^j}{n})^2}{\frac{K_n^i L_n^j}{n}}.$$

Satz 4.6.1 (Verallgemeinerter Satz von Pearson)**.** *Falls Y_1 und Z_1 unabhängig sind, dann gilt für $n \to \infty$*

$$\mathbb{P} \circ D_n \xrightarrow{w} \chi^2_{(d-1)(e-1)}.$$

Beweis. Siehe [Geo04, Satz 11.18]. □

Nun betrachten wir das statistische Modell

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $(\Omega, \mathscr{F}) = (E \times F, \mathfrak{P}(E \times F))$ und Θ der Menge aller stochastischen Vektoren auf $E \times F$. Für jedes $\pi \in \Theta$ ist $\mathbb{P}_\pi$ das Wahrscheinlichkeitsmaß mit stochastischem Vektor π.

Definition 4.6.2. *Es sei $\pi \in \Theta$ beliebig.*

(a) $\pi_1 : E \to [0,1]$ ist der stochastische Vektor gegeben durch

$$\pi_1^i = \sum_{j=1}^e \pi^{ij}, \quad i \in E.$$

(b) $\pi_2 : F \to [0,1]$ *ist der stochastische Vektor gegeben durch*

$$\pi_2^j = \sum_{i=1}^{d} \pi^{ij}, \quad j \in F.$$

Definition 4.6.3. *Für zwei stochastische Vektoren* $\pi_1 : E \to [0,1]$ *und* $\pi_2 : F \to [0,1]$ *ist* $\pi_1 \otimes \pi_2 \in \Theta$ *der stochastische Vektor gegeben durch*

$$(\pi_1 \otimes \pi_2)^{ij} = \pi_1^i \pi_2^j, \quad (i,j) \in E \times F.$$

Wir setzen

$$\Theta_0 = \{\pi \in \Theta : \pi = \pi_1 \otimes \pi_2\} \quad \text{und} \quad \Theta_1 = \Theta \setminus \Theta_0.$$

Wir betrachten also die Null-Hypothese $H_0 = \{\pi = \pi_1 \otimes \pi_2\}$ und die Alternative $H_1 = \{\pi \neq \pi_1 \otimes \pi_2\}$. Für jedes $\pi \in \Theta$ sind folgende Aussagen äquivalent:

(i) Die Stichprobe $X = ((Y_1, Z_1), \ldots, (Y_n, Z_n))$ enthält unter $\mathbb{P}_\pi$ unabhängige Komponenten Y_i und Z_i.

(ii) Es gilt $\pi \in \Theta_0$.

Definition 4.6.4. *Wir nennen*

$$\varphi = \mathbb{1}_{\{D_n > \chi^2_{(d-1)(e-1),1-\alpha}\}}$$

den χ^2*-Test auf Unabhängigkeit zum Irrtumsniveau* α*.*

Bemerkung 4.6.5. *Der* χ^2*-Test auf Unabhängigkeit ist ein asymptotischer Test. Für große* $n \in \mathbb{N}$ *gilt nach dem verallgemeinerten Satz von Pearson (Satz 4.6.1) und Lemma 4.2.2 für alle* $\pi \in \Theta_0$

$$\begin{aligned} G_\varphi(\pi) &= \mathbb{E}_\pi[\varphi] = \mathbb{E}_\pi[\mathbb{1}_{\{D_n > \chi^2_{(d-1)(e-1),1-\alpha}\}}] = \mathbb{P}_\pi(D_n > \chi^2_{(d-1)(e-1),1-\alpha}) \\ &\approx \mathbb{P}_\pi(D > \chi^2_{(d-1)(e-1),1-\alpha}) \leq \alpha, \end{aligned}$$

wobei $\mathbb{P}_\pi \circ D = \chi^2_{(d-1)(e-1)}$*. In diesem Sinne ist* φ *asymptotisch zulässig zum Irrtumsniveau* α*.*

Kapitel 5

Lineare Modelle

5.1 Einfache lineare Regression

Es sei

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

das statistische Modell gegeben durch:

- $(\Omega, \mathscr{F}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ für ein $n \in \mathbb{N}$.
- $\Theta = \Theta_0 \times (0, \infty)$ mit $\Theta_0 = \mathbb{R}^2$ und

  $$\mathbb{P}_\vartheta = \bigotimes_{i=1}^n \mathrm{N}(\beta_0 + \beta_1 y_i, \sigma^2), \quad \vartheta = (\beta, \sigma^2) \in \Theta$$

  mit einem Vektor $y \in \mathbb{R}^n$, so dass $y \notin \mathrm{lin}\{\mathbb{1}\}$, wobei $\mathbb{1} = (1, \dots, 1)$. Wir können die Einträge von y als Kontrollvariablen interpretieren.

Wir vermuten also, dass zwischen der Stichprobe X und den Kontrollvariablen y ein linearer Zusammenhang besteht. Genauer

$$X_i = \beta_0 + \beta_1 y_i + \sigma \epsilon_i, \quad i = 1, \dots, n$$

wobei $\epsilon_1, \dots, \epsilon_n \sim \mathrm{N}(0,1)$ unabhängige, standardnormalverteilte Zufallsvariablen sind. Wir können dies kompakter schreiben als

$$X = \beta_0 \mathbb{1} + \beta_1 y + \sigma \epsilon.$$

Im Folgenden betrachten wir auf $\mathbb{R}^n$ das euklische Skalarprodukt

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i$$

und die euklische Norm

$$\|x\| = \sqrt{\langle x, x\rangle} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

Definition 5.1.1. *Ein Schätzer $\hat{\beta} : \Omega \to \Theta_0$ heißt ein* <u>*Kleinste-Quadrate-Schätzer (KQS)*</u> *für β, falls*

$$\|X - (\hat{\beta}_0 \mathbb{1} + \hat{\beta}_1 y)\| = \min_{\beta \in \Theta_0} \|X - (\beta_0 \mathbb{1} + \beta_1 y)\|.$$

Satz 5.1.2. *Es existiert ein eindeutig bestimmter KQS $\hat{\beta}$ für β; dieser ist gegeben durch*

$$\begin{aligned} \hat{\beta}_1 &= \frac{\langle y - \bar{y}\mathbb{1}, X - \bar{X}\mathbb{1}\rangle}{\|y - \bar{y}\mathbb{1}\|^2} = \frac{\sum_{i=1}^n (y_i - \bar{y})(X_i - \bar{X})}{\sum_{i=1}^n (y_i - \bar{y})^2}, \\ \hat{\beta}_0 &= \bar{X} - \hat{\beta}_1 \bar{y}. \end{aligned}$$

Beweis. Für eine beliebige Stichprobe $X \in \mathbb{R}^n$ definieren wir $f : \Theta_0 = \mathbb{R}^2 \to \mathbb{R}_+$ durch

$$f(\beta) = \|X - (\beta_0 \mathbb{1} + \beta y)\|^2.$$

Dann gilt

$$\nabla f(\beta) = -2\big(\langle X - (\beta_0 \mathbb{1} + \beta_1 y), \mathbb{1}\rangle, \langle X - (\beta_0 \mathbb{1} + \beta_1 y), y\rangle\big), \quad \beta = (\beta_0, \beta_1) \in \mathbb{R}^2.$$

Es sei $\beta \in \mathbb{R}^2$ mit $\nabla f(\beta) = 0$. Dann gilt

$$\begin{aligned} &\langle X - \beta_0 \mathbb{1} - \beta_1 y, \mathbb{1}\rangle = 0 \\ \Leftrightarrow\ &\langle X, \mathbb{1}\rangle - n\beta_0 - \beta_1 \langle y, \mathbb{1}\rangle = 0 \\ \Leftrightarrow\ &n\bar{X} - n\beta_0 - n\beta_1 \bar{y} = 0 \\ \Leftrightarrow\ &\beta_0 = \bar{X} - \beta_1 \bar{y}. \end{aligned}$$

Weiterhin gilt

$$\begin{aligned} &\langle X - \beta_0 \mathbb{1} - \beta_1 y, y\rangle = 0 \\ \Leftrightarrow\ &\langle X, y\rangle - \beta_0 \langle \mathbb{1}, y\rangle - \beta_1 \langle y, y\rangle = 0 \\ \Leftrightarrow\ &\langle X, y\rangle - n\beta_0 \bar{y} - \beta_1 \|y\|^2 = 0. \end{aligned}$$

Durch Einsetzen erhalten wir

$$\begin{aligned} &\langle X, y\rangle - n(\bar{X} - \beta_1 \bar{y})\bar{y} - \beta_1 \|y\|^2 = 0 \\ \Leftrightarrow\ &\langle X, y\rangle - n\bar{X}\bar{y} + n\beta_1 \bar{y}^2 - \beta_1 \|y\|^2 = 0 \\ \Leftrightarrow\ &\beta_1 \big(\|y\|^2 - \|\bar{y}\mathbb{1}\|^2\big) = \langle X, y\rangle - \langle \bar{X}\mathbb{1}, \bar{y}\mathbb{1}\rangle. \end{aligned}$$

Nun gilt nach Lemma 1.1.8

$$\|x - \bar{x}\mathbb{1}\|^2 = \|x\|^2 - \|\bar{x}\mathbb{1}\|^2.$$

Durch Polarisierung folgt allgemeiner

$$\langle x - \bar{x}\mathbb{1}, y - \bar{y}\mathbb{1}\rangle = \langle x, y\rangle - \langle \bar{x}\mathbb{1}, \bar{y}\mathbb{1}\rangle.$$

Damit erhalten wir die gewünschte Formel. □

Bemerkung 5.1.3. *Es gilt*

$$s^2(X) = \frac{1}{n-1}\|X - \bar{X}\mathbb{1}\|^2.$$

Wir benutzen auch die Notation $s_X^2 = s^2(X)$. *Allgemeiner können wir die korrigierte Stichprobenkovarianz*

$$s_{XY} = \frac{1}{n-1}\langle X - \bar{X}\mathbb{1}, Y - \bar{Y}\mathbb{1}\rangle$$

einführen. Dann erhalten wir die Darstellung

$$\hat{\beta}_1 = \frac{s_{yX}}{s_y^2}.$$

5.2 Lineare Modelle mit koordinatengebundener Darstellung

Es sei

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

das statistische Modell gegeben durch:

- $(\Omega, \mathscr{F}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ für ein $n \in \mathbb{N}$.
- $\Theta = \Theta_0 \times (0, \infty)$ mit $\Theta_0 = \mathbb{R}^d$ für ein $d \in \mathbb{N}$ mit $d \leq n$ und

$$\mathbb{P}_\vartheta = \mathrm{N}(A\beta, \sigma^2 \mathrm{Id}), \quad \vartheta = (\beta, \sigma^2) \in \Theta$$

  mit einer Matrix $A \in \mathbb{R}^{n \times d}$, so dass $\mathrm{rg}(A) = d$.

Wir vermuten für die Stichprobe X also den linearen Zusammenhang

$$X = A\beta + \sigma\epsilon$$

wobei $\epsilon_1, \ldots, \epsilon_n \sim \mathrm{N}(0, 1)$ unabhängige, standardnormalverteilte Zufallsvariablen sind.

Definition 5.2.1. *Wir nennen $\mathscr{M}$ ein lineares Modell mit koordinatengebundener Darstellung.*

Beispiel 5.2.2. *Es sei $A = (\mathbb{1}) \in \mathbb{R}^{n\times 1}$. Dann gilt $\mathrm{rg}(A) = 1$. Wir haben in dem Fall das bekannte Gauß'sche Produktmodell*

$$\mathbb{P}_\vartheta = \mathrm{N}(\beta\mathbb{1}, \sigma^2\mathrm{Id}) = \mathrm{N}(\beta, \sigma^2)^{\otimes n}, \quad \vartheta = (\beta, \sigma^2) \in \Theta$$

vorliegen. Wir sprechen bei diesem Modell auch vom Einstichprobenproblem.

Beispiel 5.2.3 (Einfache lineare Regression). *Es sei $y \in \mathbb{R}^n$ mit $y \notin \mathrm{lin}\{\mathbb{1}\}$ für ein $n \geq 2$. Wir setzen*

$$A = \begin{pmatrix} \mathbb{1} & y \end{pmatrix} \in \mathbb{R}^{n\times 2}.$$

Dann gilt $\mathrm{rg}(A) = 2$. Dies ist die einfache lineare Regression, die wir in Abschnitt 5.1 kennengelernt haben.

Bemerkung 5.2.4. *Für die spezielle Parameterwahl $\vartheta = (\beta, \sigma^2)$ mit $\beta = (\mu, 0)$ für ein $\mu \in \mathbb{R}$ gilt $\mathbb{P}_\vartheta = \mathrm{N}(\mu, \sigma^2)^{\otimes n}$ wie beim Einstichprobenproblem.*

Beispiel 5.2.5 (Polynomiale Regression). *Es seien $d \in \mathbb{N}$ mit $d \leq n$ und $y \in \mathbb{R}^n$, welches d paarweise verschiedene Elemente besitzt. Wir setzen*

$$A = \begin{pmatrix} \mathbb{1} & y & y^2 & \dots & y^{d-1} \end{pmatrix} \in \mathbb{R}^{n\times d},$$

wobei $y^2, \dots, y^{d-1}$ komponentenweise gemeint sind. Nach dem Resultat über die Vandermonde-Matrizen gilt $\mathrm{rg}(A) = d$. In diesem Modell vermuten wir zwischen der Stichprobe X und den Kontrollvariablen y einen polynomiellen Zusammenhang

$$X_i = \beta_0 + \beta_1 y_i + \ldots + \beta_{d-1} y_i^{d-1} + \sigma\epsilon_i, \quad i = 1, \dots, n.$$

Beispiel 5.2.6 (Mehrfache lineare Regression). *Es seien $d \in \mathbb{N}$ mit $d \leq n$ und $y_1, \dots, y_{d-1} \in \mathbb{R}^n$ linear unabhängige Vektoren mit $\mathbb{1} \notin \mathrm{lin}\{y_1, \dots, y_{d-1}\}$. Wir setzen*

$$A = \begin{pmatrix} \mathbb{1} & y_1 & \dots & y_{d-1} \end{pmatrix} \in \mathbb{R}^{n\times d}.$$

Dann gilt $\mathrm{rg}(A) = d$. In diesem Modell vermuten wir zwischen der Stichprobe X und den Kontrollvariablen $y_1, \dots, y_{d-1}$ einen linearen Zusammenhang

$$X_i = \beta_0 + \beta_1 y_{1,i} + \ldots + \beta_{d-1} y_{d-1,i} + \sigma\epsilon_i, \quad i = 1, \dots, n.$$

Beispiel 5.2.7 (Mehrstichprobenproblem). *Es seien $d \in \mathbb{N}$ und $n_1, \ldots, n_d \in \mathbb{N}$, so dass $n = n_1 + \ldots + n_d$. Wir setzen*

$$A = \begin{pmatrix} \mathbb{1}_{n_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbb{1}_{n_d} \end{pmatrix} \in \mathbb{R}^{n \times d}.$$

Dann gilt $\operatorname{rg}(A) = d$. *In diesem Modell liegen bei der Stichprobe verschiedene Gruppen $X = (X_1, \ldots, X_d)$ mit $X_i \in \mathbb{R}^{n_i}$ für $i = 1, \ldots, d$ vor. Wir vermuten für jede Gruppe einen Zusammenhang*

$$X_i = \beta_i \mathbb{1} + \sigma^2 \epsilon_i.$$

Es gilt

$$\mathbb{P}_\vartheta = \bigotimes_{i=1}^{d} \mathrm{N}(\beta_i, \sigma^2)^{\otimes n_i}, \quad \vartheta = (\beta, \sigma^2) \in \Theta.$$

Für $i = 1, \ldots, d$ wird der Erwartungswert $\beta_i \in \mathbb{R}$ das i-te Gruppenmittel genannt.

Definition 5.2.8. *Ein Schätzer $\hat{\beta} : \Omega \to \Theta_0$ heißt ein Kleinste-Quadrate-Schätzer (KQS) für β, falls*

$$\|X - A\hat{\beta}\| = \min_{\beta \in \Theta_0} \|X - A\beta\|.$$

Definition 5.2.9. *Es sei $\varphi : \mathbb{R}^d \to \mathbb{R}^n$ die lineare Abbildung gegeben durch $\varphi(x) = Ax$ für alle $x \in \mathbb{R}^d$.*

Bemerkung 5.2.10. *Wegen* $\operatorname{rg}(A) = d$ *ist φ injektiv.*

Definition 5.2.11. *Wir setzen $W := \varphi(\mathbb{R}^d)$.*

Bemerkung 5.2.12. *Es gilt* $\dim W = d$, *und $\varphi : \mathbb{R}^d \to W$ ist ein Isomorphismus.*

Lemma 5.2.13. *Die Matrix $A^\top A \in \mathbb{R}^{d \times d}$ ist symmetrisch und positiv definit, und damit insbesondere invertierbar.*

Beweis. Die Symmetrie ist klar. Für jedes $b \in \mathbb{R}^d \setminus \{0\}$ gilt

$$\langle A^\top A b, b \rangle = \langle Ab, Ab \rangle = \|Ab\|^2 > 0,$$

da φ injektiv ist. □

Definition 5.2.14. *Die Moore-Penrose-Inverse von $A \in \mathbb{R}^{n \times d}$ ist definiert durch*

$$A^\dagger := (A^\top A)^{-1} A^\top \in \mathbb{R}^{d \times n}.$$

Lemma 5.2.15. *Es gilt* $A^\dagger A = \mathrm{Id} \in \mathbb{R}^{d\times d}$.

Beweis. In der Tat, es gilt

$$A^\dagger A = (A^\top A)^{-1} A^\top A = \mathrm{Id}.$$

□

Lemma 5.2.16. *Es existiert eine eindeutig bestimmte lineare Abbildung* $\Pi_W : \mathbb{R}^n \to W$*, so dass*

$$\|x - \Pi_W x\| = \min_{y\in W} \|x - y\|.$$

Definition 5.2.17. *Wir nennen* $\Pi_W : \mathbb{R}^n \to W$ *die* <u>*Orthogonalprojektion*</u> *auf* W.

Lemma 5.2.18. *Für einen linearen Operator* $\psi : \mathbb{R}^n \to W$ *sind folgende Aussagen äquivalent:*

(i) Es gilt $\psi = \Pi_W$.

(ii) Es gilt $\langle x - \psi(x), y\rangle = 0$ *für alle* $x \in \mathbb{R}^n$ *und* $y \in W$.

(iii) Es gilt $A^\top(x - \psi(x)) = 0$ *für alle* $x \in \mathbb{R}^n$.

Lemma 5.2.19. *Es gilt* $\Pi_W x = AA^\dagger x$ *für alle* $x \in \mathbb{R}^n$.

Beweis. Es gilt $AA^\dagger x \in W$ für alle $x \in \mathbb{R}^n$, da $W = \varphi(\mathbb{R}^d)$. Außerdem gilt

$$A^\top(x - AA^\dagger x) = A^\top(x - A(A^\top A)^{-1}A^\top x) = A^\top x - A^\top x = 0.$$

Also folgt die Behauptung mit Lemma 5.2.18. □

Lemma 5.2.20. *Für jedes* $x \in \mathbb{R}^n$ *hat die lineare Gleichung*

$$Ab = \Pi_W x, \quad b \in \mathbb{R}^d$$

die eindeutig bestimmte Lösung $b = A^\dagger x$.

Beweis. Nach Lemmas 5.2.19 und 5.2.15 gilt für jedes $b \in \mathbb{R}^d$

$$\begin{aligned} & Ab = \Pi_W x \\ \Leftrightarrow\ & Ab = AA^\dagger x \\ \Leftrightarrow\ & A^\dagger Ab = A^\dagger AA^\dagger x \\ \Leftrightarrow\ & b = A^\dagger x. \end{aligned}$$

□

Satz 5.2.21.

(a) *Es gibt einen eindeutig bestimmten KQS $\hat{\beta}$ für β; dieser ist gegeben durch*

$$\hat{\beta} = A^\dagger X.$$

(b) *Außerdem gilt die Darstellung $\hat{\beta} = A^\dagger \Pi_W X$.*

Beweis.

(a) Nach Lemma 5.2.16 gilt

$$\|X - A\hat{\beta}\| = \min_{\beta \in \mathbb{R}^d} \|X - A\beta\|$$

genau dann, wenn

$$A\hat{\beta} = \Pi_W X.$$

Dies ist nach Lemma 5.2.20 äquivalent zu

$$\hat{\beta} = A^\dagger X.$$

(b) Nach Lemmas 5.2.19 und 5.2.15 gilt

$$A^\dagger \Pi_W X = A^\dagger A A^\dagger X = A^\dagger X.$$

□

5.3 Lineare Modelle mit koordinatenfreier Darstellung

Es sei

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

das statistische Modell gegeben durch:

- $(\Omega, \mathscr{F}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ für ein $n \in \mathbb{N}$.
- $\Theta = \Theta_0 \times (0, \infty)$ mit $\Theta_0 = W$ für einen Unterraum $W \subset \mathbb{R}^n$ mit $\dim(W) = d \leq n$ und

$$\mathbb{P}_\vartheta = \mathrm{N}(\zeta, \sigma^2 \mathrm{Id}), \quad \vartheta = (\zeta, \sigma^2) \in \Theta.$$

Für die Stichprobe X gilt also

$$X = \zeta + \sigma\epsilon$$

wobei $\epsilon_1, \dots, \epsilon_n \sim \mathrm{N}(0,1)$ unabhängige, standardnormalverteilte Zufallsvariablen sind.

Definition 5.3.1. *Wir nennen $\mathscr{M}$ ein lineares Modell mit koordinatenfreier Darstellung.*

Bemerkung 5.3.2. *Es sei*

$$\mathscr{N} = \big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{Q}_\gamma : \gamma \in \Gamma)\big)$$

ein lineares Modell mit koordinatengebundener Darstellung. Es gilt also $\Gamma = \mathbb{R}^d \times (0,\infty)$ und

$$\mathbb{Q}_\gamma = \mathrm{N}(A\beta, \sigma^2 \mathrm{Id}), \quad \gamma = (\beta, \sigma^2) \in \Gamma.$$

Wir setzen $W = A\mathbb{R}^d$. Dann ist

$$\mathscr{M} = \big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

mit $\Theta = W \times (0,\infty)$ und

$$\mathbb{P}_\vartheta = \mathbb{Q}_{(A^\dagger\zeta, \sigma^2)} = \mathrm{N}(\zeta, \sigma^2 \mathrm{Id}), \quad \vartheta = (\zeta, \sigma^2) \in \Theta$$

ein lineares Modell mit koordinatenfreier Darstellung. Für den KQS $\hat{\beta}$ in $\mathscr{N}$ hat der Schätzer $\hat{\zeta} = A\hat{\beta}$ in $\mathscr{M}$ die Darstellung $\hat{\zeta} = \Pi_W X$. Letzteres folgt aus Satz 5.2.21(b).

Bemerkung 5.3.3. *Es sei*

$$\mathscr{M} = \big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big)$$

ein lineares Modell mit koordinatenfreier Darstellung. Es gilt also $\Theta = W \times (0,\infty)$ und

$$\mathbb{P}_\vartheta = \mathrm{N}(\zeta, \sigma^2 \mathrm{Id}), \quad \vartheta = (\zeta, \sigma^2) \in \Theta.$$

Es sei $A \in \mathbb{R}^{n \times d}$ eine Matrix, so dass $W = A\mathbb{R}^d$. Dann ist

$$\mathscr{N} = \big(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), (\mathbb{Q}_\gamma : \gamma \in \Gamma)\big)$$

mit $\Gamma = \mathbb{R}^d \times (0,\infty)$ und

$$\mathbb{Q}_\gamma = \mathbb{P}_{(A\beta, \sigma^2)} = \mathrm{N}(A\beta, \sigma^2 \mathrm{Id}), \quad \gamma = (\beta, \sigma^2) \in \Gamma$$

ein lineares Modell mit koordinatengebundener Darstellung. Für den Schätzer $\hat{\zeta} = \Pi_W X$ in $\mathscr{M}$ ist der Schätzer $\hat{\beta} = A^\dagger \hat{\zeta}$ in $\mathscr{N}$ der KQS für β. Dieser charakterisiert als die eindeutig bestimmte Lösung von $A\hat{\beta} = \hat{\zeta}$. Dies folgt aus Satz 5.2.21(b).

Beispiel 5.3.4. *Beim Gauß'schen Produktmodell (Beispiel 5.2.2) gilt* $W = \mathrm{lin}\{\mathbb{1}\}$.

Beispiel 5.3.5. *Bei der einfachen linearen Regression (Beispiel 5.2.3) gilt*

$$W = \mathrm{lin}\{\mathbb{1}, y\}.$$

Beispiel 5.3.6. *Bei der polynomialen Regression (Beispiel 5.2.5) gilt*

$$W = \mathrm{lin}\{\mathbb{1}, y, y^2, \dots, y^{d-1}\}.$$

Beispiel 5.3.7. *Bei der mehrfachen linearen Regression (Beispiel 5.2.6) gilt*

$$W = \mathrm{lin}\{\mathbb{1}, y_1, \dots, y_{d-1}\}.$$

Beispiel 5.3.8. *Beim Mehrstichprobenproblem (Beispiel 5.2.7) gilt*

$$W = \mathrm{lin}\left\{\begin{pmatrix} \mathbb{1}_{n_1} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \mathbb{1}_{n_d} \end{pmatrix}\right\}.$$

Beispiel 5.3.9. *Wir betrachten die einfache lineare Regression*

$$A = \begin{pmatrix} \mathbb{1} & y \end{pmatrix} \in \mathbb{R}^{n \times 2}.$$

Die Menge $\{v_1, v_2\}$ *mit*

$$v_1 = \frac{\mathbb{1}}{\sqrt{n}} \quad \textit{und} \quad v_2 = \frac{y - \bar{y}\mathbb{1}}{\sqrt{n-1}s_y}$$

ist eine Orthonormalbasis von $W = \mathrm{lin}\{\mathbb{1}, y\}$. *Wegen der Gleichung*

$$\langle y - \bar{y}\mathbb{1}, X\mathbb{1}\rangle = \langle y - \bar{y}\mathbb{1}, X - \bar{X}\mathbb{1}\rangle$$

folgt

$$\begin{aligned}
\Pi_W X &= \langle X, v_1\rangle v_1 + \langle X, v_2\rangle v_2 = \frac{\langle X, \mathbb{1}\rangle}{n}\mathbb{1} + \frac{1}{(n-1)s_y^2}\langle X, y - \bar{y}\mathbb{1}\rangle(y - \bar{y}\mathbb{1}) \\
&= \bar{X}\mathbb{1} + \frac{\langle y - \bar{y}\mathbb{1}, X - \bar{X}\mathbb{1}\rangle}{(n-1)s_y^2}(y - \bar{y}\mathbb{1}) = \bar{X}\mathbb{1} + \frac{s_{yX}}{s_y^2}(y - \bar{y}\mathbb{1}) \\
&= \left(\bar{X} - \frac{s_{yX}}{s_y^2}\bar{y}\right)\mathbb{1} + \frac{s_{yX}}{s_y^2}y.
\end{aligned}$$

Die Gleichung

$$A\hat{\beta} = \hat{\zeta}$$

aus Bemerkung 5.3.3 lautet hier also

$$\hat{\beta}_0 \mathbb{1} + \hat{\beta}_1 y = \left(\bar{X} - \frac{s_{yX}}{s_y^2}\bar{y}\right)\mathbb{1} + \frac{s_{yX}}{s_y^2} y.$$

Also erhalten wir den KQS $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ *für* β *als*

$$\hat{\beta}_1 = \frac{s_{yX}}{s_y^2}$$

und

$$\hat{\beta}_0 = \bar{X} - \hat{\beta}_1 \bar{y}.$$

Beispiel 5.3.10. *Wir betrachten das Mehrstichprobenproblem*

$$A = \begin{pmatrix} \mathbb{1}_{n_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbb{1}_{n_d} \end{pmatrix} = \begin{pmatrix} f_1 & \dots & f_d \end{pmatrix} \in \mathbb{R}^{n \times d}.$$

Die Menge $\{v_1, \dots, v_d\}$ *mit*

$$v_i = \frac{f_i}{\sqrt{n_i}}$$

ist eine Orthonormalbasis von $W = \operatorname{lin}\{f_1, \dots, f_d\}$. *Es folgt*

$$\Pi_W X = \sum_{i=1}^d \langle X, v_i \rangle v_i = \sum_{i=1}^d \frac{1}{n_i} \langle X, f_i \rangle f_i = \sum_{i=1}^d \bar{X}_i f_i,$$

wobei die Stichprobe $X = (X_1, \dots, X_d)$ *die Gruppen* $X_i \in \mathbb{R}^{n_i}$ *für* $i = 1, \dots, d$ *enthält. Die Gleichung*

$$A\hat{\beta} = \hat{\zeta}$$

aus Bemerkung 5.3.3 lautet hier also

$$\sum_{i=1}^d \hat{\beta}_i f_i = \sum_{i=1}^d \bar{X}_i f_i.$$

Also erhalten wir den KQS

$$\hat{\beta} = (\bar{X}_1, \dots, \bar{X}_d).$$

5.4 Punktschätzer

Wir betrachten ein lineares Modell $\mathscr{M}$ mit koordinatenfreier Darstellung. Wir nehmen an, dass $d < n$, wobei $d = \dim(W)$.

Definition 5.4.1. *Wir setzen*

$$\hat{\zeta} := \Pi_W X \quad \textit{und} \quad \hat{s}^2 := \frac{1}{n-d}\|X - \hat{\zeta}\|^2.$$

Lemma 5.4.2. *Es gilt $\langle b, v\rangle = \langle \Pi_W b, v\rangle$ für alle $b \in \mathbb{R}^n$ und $v \in W$.*

Beweis. Es gilt

$$\langle b, v\rangle = \langle \Pi_W b, v\rangle + \underbrace{\langle \Pi_{W^\perp} b, v\rangle}_{=0} = \langle \Pi_W b, v\rangle.$$

□

Wir fixieren ein $\vartheta = (\zeta, \sigma^2) \in \Theta$, und betrachten das Wahrscheinlichkeitsmaß $\mathbb{P}_\vartheta$. Weiterhin fixieren wir eine Orthonormalbasis $\{v_1, \ldots, v_n\}$ von $\mathbb{R}^n$, so dass $\{v_1, \ldots, v_d\}$ eine Orthonormalbasis von W ist.

Definition 5.4.3.

(a) *Wir setzen $\widetilde{X}_i := \langle X, v_i\rangle$ für $i = 1, \ldots, n$.*

(b) *Wir setzen $Y_i := \langle X - \zeta, v_i\rangle$ für $i = 1, \ldots, n$.*

(c) *Wir setzen $Z_i := \frac{Y_i}{\sigma}$ für $i = 1, \ldots, n$.*

Lemma 5.4.4.

(a) *Die Zufallsvariablen $\widetilde{X}_1, \ldots, \widetilde{X}_n$ sind unabhängig mit $\widetilde{X}_i \sim \mathrm{N}(\langle \zeta, v_i\rangle, \sigma^2)$ für alle $i = 1, \ldots, n$.*

(b) *Die Zufallsvariablen $Y_1, \ldots, Y_n$ sind unabhängig mit $Y_i \sim \mathrm{N}(0, \sigma^2)$ für alle $i = 1, \ldots, n$.*

(c) *Die Zufallsvariablen $Z_1, \ldots, Z_n$ sind unabhängig mit $Y_i \sim \mathrm{N}(0, 1)$ für alle $i = 1, \ldots, n$.*

Beweis.

(a) Folgt aus Korollar 4.3.11, da $\widetilde{X} = AX$ mit der orthogonalen Matrix

$$A = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

(b) Folgt aus Korollar 4.3.11, da $Y = A(X - \zeta)$ mit der orthogonalen Matrix

$$A = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

(c) ✓

□

Lemma 5.4.5. *Es gelten die Darstellungen*

$$\hat{\zeta} = \sum_{i=1}^{d} \widetilde{X}_i v_i = \zeta + \sum_{i=1}^{d} Y_i v_i \quad \textit{und} \quad \hat{s}^2 = \frac{1}{n-d} \sum_{i=d+1}^{n} \widetilde{X}_i^2 = \frac{1}{n-d} \sum_{i=d+1}^{n} Y_i^2.$$

Beweis. Es gilt

$$\hat{\zeta} = \Pi_W X = \sum_{i=1}^{d} \langle X, v_i \rangle v_i = \sum_{i=1}^{d} \widetilde{X}_i v_i,$$

und wegen $\zeta \in W$ gilt

$$\hat{\zeta} = \sum_{i=1}^{d} \langle \zeta, v_i \rangle v_i + \sum_{i=1}^{d} \langle X - \zeta, v_i \rangle v_i = \zeta + \sum_{i=1}^{d} Y_i v_i.$$

Weiterhin gilt

$$X - \hat{\zeta} = \sum_{i=d+1}^{n} \widetilde{X}_i v_i.$$

Wegen $\zeta \in W$ gilt $\langle \zeta, v_i \rangle = 0$ für alle $i = d+1, \ldots, n$, und damit $\widetilde{X}_i = Y_i$ für alle $i = d+1, \ldots, n$. Es folgt

$$\hat{s}^2 = \frac{1}{n-d} \|X - \hat{\zeta}\|^2 = \frac{1}{n-d} \sum_{i=d+1}^{n} \widetilde{X}_i^2 = \frac{1}{n-d} \sum_{i=d+1}^{n} Y_i^2.$$

□

Lemma 5.4.6. *Es gilt* $\mathbb{E}[\hat{\zeta}] = \zeta$.

Beweis. Nach Lemma 5.4.5 gilt

$$\mathbb{E}[\hat{\zeta}] = \mathbb{E}\left[\zeta + \sum_{i=1}^{d} Y_i v_i\right] = \zeta + \sum_{i=1}^{d} \mathbb{E}[Y_i] v_i = \zeta.$$

□

Lemma 5.4.7. *Für alle $b, c \in \mathbb{R}^n$ gilt*

$$\mathrm{Cov}\big(\langle b, \hat{\zeta}\rangle, \langle c, \hat{\zeta}\rangle\big) = \langle \Pi_W b, \Pi_W c\rangle \sigma^2.$$

Beweis. Nach Lemmas 5.4.5 und 5.4.2 gilt

$$\begin{aligned}
&\mathrm{Cov}\big(\langle b, \hat{\zeta}\rangle, \langle c, \hat{\zeta}\rangle\big) = \mathrm{Cov}\left(\left\langle b, \sum_{i=1}^{d} Y_i v_i\right\rangle, \left\langle c, \sum_{j=1}^{d} Y_j v_j\right\rangle\right) \\
&= \sum_{i=1}^{d}\sum_{j=1}^{d} \langle b, v_i\rangle \langle c, v_j\rangle \mathrm{Cov}(Y_i, Y_j) = \sum_{i=1}^{d} \langle b, v_i\rangle \langle v_i, c\rangle \mathrm{Var}[Y_i] \\
&= \sum_{i=1}^{d} \langle b, v_i\rangle \langle v_i, c\rangle \sigma^2 = \sum_{i=1}^{d} \langle \Pi_W b, v_i\rangle \langle v_i, \Pi_W c\rangle \sigma^2 = \langle \Pi_W b, \Pi_W c\rangle \sigma^2.
\end{aligned}$$

□

Lemma 5.4.8. *Es gilt*

$$\hat{\zeta} \sim \mathrm{N}\big(\zeta, (\langle \Pi_W e_i, \Pi_W e_j\rangle \sigma^2)_{i,j=1,\dots,n}\big).$$

Beweis. Folgt aus Lemmas 5.4.6 und 5.4.7. □

Lemma 5.4.9. *Für alle $b \in \mathbb{R}^n$ gilt*

$$\langle b, \hat{\zeta}\rangle \sim \mathrm{N}(\langle b, \zeta\rangle, \|\Pi_W b\|^2 \sigma^2).$$

Beweis. Folgt aus Lemmas 5.4.6 und 5.4.7. □

Satz 5.4.10.

(a) Die Zufallsvariablen $\hat{\zeta}$ und $\hat{s}^2$ sind unabhängig.

(b) Für jedes $b \in \mathbb{R}^n \setminus W^\perp$ gilt

$$\frac{\langle b, \hat{\zeta}\rangle - \langle b, \zeta\rangle}{\sigma \|\Pi_W b\|} \sim \mathrm{N}(0, 1).$$

(c) Es gilt

$$\frac{n-d}{\sigma^2}\hat{s}^2 \sim \chi^2_{n-d}.$$

(d) Für jedes $b \in \mathbb{R}^n \setminus W^\perp$ gilt

$$\frac{\langle b, \hat{\zeta}\rangle - \langle b, \zeta\rangle}{\hat{s}\|\Pi_W b\|} \sim t_{n-d}.$$

Beweis.

(a) Folgt aus Lemmas 5.4.4 und 5.4.5.

(b) Folgt aus Lemma 5.4.9.

(c) Mit Lemma 5.4.5 und Satz 2.2.2(a) gilt

$$\frac{n-d}{\sigma^2}\hat{s}^2 = \frac{1}{\sigma^2}\sum_{i=d+1}^{n} Y_i^2 = \sum_{i=d+1}^{n} Z_i^2 \sim \chi^2_{n-d}.$$

(d) Wir setzen

$$U := \frac{\langle b, \hat{\zeta}\rangle - \langle b, \zeta\rangle}{\sigma\|\Pi_W b\|} \quad \text{und} \quad V := \frac{n-d}{\sigma^2}\hat{s}^2$$

Nach den Teilen (a)–(c) sind U und V unabhängig mit $U \sim \mathrm{N}(0,1)$ und $V \sim \chi^2_{n-d}$. Also folgt mit Satz 2.2.2(b)

$$\frac{\langle b, \hat{\zeta}\rangle - \langle b, \zeta\rangle}{\hat{s}\|\Pi_W b\|} = \frac{\langle b, \hat{\zeta}\rangle - \langle b, \zeta\rangle}{\sigma\|\Pi_W b\|} \cdot \sqrt{\frac{\sigma^2}{\hat{s}^2}} = \frac{U}{\sqrt{V/(n-d)}} \sim t_{n-d}.$$

□

Lemma 5.4.11. *Für $W = \mathrm{lin}\{\mathbb{1}\}$ gilt $\hat{\zeta} = \bar{X}\mathbb{1}$ und $\hat{s}^2 = s^2$.*

Beweis. Wir setzen $v_1 = \frac{\mathbb{1}}{\sqrt{n}} \in W$. Dann gilt $\|v_1\| = 1$, und folglich

$$\hat{\zeta} = \Pi_W X = \langle X, v_1\rangle v_1 = \left\langle X, \frac{\mathbb{1}}{\sqrt{n}}\right\rangle \frac{\mathbb{1}}{\sqrt{n}} = \frac{1}{n}\langle X, \mathbb{1}\rangle \mathbb{1} = \bar{X}\mathbb{1}.$$

Weiterhin gilt

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\|X - \bar{X}\mathbb{1}\| \\ &= \frac{1}{n-1}\|X - \hat{\zeta}\|^2 = \hat{s}^2. \end{aligned}$$

□

Bemerkung 5.4.12. *Wir betrachten den Fall* $W = \mathrm{lin}\{\mathbb{1}\}$. *Dann gilt* $\hat{\zeta} = \bar{X}\mathbb{1}$ *und* $\zeta = \mu\mathbb{1}$ *mit* $\mu \in \mathbb{R}$. *Wir setzen* $b := \frac{\mathbb{1}}{n}$. *Dann gilt* $b \in W$ *mit*

$$\|b\| = \left\| \frac{\mathbb{1}}{n} \right\| = \frac{1}{n}\|\mathbb{1}\| = \frac{1}{n}\sqrt{1^2 + \ldots + 1^2} = \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}},$$

und für alle $\gamma \in \mathbb{R}$ *gilt*

$$\langle b, \gamma\mathbb{1}\rangle = \left\langle \frac{\mathbb{1}}{n}, \gamma\mathbb{1} \right\rangle = \frac{\gamma}{n}\langle \mathbb{1}, \mathbb{1}\rangle = \gamma.$$

Also verallgemeinert Satz 5.4.10 den Satz 2.2.3.

Lemma 5.4.13. *Es gilt* $\mathbb{E}[\hat{s}^2] = \sigma^2$.

Beweis. Wir setzen

$$V := \frac{n-d}{\sigma^2}\hat{s}^2.$$

Nach Satz 5.4.10(c) gilt $V \sim \chi^2_{n-d}$, und daher $\mathbb{E}[V] = n - d$. Also folgt

$$\mathbb{E}[\hat{s}^2] = \frac{\sigma^2}{n-d}\mathbb{E}[V] = \sigma^2.$$

□

Definition 5.4.14. *Zwei statistische Modelle*

$$\mathscr{M} = \big(\Omega, \mathscr{F}, (\mathbb{P}_\vartheta : \vartheta \in \Theta)\big) \quad \text{und} \quad \mathscr{N} = \big(\Omega, \mathscr{F}, (\mathbb{Q}_\gamma : \gamma \in \Gamma)\big)$$

heißen äquivalent, falls

$$\{\mathbb{P}_\vartheta : \vartheta \in \Theta\} = \{\mathbb{Q}_\gamma : \gamma \in \Gamma\}.$$

Lemma 5.4.15. *Es seien* $\mathscr{M}$ *und* $\mathscr{N}$ *zwei äquivalente statistische Modelle, und es sei* $S : \Omega \to E$ *eine Statistik mit Werten in einem messbaren Raum* $(E, \mathscr{E})$. *Dann sind folgende Aussagen äquivalent:*

(i) S *ist suffizient bezüglich* $\mathscr{M}$.

(ii) S *ist suffizient bezüglich* $\mathscr{N}$.

Beweis. (i) ⇒ (ii): Es sei $f : \Omega \to \mathbb{R}$ eine messbare, beschränkte Funktion. Nach Voraussetzung existiert eine $\sigma(S)$-messbare Zufallsvariable $g : \Omega \to \mathbb{R}$, so dass

$$\mathbb{E}_{\mathbb{P}_\vartheta}[f \mid S] = g \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für jedes } \vartheta \in \Theta.$$

Für jedes $\gamma \in \Gamma$ existiert ein $\vartheta \in \Theta$, so dass $\mathbb{Q}_\gamma = \mathbb{P}_\vartheta$. Also folgt

$$\mathbb{E}_{\mathbb{Q}_\gamma}[f \mid S] = g \quad \mathbb{Q}_\gamma\text{-fast sicher} \quad \text{für jedes } \gamma \in \Gamma.$$

(ii) ⇒ (i): verläuft analog. □

Lemma 5.4.16. *Es seien $\mathscr{M}$ und $\mathscr{N}$ zwei äquivalente statistische Modelle, und es sei $S : \Omega \to E$ eine Statistik mit Werten in einem messbaren Raum $(E, \mathscr{E})$. Dann sind folgende Aussagen äquivalent:*

(i) S ist vollständig bezüglich $\mathscr{M}$.

(ii) S ist vollständig bezüglich $\mathscr{N}$.

Beweis. (i) $\Rightarrow$ (ii): Es sei $h : E \to \mathbb{R}$ eine messbare Funktion mit $h(S) \in \mathscr{L}^2(\mathbb{Q}_\gamma)$ für alle $\gamma \in \Gamma$, so dass

$$\mathbb{E}_{\mathbb{Q}_\gamma}[h(S)] = 0 \quad \text{für alle } \gamma \in \Gamma.$$

Für jedes $\vartheta \in \Theta$ existiert ein $\gamma \in \Gamma$, so dass $\mathbb{P}_\vartheta = \mathbb{Q}_\gamma$. Also folgt $h(S) \in \mathscr{L}^2(\mathbb{P}_\vartheta)$ für alle $\vartheta \in \Theta$, sowie

$$\mathbb{E}_{\mathbb{P}_\vartheta}[h(S)] = 0 \quad \text{für alle } \vartheta \in \Theta.$$

Nach Voraussetzung folgt

$$h(S) = 0 \quad \mathbb{P}_\vartheta\text{-fast sicher} \quad \text{für alle } \vartheta \in \Theta.$$

Für jedes $\gamma \in \Gamma$ existiert ein $\vartheta \in \Theta$, so dass $\mathbb{Q}_\gamma = \mathbb{P}_\vartheta$. Also folgt

$$h(S) = 0 \quad \mathbb{Q}_\gamma\text{-fast sicher} \quad \text{für alle } \gamma \in \Gamma.$$

(ii) $\Rightarrow$ (i): verläuft analog. $\square$

Satz 5.4.17. *Die Statistik*

$$S = \left(\widetilde{X}_1, \dots, \widetilde{X}_d, \sum_{i=1}^n \widetilde{X}_i^2 \right)$$

ist vollständig und suffizient.

Beweis. Da die Statistik S von der transformierten Stichprobe $\widetilde{X}$ abhängt, dürfen wir nach Lemma 5.4.4 das statistische Modell mit

$$\mathbb{P}_\vartheta = \bigotimes_{i=1}^n \mathrm{N}(\langle \zeta, v_i \rangle, \sigma^2), \quad \vartheta = (\zeta, \sigma^2) \in W \times (0, \infty) = \Theta$$

betrachten. Gemäß Lemmas 5.4.15 und 5.4.16 dürfen wir das statistische Modell mit

$$\mathbb{Q}_\vartheta = \bigotimes_{i=1}^d \mathrm{N}(\eta_i, \sigma^2) \times \bigotimes_{i=d+1}^n \mathrm{N}(0, \sigma^2), \quad \vartheta = (\eta, \sigma^2) \in \mathbb{R}^d \times (0, \infty) =: \Gamma$$

betrachten. Die Likelihood-Funktion ist gegeben durch

$$\begin{aligned} L(\vartheta, x) &= \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\eta_i)^2}{2\sigma^2} \right) \cdot \prod_{i=d+1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2} \right) \\ &= \exp\left(-\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{1}{\sigma^2}\sum_{i=1}^{d} x_i\eta_i - \frac{1}{2\sigma^2}\sum_{i=1}^{d}\eta_i^2 \right). \end{aligned}$$

Also ist $\mathscr{M}$ eine mehrdimensionale exponentielle Familie der Form

$$L(\vartheta, x) = \exp\big(\langle a(\vartheta), T(x)\rangle_{\mathbb{R}^{d+1}} - b(\vartheta)\big) \cdot h(x) \quad \text{für alle } (\vartheta, x) \in \Gamma \times \Omega,$$

wobei

$$\begin{aligned} a(\vartheta) &= \left(\frac{\eta_1}{\sigma^2}, \ldots, \frac{\eta_d}{\sigma^2}, -\frac{1}{2\sigma^2} \right) = \left(\frac{\eta}{\sigma^2}, -\frac{1}{2\sigma^2} \right), \\ b(\vartheta) &= \frac{n}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{d}\eta_i^2 = \frac{n}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\|\eta\|_{\mathbb{R}^d}^2, \\ T(x) &= \left(x_1, \ldots, x_d, \sum_{i=1}^{n} x_i^2 \right) = \left((x_i)_{i=1,\ldots,d}, \|x\|_{\mathbb{R}^n}^2 \right), \\ h(x) &= 1. \end{aligned}$$

Für jedes $\vartheta \in \Gamma$ gilt

$$J_a(\vartheta) = \begin{pmatrix} \frac{\mathrm{Id}}{\sigma^2} & -\frac{\eta}{\sigma^4} \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix} \in \mathbb{R}^{(d+1)\times(d+1)},$$

und daher

$$\det J_a(\vartheta) = \frac{1}{2\sigma^{2d+4}} > 0.$$

Nun folgt mit Satz 1.10.10, dass S vollständig und suffizient ist. □

Satz 5.4.18.

(a) *Für jedes $b \in \mathbb{R}^n$ ist $\langle b, \hat{\zeta}\rangle$ der eindeutig bestimmte gleichmäßig beste Schätzer für $\langle b, \zeta\rangle$.*

(b) *$\hat{s}^2$ ist der eindeutig bestimmte gleichmäßig beste Schätzer für σ^2.*

Beweis. Nach Satz 5.4.17 ist die Statistik

$$S = \left(\widetilde{X}_1, \ldots, \widetilde{X}_d, \sum_{i=1}^{n} \widetilde{X}_i^2 \right)$$

vollständig und suffizient.

(a) Wir definieren

$$h : \mathbb{R}^{d+1} \to \mathbb{R}, \quad h(x) = \left\langle b, \sum_{i=1}^{d} x_i v_i \right\rangle.$$

Nach Lemma 5.4.5 gilt

$$\langle b, \hat{\zeta} \rangle = \left\langle b, \sum_{i=1}^{d} \widetilde{X}_i v_i \right\rangle = h(S).$$

Nach Lemma 5.4.9 gilt außerdem für alle $\vartheta = (\zeta, \sigma^2) \in \Theta$

$$\mathbb{E}_\vartheta[h(S)] = \mathbb{E}_\vartheta[\langle b, \hat{\zeta} \rangle] = \langle b, \zeta \rangle.$$

Also folgt mit Korollar 1.9.4, dass $\langle b, \hat{\zeta} \rangle$ der eindeutig bestimmte gleichmäßig beste Schätzer für $\langle b, \zeta \rangle$ ist.

(b) Wir definieren

$$h : \mathbb{R}^{d+1} \to \mathbb{R}, \quad h(x) = \frac{1}{n-d} \left(x_{d+1} - \sum_{i=1}^{d} x_i^2 \right).$$

Nach Lemma 5.4.5 gilt

$$\hat{s}^2 = \frac{1}{n-d} \sum_{i=d+1}^{n} \widetilde{X}_i^2 = h(S).$$

Nach Lemma 5.4.13 gilt außerdem für alle $\vartheta = (\zeta, \sigma^2) \in \Theta$

$$\mathbb{E}_\vartheta[h(S)] = \sigma^2.$$

Also folgt mit Korollar 1.9.4, dass $\hat{s}^2$ der eindeutig bestimmte gleichmäßig beste Schätzer für σ^2 ist.

$\square$

Definition 5.4.19. *Ein Schätzer $T : \Omega \to \mathbb{R}$ heißt linear, falls $T = \langle b, X \rangle$ für ein $b \in \mathbb{R}^n$.*

Definition 5.4.20. *Es seien $\tau : \Theta \to \mathbb{R}$ eine Kenngröße und $T : \Omega \to \mathbb{R}$ ein erwartungstreuer, linearer Schätzer für τ. Dann heißt T ein gleichmäßig bester linearer Schätzer (oder BLUE für "best linear unbiased estimator"), wenn*

$$\mathrm{Var}_\vartheta[T] \le \mathrm{Var}_\vartheta[S]$$

für alle $\vartheta \in \Theta$ und für jeden erwartungstreuen, linearen Schätzer $S : \Omega \to \mathbb{R}$ für τ.

Satz 5.4.21 (Satz von Gauß-Markov). *Für jedes $b \in \mathbb{R}^n$ ist $\langle b, \hat{\zeta} \rangle$ der eindeutig bestimmte gleichmäßig beste lineare Schätzer für $\langle b, \zeta \rangle$, und es gilt*

$$\mathrm{Var}_\vartheta[\langle b, \hat{\zeta} \rangle] = \|\Pi_W b\|^2 \sigma^2 \quad \textit{für alle } \vartheta = (\zeta, \sigma^2) \in \Theta.$$

Beweis. Wir setzen $T := \langle b, \hat{\zeta} \rangle$. Dann gilt

$$T = \langle b, \Pi_W X \rangle = \langle \Pi_W b, X \rangle.$$

Also ist T ein linearer Schätzer, der nach Lemma 5.4.9 erwartungstreu für $\langle b, \zeta \rangle$ ist. Nun sei $c \in \mathbb{R}^n$, so dass der Schätzer $S = \langle c, X \rangle$ erwartungstreu für $\langle b, \zeta \rangle$ ist. Dann gilt für alle $\vartheta = (\zeta, \sigma^2) \in \Theta$

$$\langle b, \zeta \rangle = \mathbb{E}_\vartheta[S] = \mathbb{E}_\vartheta[\langle c, X \rangle] = \langle c, \zeta \rangle.$$

Also gilt

$$\langle b - c, \zeta \rangle = 0 \quad \text{für alle } \zeta \in W.$$

Es folgt $b - c \in W^\perp$, und damit $\Pi_W b = \Pi_W c$. Nun sei $\vartheta = (\zeta, \sigma^2) \in \Theta$ beliebig. Mit Lemma 5.4.9 folgt

$$\mathrm{Var}_\vartheta[T] = \mathrm{Var}_\vartheta[\langle b, \hat{\zeta} \rangle] = \|\Pi_W b\|^2 \sigma^2 = \|\Pi_W c\|^2 \sigma^2.$$

Außerdem gilt

$$\mathrm{Var}_\vartheta[S] = \mathrm{Var}_\vartheta[\langle c, X \rangle] = \mathrm{Var}_\vartheta\left[\sum_{i=1}^n c_i X_i\right] = \sum_{i=1}^n c_i^2 \sigma^2 = \|c\|^2 \sigma^2.$$

Wegen

$$\|c\|^2 = \|\Pi_W c\|^2 + \|\Pi_{W^\perp} c\|^2$$

gilt also

$$\mathrm{Var}_\vartheta[T] \leq \mathrm{Var}_\vartheta[S] \quad \text{für alle } \vartheta \in \Theta,$$

und folglich ist T ein gleichmäßig bester linearer Schätzer für $\langle b, \zeta \rangle$. Zum Beweis der Eindeutigkeit nehmen wir an, dass

$$\mathrm{Var}_\vartheta[T] = \mathrm{Var}_\vartheta[S] \quad \text{für alle } \vartheta \in \Theta.$$

Dann folgt $\|\Pi_{W^\perp} c\| = 0$, und daher $c \in W$. Nun folgt aber $c = \Pi_W c = \Pi_W b$, und damit $T = S$. □

Bemerkung 5.4.22. *Selbstverständlich läßt sich der Satz von Gauß-Markov (Satz 5.4.21) einfacher mit Hilfe von Satz 5.4.18 beweisen.*

5.5 Bereichsschätzer

Wir betrachten ein lineares Modell $\mathscr{M}$ mit koordinatenfreier Darstellung. Wir nehmen an, dass $d < n$, wobei $d = \dim(W)$.

Satz 5.5.1.

(a) *Für jedes $b \in \mathbb{R}^n \setminus W$ ist die Zufallsvariable*

$$\pi(\vartheta, \cdot) = \frac{\langle b, \hat{\zeta}\rangle - \langle b, \zeta\rangle}{\hat{s}\|\Pi_W b\|}, \quad \vartheta = (\zeta, \sigma^2)$$

eine Pivot-Statistik mit Pivot-Verteilung $\eta = t_{n-d}$.

(b) *Für jedes $b \in \mathbb{R}^n \setminus W$ und jedes $\alpha \in (0,1)$ ist das Zufallsintervall $C_b : \Omega \to \mathcal{B}(\mathbb{R})$ gegeben durch*

$$C_b = \langle b, \hat{\zeta}\rangle \pm t_{n-d,1-\alpha/2}\hat{s}\|\Pi_W b\|$$

ein Konfidenzintervall für $\langle b, \zeta\rangle$ zum Niveau α.

(c) *Für jedes $\alpha \in (0,1)$ ist die zufällige Kugel $C : \Omega \to \mathcal{B}(W)$ gegeben durch*

$$C = \hat{\zeta} + B_W(\sqrt{d}\, t_{n-d,1-\alpha/(2d)}\hat{s})$$

mit

$$B_W(\epsilon) := \{x \in W : \|x\| \leq \epsilon\}$$

ein Konfidenzbereich für ζ zum Niveau α.

Beweis.

(a) Folgt aus Satz 5.4.10(d).

(b) Es sei F die Verteilungsfunktion von η. Mit $c := t_{n-d,1-\alpha/2}$ und $B = [-c, c]$ gilt wegen der Symmetrie der Dichte

$$\begin{aligned}\eta(B) &= \eta((-\infty, c]) - \eta((-\infty, -c)) = F(c) - F(-c) = F(c) - (1 - F(c)) \\ &= 2F(c) - 1 = 2\Big(1 - \frac{\alpha}{2}\Big) - 1 = 1 - \alpha,\end{aligned}$$

und für alle $\vartheta = (\zeta, \sigma^2) \in \Theta$ gilt

$$\begin{aligned}
\{\langle b, \zeta\rangle \in C_b\} &= \Big\{\langle b, \zeta\rangle \in \langle b, \hat{\zeta}\rangle \pm t_{n-d,1-\alpha/2}\hat{s}\|\Pi_W b\|\Big\} \\
&= \Big\{\langle b, \hat{\zeta}\rangle - c\hat{s}\|\Pi_W b\| \le \langle b, \zeta\rangle \le \langle b, \hat{\zeta}\rangle + c\hat{s}\|\Pi_W b\|\Big\} \\
&= \Big\{\langle b, \hat{\zeta}\rangle - \langle b, \zeta\rangle - c\hat{s}\|\Pi_W b\| \le 0 \le \langle b, \hat{\zeta}\rangle - \langle b, \zeta\rangle + c\hat{s}\|\Pi_W b\|\Big\} \\
&= \{\pi(\vartheta, \cdot) - c \le 0 \le \pi(\vartheta, \cdot) + c\} \\
&= \{|\pi(\vartheta, \cdot)| \le c\} = \{\pi(\vartheta, \cdot) \in B\}.
\end{aligned}$$

Also folgt dieser Teil mit Satz 2.1.5.

(c) Es sei $c := t_{n-d,1-\alpha/(2d)}$. Es sei $\{v_1, \ldots, v_d\}$ eine Orthonormalbasis von W. Nach Teil (b) und Satz 2.1.3 ist das Zufallsrechteck $\widetilde{C} : \Omega \to \mathcal{B}(\mathbb{R}^d)$ gegeben durch

$$\widetilde{C} := \Big(\langle v_1, \hat{\zeta}\rangle \pm c\hat{s}\Big) \times \ldots \times \Big(\langle v_d, \hat{\zeta}\rangle \pm c\hat{s}\Big)$$

ist ein Konfidenzbereich für $\langle v, \zeta\rangle := (\langle v_1, \zeta\rangle, \ldots, \langle v_d, \zeta\rangle)$ zum Niveau α. Für alle $\vartheta = (\zeta, \sigma^2) \in \Theta$ folgt

$$\begin{aligned}
\mathbb{P}_\vartheta(\zeta \in C) &= \mathbb{P}_\vartheta\big(\|\hat{\zeta} - \zeta\| \le \sqrt{d}c\hat{s}\big) = \mathbb{P}_\vartheta\big(\|\hat{\zeta} - \zeta\|^2 \le dc^2\hat{s}^2\big) \\
&= \mathbb{P}_\vartheta\bigg(\sum_{i=1}^d |\langle v_i, \hat{\zeta} - \zeta\rangle|^2 \le dc^2\hat{s}^2\bigg) \ge \mathbb{P}_\vartheta\bigg(\bigcap_{i=1}^d \big\{|\langle v_i, \hat{\zeta} - \zeta\rangle|^2 \le c^2\hat{s}^2\big\}\bigg) \\
&= \mathbb{P}_\vartheta\bigg(\bigcap_{i=1}^d \big\{|\langle v_i, \hat{\zeta} - \zeta\rangle| \le c\hat{s}\big\}\bigg) = \mathbb{P}_\vartheta\big(\langle v, \zeta\rangle \in \widetilde{C}\big) \ge 1 - \alpha.
\end{aligned}$$

□

Satz 5.5.2.

(a) Die Zufallsvariable

$$\pi(\vartheta, \cdot) = \frac{n-d}{\sigma^2}\hat{s}^2, \quad \vartheta = (\zeta, \sigma^2)$$

ist eine Pivot-Statistik mit Pivot-Verteilung $\eta = \chi^2_{n-d}$.

(b) Für jedes $\alpha \in (0,1)$ ist das Zufallsintervall $C : \Omega \to \mathcal{B}(\mathbb{R})$ gegeben durch

$$C = \left[\frac{n-d}{\chi^2_{n-d,1-\alpha/2}}\hat{s}^2, \frac{n-d}{\chi^2_{n-d,\alpha/2}}\hat{s}^2\right]$$

ein Konfidenzintervall für σ^2 zum Niveau α.

Beweis.

(a) Folgt aus Satz 5.4.10(c).

(b) Es sei F die Verteilungsfunktion von η. Mit $B = [\chi^2_{n-d,\alpha/2}, \chi^2_{n-d,1-\alpha/2}]$ gilt

$$\begin{aligned}\eta(B) &= \eta\big((-\infty, \chi^2_{n-d,1-\alpha/2}]\big) - \eta\big((-\infty, \chi^2_{n-d,\alpha/2})\big)\\ &= F(\chi^2_{n-d,1-\alpha/2}) - F(\chi^2_{n-d,\alpha/2}) = \Big(1 - \frac{\alpha}{2}\Big) - \frac{\alpha}{2} = 1 - \alpha,\end{aligned}$$

und für alle $\vartheta = (\zeta, \sigma^2) \in \Theta$ gilt

$$\begin{aligned}\{\sigma^2 \in C\} &= \left\{\sigma^2 \in \left[\frac{n-d}{\chi^2_{n-d,1-\alpha/2}}\hat{s}^2, \frac{n-d}{\chi^2_{n-d,\alpha/2}}\hat{s}^2\right]\right\}\\ &= \left\{\frac{n-d}{\chi^2_{n-d,1-\alpha/2}}\hat{s}^2 \le \sigma^2 \le \frac{n-d}{\chi^2_{n-d,\alpha/2}}\hat{s}^2\right\}\\ &= \left\{\frac{1}{\chi^2_{n-d,1-\alpha/2}} \le \frac{1}{\pi(\vartheta,\cdot)} \le \frac{1}{\chi^2_{n-d,\alpha/2}}\right\}\\ &= \{\chi^2_{n-d,\alpha/2} \le \pi(\vartheta,\cdot) \le \chi^2_{n-d,1-\alpha/2}\}\\ &= \{\pi(\vartheta,\cdot) \in B\}.\end{aligned}$$

Also folgt dieser Teil mit Satz 2.1.5.

□

Korollar 5.5.3. *Für jedes $\alpha \in (0,1)$ ist der Bereichsschätzer $C : \Omega \to \mathcal{B}(\Theta)$ gegeben durch*

$$C = \big(\hat{\zeta} + B_W(\sqrt{d}\, t_{n-d,1-\alpha/(4d)}\hat{s})\big) \times \left[\frac{n-d}{\chi^2_{n-d,1-\alpha/4}}\hat{s}^2, \frac{n-d}{\chi^2_{n-d,\alpha/4}}\hat{s}^2\right]$$

ein Konfidenzbereich für (ζ, σ^2) zum Niveau α.

Beweis. Folgt aus den Sätzen 2.1.3 und 5.5.1, 5.5.2. □

5.6 Hypothesentests

Wir betrachten ein lineares Modell $\mathscr{M}$ mit koordinatenfreier Darstellung. Wir nehmen an, dass $d < n$, wobei $d = \dim(W)$.

Satz 5.6.1 (t-Test für den Erwartungswert)**.** *Es sei $\zeta_0 \in W$ beliebig. Wir zerlegen Θ in*

$$\Theta_0 = \{\zeta_0\} \times (0,\infty) \quad \textit{und} \quad \Theta_1 = (W \setminus \{\zeta_0\}) \times (0,\infty).$$

Wir betrachten also die Null-Hypothese $H_0 = \{\zeta = \zeta_0\}$ und die Alternative $H_1 = \{\zeta \neq \zeta_0\}$. Für jedes $\alpha \in (0,1)$ ist $\varphi = \mathbb{1}_B$ mit

$$B := \big\{\|\hat{\zeta} - \zeta_0\| > \sqrt{d}\, t_{n-d,1-\alpha/(2d)}\hat{s}\big\}$$

ein zulässiger Test zum Irrtumsniveau α.

Beweis. Wir setzen $c := t_{n-d,1-\alpha/(2d)}$. Es sei $b \in \mathbb{R}^n \setminus W^\perp$ beliebig. Weiterhin sei $\vartheta = (\zeta_0, \sigma^2) \in \Theta_0$ beliebig. Nach Satz 5.4.10(d) gilt

$$\frac{\langle b, \hat{\zeta}\rangle - \langle b, \zeta_0\rangle}{\hat{s}\|\Pi_W b\|} \sim t_{n-d} \quad \text{unter } \mathbb{P}_\vartheta.$$

Es folgt

$$\mathbb{P}_\vartheta\big(|\langle b, \hat{\zeta} - \zeta_0\rangle| > c\hat{s}\|\Pi_W b\|\big) = \mathbb{P}_\vartheta\left(\left|\frac{\langle b, \hat{\zeta}\rangle - \langle b, \zeta_0\rangle}{\hat{s}\|\Pi_W b\|}\right| > c\right) \leq \frac{\alpha}{d}.$$

Nun sei $\{v_1, \ldots, v_d\}$ eine Orthonormalbasis von W. Dann folgt

$$\begin{aligned}
G_\varphi(\vartheta) &= \mathbb{E}_\vartheta[\varphi] = \mathbb{E}_\vartheta[\mathbb{1}_B] = \mathbb{P}_\vartheta(B) = \mathbb{P}_\vartheta\big(\|\hat{\zeta} - \zeta_0\| > \sqrt{d}c\hat{s}\big) \\
&= \mathbb{P}_\vartheta\big(\|\hat{\zeta} - \zeta_0\|^2 > dc^2\hat{s}^2\big) = \mathbb{P}_\vartheta\left(\sum_{i=1}^d |\langle v_i, \hat{\zeta} - \zeta_0\rangle|^2 > dc^2\hat{s}^2\right) \\
&\leq \mathbb{P}_\vartheta\left(\bigcup_{i=1}^d \big\{|\langle v_i, \hat{\zeta} - \zeta_0\rangle|^2 > c^2\hat{s}^2\big\}\right) \leq \sum_{i=1}^d \mathbb{P}_\vartheta\big(|\langle v_i, \hat{\zeta} - \zeta_0\rangle| > c\hat{s}\big) \leq d \cdot \frac{\alpha}{d} = \alpha.
\end{aligned}$$

□

Bemerkung 5.6.2. *Satz 5.6.1 folgt auch aus Satz 3.5.1 und Satz 5.5.1(c).*

Bemerkung 5.6.3. *Satz 5.5.1(c) folgt auch aus Satz 3.5.3 und Satz 5.6.1.*

Satz 5.6.4 (χ^2-Test für die Varianz)**.** *Es sei $\sigma_0^2 \in (0,\infty)$ beliebig. Wir zerlegen Θ in*

$$\Theta_0 = W \times (0, \sigma_0^2] \quad \textit{und} \quad \Theta_1 = W \times (\sigma_0^2, \infty).$$

Wir betrachten also die Null-Hypothese $H_0 = \{\sigma^2 \leq \sigma_0^2\}$ und die Alternative $H_1 = \{\sigma^2 > \sigma_0^2\}$. Für jedes $\alpha \in (0,1)$ ist $\varphi = \mathbb{1}_B$ mit

$$B := \{(n-d)\hat{s}^2 > \sigma_0^2\chi^2_{n-d,1-\alpha}\}$$

ein zulässiger Test zum Irrtumsniveau α.

Beweis. Wir setzen $c := \chi^2_{n-d,1-\alpha}$. Weiterhin sei $\vartheta = (\zeta, \sigma^2) \in \Theta_0$ beliebig. Es gilt also $\sigma^2 \leq \sigma_0^2$. Nach Satz 5.4.10(c) gilt

$$\frac{n-d}{\sigma^2}\hat{s}^2 \sim \chi^2_{n-d} \quad \text{unter } \mathbb{P}_\vartheta,$$

und es folgt

$$\begin{aligned} G_\varphi(\vartheta) &= \mathbb{E}_\vartheta[\varphi] = \mathbb{E}_\vartheta[\mathbb{1}_B] = \mathbb{P}_\vartheta(B) \\ &= \mathbb{P}_\vartheta((n-d)\hat{s}^2 > \sigma_0^2 c) \leq \mathbb{P}_\vartheta((n-d)\hat{s}^2 > \sigma^2 c) = \mathbb{P}_\vartheta\left(\frac{n-d}{\sigma^2}\hat{s}^2 > c\right) = \alpha. \end{aligned}$$

□

Nun sei $W_0 \subset W$ ein weiterer Unterraum mit $e < d$, wobei $e = \dim(W_0)$. Wir zerlegen $\Theta = W \times (0, \infty)$ in

$$\Theta_0 = W_0 \times (0, \infty) \quad \text{und} \quad \Theta_1 = (W \setminus W_0) \times (0, \infty).$$

Wir haben also die Null-Hypothese $H_0 = \{\zeta \in W_0\}$ und die Alternative $H_1 = \{\zeta \notin W_0\}$.

Beispiel 5.6.5. *Bei der einfachen linearen Regression (Beispiel 5.2.3) mit*

$$W = \operatorname{lin}\{\mathbb{1}, y\}$$

setzen wir

$$W_0 = \operatorname{lin}\{\mathbb{1}\}.$$

Wir testen also, ob tatsächlich eine nichttriviale lineare Abhängigkeit tatsächlich vorliegt. Genauer sind hier:

- H_0*: Gauß'sches Produktmodell. Es liegt also keine Abhängigkeit zwischen der Stichprobe* X *und den Kontrollvariablen* y *vor.*
- H_1*: Lineares Regressionsmodell, das kein Produktmodell ist. Es liegt also eine Abhängigkeit zwischen der Stichprobe* X *und den Kontrollvariablen* y *vor; und zwar eine lineare.*

Beispiel 5.6.6. *Bei der polynomialen Regression (Beispiel 5.2.5) mit*

$$W = \operatorname{lin}\{\mathbb{1}, y, y^2, \ldots, y^{d-1}\}$$

setzen wir

$$W_0 = \operatorname{lin}\{\mathbb{1}, y, y^2, \ldots, y^{e-1}\}$$

für ein $e < d$. *Wir testen also, ob der Grad des Polynoms reduziert werden kann. Im Fall* $e = 2$ *ist* $W_0 = \operatorname{lin}\{\mathbb{1}, y\}$, *und wir erhalten:*

- *H_0: Lineares Regressionsmodell.*
- *H_1: Polynomiales Regressionsmodell, das nicht linear ist.*

Beispiel 5.6.7. *Bei der mehrfachen linearen Regression (Beispiel 5.2.6) mit*

$$W = \operatorname{lin}\{\mathbb{1}, y_1, \ldots, y_{d-1}\}$$

setzen wir

$$W_0 = \operatorname{lin}\big(\{\mathbb{1}\} \cup \{y_i : i \in I\}\big)$$

mit einer echten Teilmenge $I \subsetneq \{1, \ldots, d-1\}$. Wir testen also, ob die Anzahl der linearen Einflüsse verkleinert werden kann. Wichtige Spezialfälle sind $W_0 = \operatorname{lin}\{\mathbb{1}, y_i\}$ für ein $i = 1, \ldots, d-1$ sowie $W_0 = \operatorname{lin}\{\mathbb{1}\}$.

Beispiel 5.6.8. *Beim Mehrstichprobenproblem (Beispiel 5.2.7) mit*

$$W = \operatorname{lin}\{f_1, \ldots, f_d\},$$

wobei die $f_1, \ldots, f_d$ wie in Beispiel 5.3.10 die „verallgemeinerten Einheitsvektoren” bezeichnen, setzen wir

$$W_0 = \operatorname{lin}\{\mathbb{1}\}.$$

Wir testen also, ob alle Gruppenmittel übereinstimmen. Genauer sind hier:

- *H_0: Alle Gruppenmittel stimmen überein.*
- *H_1: Es gibt (mindestens zwei) Gruppen, deren Gruppenmittel nicht übereinstimmt.*

Definition 5.6.9. *Es seien $X_1, \ldots, X_n \sim \mathrm{N}(\mu_i, 1)$ unabhängige Zufallsvariablen. Die Verteilung von*

$$U := \sum_{i=1}^{n} X_i^2 = \|X\|^2$$

heißt die nichtzentrale χ^2-Verteilung mit n Freiheitsgraden und Nichtzentralitätsparameter

$$\delta^2 := \sum_{i=1}^{n} \mu_i^2 = \|\mu\|^2 \in \mathbb{R}_+.$$

Wir bezeichnen sie mit $\chi_n^2(\delta^2)$.

Bemerkung 5.6.10. *Es gilt* $\chi_n^2 = \chi_n^2(0)$.

Definition 5.6.11. *Es seien* $U \sim \chi_n^2(\delta^2)$ *und* $V \sim \chi_m^2$ *unabhängige Zufallsvariablen. Die Verteilung von*

$$Z := \frac{m}{n}\frac{U}{V}$$

heißt die *nichtzentrale Fischer-Verteilung* *mit* (n, m) *Freiheitsgraden und Nichtzentralitätsparameter* δ^2. *Wir bezeichnen sie mit* $\mathrm{F}_{n,m}(\delta^2)$.

Definition 5.6.12. *Wir nennen* $\mathrm{F}_{n,m} := \mathrm{F}_{n,m}(0)$ *die die* *Fischer-Verteilung* *mit* (n, m) *Freiheitsgraden.*

Satz 5.6.13. *Die Fischer-Verteilung* $\mathrm{F}_{n,m}$ *ist absolutstetig mit Dichte* $f : \mathbb{R} \to \mathbb{R}_+$ *gegeben durch*

$$f(x) = \frac{n^{n/2}m^{m/2}}{B(n/2, m/2)} \frac{x^{\frac{n}{2}-1}}{(m + nx)^{(n+m)/2}} \mathbb{1}_{(0,\infty)}(x),$$

wobei

$$B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r + s)}$$

die Betafunktion bezeichnet.

Wir fixieren ein $\vartheta = (\zeta, \sigma^2) \in \Theta$.

Definition 5.6.14.

(a) *Wir setzen* $\zeta_0 := \Pi_{W_0}\zeta$ *und* $\hat{\zeta}_0 := \Pi_{W_0}X$.

(b) *Wir setzen*

$$\hat{s}_0^2 := \frac{1}{d - e}\|\hat{\zeta} - \hat{\zeta}_0\|^2.$$

(c) *Wir definieren die Quotienten-Statistik*

$$V_n := \frac{\hat{s}_0^2}{\hat{s}^2} = \frac{n - d}{d - e}\frac{\|\hat{\zeta} - \hat{\zeta}_0\|^2}{\|X - \hat{\zeta}\|^2}.$$

Definition 5.6.15. *Für zwei Unterräume* $U \subset V \subset \mathbb{R}^n$ *mit* $\dim(U) = e$ *und* $\dim(V) = d$ *setzen wir*

$$\hat{s}_{V,U}^2 := \frac{1}{d - e}\|\Pi_V X - \Pi_U X\|^2.$$

Bemerkung 5.6.16. *Es gilt* $\hat{s}^2 = \hat{s}^2_{\mathbb{R}^n,W}$ *und* $\hat{s}^2_0 = \hat{s}^2_{W,W_0}$*, und daher*

$$V_n = \frac{\hat{s}^2_{W,W_0}}{\hat{s}^2_{\mathbb{R}^n,W}}.$$

Satz 5.6.17 (Pythagoras)**.** *Es gilt die Zerlegung*

$$\begin{aligned}(n-e)\hat{s}^2_{\mathbb{R}^n,W_0} &= (n-d)\hat{s}^2_{\mathbb{R}^n,W} + (d-e)\hat{s}^2_{W,W_0} \\ &= (n-d)\hat{s}^2 + (d-e)\hat{s}^2_0.\end{aligned}$$

Beweis. Es gilt

$$\begin{aligned}(n-e)\hat{s}^2_{\mathbb{R}^n,W_0} &= \|X - \Pi_{W_0} X \mathbb{1}\|^2 = \|X - \Pi_W X\|^2 + \|\Pi_W X - \Pi_{W_0} X \mathbb{1}\|^2 \\ &= (n-d)\hat{s}^2_{\mathbb{R}^n,W} + (d-e)\hat{s}^2_{W_0,W}.\end{aligned}$$

□

Wir setzen

$$\delta^2 := \frac{\|\zeta - \zeta_0\|^2}{\sigma^2}.$$

Satz 5.6.18. *Die Schätzer* $\hat{s}^2$ *und* $\hat{s}^2_0$ *sind unabhängig mit*

$$\frac{n-d}{\sigma^2}\hat{s}^2 \sim \chi^2_{n-d} \quad \textit{und} \quad \frac{d-e}{\sigma^2}\hat{s}^2_0 \sim \chi^2_{d-e}(\delta^2).$$

Folglich gilt $V_n \sim \mathrm{F}_{d-e,n-d}(\delta^2)$*.*

Beweis. Nach Satz 5.4.10(c) gilt

$$\frac{n-d}{\sigma^2}\hat{s}^2 \sim \chi^2_{n-d}.$$

Es sei $\{v_1, \dots, v_n\}$ eine Orthonormalbasis von $\mathbb{R}^n$, so dass $\{v_1, \dots, v_e\}$ eine Orthonormalbasis von W_0 ist, und $\{v_1, \dots, v_d\}$ eine Orthonormalbasis von W ist. Wir setzen $Y_i := \langle X, v_i \rangle$ und $Z_i := \frac{Y_i}{\sigma}$ für $i = 1, \dots, n$. Dann sind die Zufallsvariablen $Z_1, \dots, Z_n$ unabhängig mit $Z_i \sim \mathrm{N}(\mu_i, 1)$ für alle $i = 1, \dots, n$, wobei

$$\mu_i := \frac{\langle \zeta, v_i \rangle}{\sigma}, \quad i = 1, \dots, n.$$

Es gilt

$$\delta^2 = \frac{\|\zeta - \zeta_0\|^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=e+1}^{d} |\langle \zeta, v_i \rangle|^2 = \sum_{i=e+1}^{d} \mu_i^2 = \|(\mu_{e+1}, \dots, \mu_d)\|^2,$$

und daher folgt

$$\frac{d-e}{\sigma^2}\hat{s}_0^2 = \frac{1}{\sigma^2}\|\hat{\zeta} - \hat{\zeta}_0\|^2 = \frac{1}{\sigma^2}\sum_{i=e+1}^{d} Y_i^2 = \sum_{i=e+1}^{d} Z_i^2 \sim \chi^2_{d-e}(\delta^2).$$

Nach Lemma 5.4.5 gilt außerdem

$$\frac{n-d}{\sigma^2}\hat{s}^2 = \sum_{i=d+1}^{n} Z_i^2 \sim \chi^2_{n-d},$$

so dass die Unabhängigkeit von $\hat{s}^2$ und $\hat{s}_0^2$ folgt. □

Satz 5.6.19. *Für alle $\vartheta = (\zeta, \sigma^2) \in \Theta$ gilt*

$$\mathbb{E}_\vartheta[\hat{s}^2_{\mathbb{R}^n, W_0}] = \sigma^2 + \frac{\|\zeta - \zeta_0\|^2}{n-e}.$$

Beweis. Nach Satz 5.6.17 gilt

$$\hat{s}^2_{\mathbb{R}^n, W_0} = \frac{n-d}{n-e}\hat{s}^2 + \frac{d-e}{n-e}\hat{s}_0^2.$$

Für $X \sim \chi_n(\delta^2)$ gilt $\mathbb{E}[X] = n + \delta^2$. Also folgt mit Satz 5.6.18

$$\begin{aligned}
\mathbb{E}_\vartheta[\hat{s}^2_{\mathbb{R}^n, W_0}] &= \frac{n-d}{n-e}\sigma^2 + \frac{\sigma^2}{n-e}\mathbb{E}_\vartheta\left[\frac{d-e}{\sigma^2}\hat{s}_0^2\right] \\
&= \frac{n-d}{n-e}\sigma^2 + \frac{\sigma^2}{n-e}(d-e+\delta^2) \\
&= \sigma^2 + \frac{\|\zeta - \zeta_0\|^2}{n-e}.
\end{aligned}$$

□

Korollar 5.6.20. *Für jedes $\vartheta = (\zeta, \sigma^2) \in \Theta$ sind folgenden Aussagen äquivalent:*

(i) *Es gilt* $\mathbb{E}_\vartheta[\hat{s}^2_{\mathbb{R}^n, W_0}] = \sigma^2$.

(ii) *Es gilt* $\vartheta \in \Theta_0$, *das heißt* $\zeta \in W_0$.

(iii) *Es gilt* $\delta^2 = 0$.

Beweis. Folgt aus Satz 5.6.19. □

Korollar 5.6.21. *Für alle $\vartheta \in \Theta_0$ gilt $V_n \sim \mathrm{F}_{d-e,n-d}$ unter $\mathbb{P}_\vartheta$.*

Beweis. Folgt aus Satz 5.6.18 und Korollar 5.6.20. □

Satz 5.6.22 (F-Test der linearen Hypothese)**.** *Für jedes $\alpha \in (0,1)$ ist*

$$\varphi := \mathbb{1}_{\{V_n \geq \mathrm{F}_{d-e,n-d,1-\alpha}\}} = \mathbb{1}_{\{\hat{s}_0^2 \geq \mathrm{F}_{d-e,n-d,1-\alpha}\hat{s}^2\}}$$

ein Level-α-Test.

Beweis. Für jedes $\vartheta \in \Theta_0$ gilt nach Korollar 5.6.21

$$G_\varphi(\vartheta) = \mathbb{E}_\vartheta[\varphi] = \mathbb{P}_\vartheta(V_n \geq \mathrm{F}_{d-e,n-d,1-\alpha}) = \alpha.$$

□

Bemerkung 5.6.23. *Große Werte von $\hat{s}_0^2 = \hat{s}_{W,W_0}^2$ führen zur Ablehung der Nullhypothese, die besagt, dass ζ in W_0 liegt.*

5.7 Varianzanalyse

Wir betrachten das Mehrstichprobenproblem. Bei der koordinatengebundenen Darstellung haben wir die Matrix

$$A = \begin{pmatrix} \mathbb{1}_{n_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \mathbb{1}_{n_d} \end{pmatrix} = \begin{pmatrix} f_1 & \dots & f_d \end{pmatrix} \in \mathbb{R}^{n \times d},$$

und bei der koordinatenfreien Darstellung haben wir den Unterraum

$$W = \mathrm{lin}\{f_1, \dots, f_d\} \subset \mathbb{R}^n.$$

In diesem Modell liegen bei der Stichprobe verschiedene Gruppen $X = (X_1, \dots, X_d)$ mit $X_i \in \mathbb{R}^{n_i}$ für $i = 1, \dots, d$ vor, wobei $n_1 + \ldots + n_d = n$.

Satz 5.7.1.

(a) Der Schätzer $\hat{\zeta}$ ist gegeben durch

$$\hat{\zeta} = \sum_{i=1}^{d} \bar{X}_i f_i.$$

(b) Der KQS $\hat{\beta}$ ist gegeben durch

$$\hat{\beta} = (\bar{X}_1, \dots, \bar{X}_d).$$

Beweis. Siehe Beispiel 5.3.10. □

Es sei $W_0 = \text{lin}\{\mathbb{1}\}$. Wir möchten also testen, ob die Gruppenmittel übereinstimmen.

Bemerkung 5.7.2. *Es gilt* $\Pi_{W_0} X = \bar{X}\mathbb{1}$*; siehe Lemma 5.4.11. Folglich gilt*

$$\Pi_{\text{lin}\{f_i\}} X_i = \bar{X}_i f_i \quad \text{für alle } i = 1, \dots, d.$$

Definition 5.7.3. *Wir nennen* $\hat{s}^2$ *die Stichprobenvarianz innerhalb der Gruppen.*

Definition 5.7.4. *Für* $i = 1, \dots, d$ *nennen wir*

$$s_{X_i}^2 = \frac{1}{n_i - 1} \|X_i - \bar{X}_i f_i\|^2$$

die Stichprobenvarianz der i-ten Gruppe.

Satz 5.7.5. *Es gilt die Zerlegung*

$$(n-d)\hat{s}^2 = \sum_{i=1}^{d} (n_i - 1) s_{X_i}^2.$$

Beweis. Wir setzen $W_i := \text{lin}\{f_i\}$ für $i = 1, \dots, d$. Dann gilt

$$W = W_1 \oplus \dots \oplus W_d.$$

Außerdem seien $V_1, \dots, V_d$ die eindeutig bestimmten Unterräume des $\mathbb{R}^n$, die von Einheitsvektoren erzeugt werden, so dass $f_i \in V_i$ für alle $i = 1, \dots, d$. Dann gilt

$$\mathbb{R}^n = V_1 \oplus \dots \oplus V_d,$$

und es folgt mit Bemerkung 5.7.2

$$\begin{aligned}
(n-d)\hat{s}^2 = \|X - \Pi_W X\|^2 &= \left\| \sum_{i=1}^{d} \left(\Pi_{V_i} X - \Pi_{W_i} X \right) \right\|^2 \\
&= \sum_{i=1}^{d} \|\Pi_{V_i} X - \Pi_{W_i} X\|^2 = \sum_{i=1}^{d} (n_i - 1) \hat{s}_{V_i, W_i}^2 = \sum_{i=1}^{d} (n_i - 1) s_{X_i}^2.
\end{aligned}$$

□

Definition 5.7.6. *Wir nennen*

$$s_X^2 = \frac{1}{n-1} \|X - \bar{X}\mathbb{1}\|^2$$

die totale empirische Varianz (oder Gesamtvarianz).

Definition 5.7.7. *Wir nennen*

$$\hat{s}_0^2 = \hat{s}_{W,W_0}^2 = \frac{1}{d-1}\sum_{i=1}^{d} \|\bar{X}_i f_i - \bar{X} f_i\|^2$$

die Stichprobenvarianz zwischen den Gruppen.

Satz 5.7.8. *Es gilt die Zerlegung*

$$(n-1)s_X^2 = (n-d)\hat{s}^2 + (d-1)\hat{s}_0^2.$$

Beweis. Folgt aus Satz 5.6.17. □

Satz 5.7.9. *Es gilt für alle* $\vartheta = (\zeta, \sigma^2) \in \Theta$

$$\mathbb{E}_\vartheta[s_X^2] = \sigma^2 + \frac{\|\zeta - \zeta_0\|^2}{n-1}.$$

Beweis. Folgt aus Satz 5.6.19. □

Korollar 5.7.10. *Für jedes* $\vartheta = (\zeta, \sigma^2) \in \Theta$ *sind folgenden Aussagen äquivalent:*

(i) *Es gilt* $\mathbb{E}_\vartheta[s_X^2] = \sigma^2$.

(ii) *Es gilt* $\vartheta \in \Theta_0$, *das heißt* $\zeta \in W_0$.

(iii) *Es gilt* $\delta^2 = 0$.

Beweis. Folgt aus Korollar 5.6.20. □

Korollar 5.7.11. *Für jedes* $\alpha \in (0,1)$ *ist*

$$\varphi := \mathbb{1}_{\{\hat{s}_0^2 \geq \mathrm{F}_{d-e,n-d,1-\alpha}\hat{s}^2\}}$$

ein Level-α*-Test.*

Beweis. Folgt aus Satz 5.6.22. □

Bemerkung 5.7.12. *Je größer die Stichprobenvarianz* $\hat{s}_0^2$ *zwischen den Gruppen, desto eher wird* H_0 *abgelehnt.*

Literaturverzeichnis

[Bib16] Bibinger, M.: *Mathematische Statistik*. 2016. – Vorlesungsskript aus dem SS 2016, Philipps-Universität Marburg

[CS11] Czado, C. ; Schmidt, T.: *Mathematische Statistik*. Springer-Verlag, Berlin, 2011

[Geo04] Georgii, H. O.: *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Walter de Gruyter & Co., Berlin, 2004

[Grü10] Grübel, R.: *Stochastik*. 2010. – Vorlesungsskript, Leibniz Universität Hannover

[JP04] Jacod, J. ; Protter, P.: *Probability Essentials*. Berlin : Springer-Verlag, 2004 (Universitext)

[Kre05] Krengel, U.: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Vieweg-Verlag, Wiesbaden, 2005

[Löw] Löwe, M.: *Mathematische Statistik I*. – Vorlesungsskript, Universität Münster

[Pru00] Pruscha, H.: *Vorlesungen über Mathematische Statistik*. Teubner-Verlag, Stuttgart, 2000

[Ren12] Renesse, M.: *Mathematische Statistik*. 2012. – Vorlesungsskript aus dem SS 2012, Universität Leipzig, aufgezeichnet von Tobias Weihrauch

[Rüs14] Rüschendorf, L.: *Mathematische Statistik*. Springer-Verlag, Berlin, 2014

[Sha03] Shao, J.: *Mathematical Statistics*. Springer-Verlag, New York, 2003

[Tap13] Tappe, S.: *Einführung in die Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin, 2013