Mathematische Statistik
Gliederung zur Vorlesung im Wintersemester 2006/07

Markus Reiß
Universität Heidelberg
reiss@statlab.uni-heidelberg.de

VORLÄUFIGE FASSUNG: 9. Februar 2007

Inhaltsverzeichnis

1 Einführende Beispiele

- Modellierung
- Modelldiagnostik (QQ-Plot, Boxplot, empirische Korrelation)
- Median, Mittelwert, Ausreißer
- Konfidenzintervall
- Hypothesentest
- Klassifikation
- Vorhersage

2 Entscheidungstheorie

2.1 Formalisierung eines statistischen Problems

2.1 Definition. *Ein Messraum $(\mathcal{X}, \mathscr{F})$ versehen mit einer Familie $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ von Wahrscheinlichkeitsmaßen, $\Theta \neq \varnothing$ beliebige Parametermenge, heißt statistisches Experiment. Jede $(\mathscr{F}, \mathscr{S})$-messbare Funktion $Y : \mathcal{X} \to S$ heißt Beobachtung oder Statistik mit Werten in $(S, \mathscr{S})$ und induziert das statistische Experiment $(S, \mathscr{S}, (\mathbb{P}^Y_\vartheta)_{\vartheta\in\Theta})$. Sind die Beobachtungen $Y_1, \ldots, Y_n$ für jedes $\mathbb{P}_\vartheta$ unabhängig und identisch verteilt, so nennt man $Y_1, \ldots, Y_n$ eine mathematische Stichprobe.*

2.2 Definition. *Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Experiment. Eine Entscheidungsregel ist eine messbare Abbildung $\rho : \mathcal{X} \to A$, wobei der Messraum $(A, \mathscr{A})$ der sogenannte Aktionsraum ist. Jede Funktion $l : \Theta \times A \to [0, \infty) =: \mathbb{R}^+$, die messbar im zweiten Argument ist, heißt Verlustfunktion. Das Risiko einer Entscheidungsregel ρ bei Vorliegen des Parameters $\vartheta \in \Theta$ ist*

$$R(\vartheta, \rho) := \mathbb{E}_\vartheta[l(\vartheta, \rho)] = \int_{\mathcal{X}} l(\vartheta, \rho(x))\, \mathbb{P}_\vartheta(dx).$$

2.3 Definition. *Die Entscheidungsregel ρ heißt besser als eine Entscheidungsregel ρ', falls $R(\vartheta, \rho) \leqslant R(\vartheta, \rho')$ für alle $\vartheta \in \Theta$ gilt und falls ein $\vartheta_0 \in \Theta$ mit $R(\vartheta_0, \rho) < R(\vartheta_0, \rho')$ existiert. Eine Entscheidungsregel heißt zulässig, wenn es keine bessere Entscheidungsregel gibt.*

2.2 Minimax- und Bayes-Ansatz

2.4 Definition. *Eine Entscheidungsregel ρ heißt minimax, falls*

$$\sup_{\vartheta\in\Theta} R(\vartheta, \rho) = \inf_{\rho'} \sup_{\vartheta\in\Theta} R(\vartheta, \rho'),$$

wobei sich das Infimum über alle Entscheidungsregeln ρ' erstreckt.

2.5 Definition. *Der Parameterraum Θ trage die σ-Algebra $\mathscr{F}_\Theta$, die Verlustfunktion l sei produktmessbar und $\vartheta \mapsto \mathbb{P}_\vartheta(B)$ sei messbar für alle $B \in \mathscr{F}$. Die a priori-Verteilung π des Parameters ϑ ist gegeben durch ein Wahrscheinlichkeitsmaß auf $(\Theta, \mathscr{F}_\Theta)$. Das zu π assoziierte Bayesrisiko einer Entscheidungsregel ρ ist*

$$R_\pi(\rho) := \mathbb{E}_\pi[R(\vartheta,\rho)] = \int_\Theta R(\vartheta,\rho)\,\pi(d\vartheta) = \int_\Theta \int_{\mathcal{X}} l(\vartheta,\rho(x))\;\mathbb{P}_\vartheta(dx)\,\pi(d\vartheta).$$

ρ heißt Bayesregel oder Bayes-optimal (bezüglich π), falls

$$R_\pi(\rho) = \inf_{\rho'} R_\pi(\rho')$$

gilt, wobei sich das Infimum über alle Entscheidungsregeln ρ' erstreckt.

2.6 Satz. *Es liege die Situation aus der vorangegangenen Definition vor.*

(a) Für jede Entscheidungsregel ρ gilt

$$\sup_{\vartheta\in\Theta} R(\vartheta,\rho) = \sup_\pi R_\pi(\rho),$$

wobei sich das zweite Supremum über alle a priori-Verteilungen π erstreckt. Insbesondere ist das Risiko einer Bayesregel stets kleiner oder gleich dem Minimaxrisiko.

(b) Für eine Minimaxregel ρ gilt $\sup_\pi R_\pi(\rho) = \inf_{\rho'} \sup_\pi R_\pi(\rho')$.

2.7 Definition. *Definiere $\Omega := \mathcal{X} \times \Theta$ und $\tilde{\mathbb{P}}$ auf $(\Omega, \mathscr{F} \otimes \mathscr{F}_\Theta)$ gemäß $\tilde{\mathbb{P}}(dx, d\vartheta) = P_\vartheta(dx)\pi(d\vartheta)$ (gemeinsame Verteilung von Beobachtung und Parameter). Bezeichne mit X und $\bar{\vartheta}$ die Koordinatenprojektionen von Ω auf $\mathcal{X}$ bzw. Θ.*

2.8 Satz. *Eine Regel ρ ist Bayes-optimal, falls für $\tilde{\mathbb{P}}$-f.a. $x \in \mathcal{X}$ gilt*

$$\rho(x) = \operatorname{argmin}_{a\in A} \mathbb{E}_{\tilde{\mathbb{P}}}[l(\bar{\vartheta}, a) \,|\, X = x].$$

2.9 Korollar. *Für $\Theta \subseteq \mathbb{R}$, $A = \mathbb{R}$ und quadratisches Risiko (d.h. $l(\vartheta, a) = (a-\vartheta)^2$) ist die bedingte Erwartung $\hat{\vartheta}_\pi := \mathbb{E}_{\tilde{\mathbb{P}}}[\bar{\vartheta} \,|\, X = x]$ Bayes-optimaler Schätzer von ϑ bezüglich der a priori-Verteilung π.*

2.10 Definition. *Es sei X eine $(S, \mathscr{S})$-wertige Zufallsvariable auf $(\Omega, \mathscr{F}, \mathbb{P})$. Eine Abbildung $K : S \times \mathscr{F} \to [0, 1]$ heißt reguläre bedingte Wahrscheinlichkeit oder Markovkern bezüglich X, falls*

(a) $A \mapsto K(x, A)$ ist Wahrscheinlichkeitsmaß für alle $x \in S$;

(b) $x \mapsto K(x, A)$ ist messbar für alle $A \in \mathscr{F}$;

(c) $K(X, A) = \mathbb{P}(A \,|\, X) := \mathbb{E}[\mathbf{1}_A \,|\, X]$ $\mathbb{P}$-f.s. für alle $A \in \mathscr{F}$.

2.11 Satz. *Es sei* (Ω, d) *ein vollständiger, separabler Raum mit Metrik* d *und Borel-*σ*-Algebra* $\mathscr{F}$ *(polnischer Raum). Für jede Zufallsvariable* X *auf* $(\Omega, \mathscr{F}, \mathbb{P})$ *existiert eine reguläre bedingte Wahrscheinlichkeit* K *bezüglich* X*.* K *ist* $\mathbb{P}$*-f.s. eindeutig bestimmt, d.h. für eine zweite solche reguläre bedingte Wahrscheinlichkeit* K' *gilt*

$$\mathbb{P}(\forall A \in \mathscr{F}: \ K(X, A) = K'(X, A)) = 1.$$

2.12 Definition. *Die Verteilung von* $\bar{\vartheta}$ *unter der regulären bedingten Wahrscheinlichkeit* $\tilde{\mathbb{P}}(\bullet \,|\, X = x)$ *von* $\tilde{\mathbb{P}}$ *heißt a posteriori-Verteilung des Parameters gegeben die Beobachtung* $X = x$*.*

2.13 Satz. *Für jede Entscheidungsregel* ρ *gilt:*

(a) *Ist* ρ *minimax und eindeutig in dem Sinn, dass jede andere Minimax-Regel die gleiche Risikofunktion besitzt, so ist* ρ *zulässig.*

(b) *Ist* ρ *zulässig mit konstanter Risikofunktion, so ist* ρ *minimax.*

(c) *Ist* ρ *eine Bayesregel (bzgl.* π*) und eindeutig in dem Sinn, dass jede andere Bayesregel (bzgl.* π*) die gleiche Risikofunktion besitzt, so ist* ρ *zulässig.*

(d) *Die Parametermenge* Θ *bilde einen metrischen Raum mit Borel-*σ*-Algebra* $\mathscr{F}_\Theta$*. Ist* ρ *eine Bayesregel (bzgl.* π*), so ist* ρ *zulässig, falls (i)* $R_\pi(\rho) < \infty$*; (ii) für jede nichtleere offene Menge* U *in* Θ *gilt* $\pi(U) > 0$*; (iii) für jede Regel* ρ' *ist* $\vartheta \mapsto R(\vartheta, \rho')$ *stetig.*

2.14 Korollar. *Es sei* $X_1, \dots, X_n$ *eine* $N(\mu, 1)$*-verteilte mathematische Stichprobe mit* $\mu \in \mathbb{R}$ *unbekannt. Bezüglich quadratischem Risiko ist das arithmetische Mittel* $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ *zulässig und minimax als Schätzer von* μ*.*

2.15 Definition. *Eine Verteilung* π *auf* $(\Theta, \mathscr{F}_\Theta)$ *heißt ungünstigste a priori-Verteilung zu einer gegebenen Verlustfunktion, falls*

$$\inf_\rho R_\pi(\rho) = \sup_{\pi'} \inf_\rho R_{\pi'}(\rho).$$

2.16 Lemma. *Gilt* $R_\pi(\rho_\pi) = \sup_{\vartheta \in \Theta} R(\vartheta, \rho_\pi)$ *für eine a priori-Verteilung* π *und ihre zugehörige Bayesregel* ρ_π*, so folgt die Sattelpunktseigenschaft*

$$\forall \pi' \, \forall \rho' : \ R_{\pi'}(\rho_\pi) \leqslant R_\pi(\rho_\pi) \leqslant R_\pi(\rho').$$

Weiterhin ist ρ_π *minimax und* π *ungünstigste a priori-Verteilung.*

2.3 Das Stein-Phänomen

2.17 Lemma. *Es sei* $f : \mathbb{R}^d \to \mathbb{R}$ *eine Funktion, die Lebesgue-f.ü. absolut stetig in jeder Koordinate ist. Dann gilt für* $Y \sim N(\mu, \sigma^2 E_d)$ *mit* $\mu \in \mathbb{R}^d$*,* $\sigma > 0$*,* $E_d = \operatorname{diag}(1, \dots, 1) \in \mathbb{R}^{d \times d}$ *und für alle* $i = 1, \dots, d$

$$\mathbb{E}[(\mu_i - Y_i) f(Y)] = -\sigma^2 \, \mathbb{E}[\tfrac{\partial f}{\partial x_i}(Y)],$$

sofern $\mathbb{E}[|\frac{\partial f}{\partial x_i}(Y)|] < \infty$*.*

2.18 Satz. *Es sei $d \geqslant 3$ und $Y_1, \dots, Y_n$ eine $N(\mu, E_d)$-verteilte mathematische Stichprobe mit $\mu \in \mathbb{R}^d$ unbekannt. Dann gilt für den James-Stein-Schätzer*

$$\hat{\mu}_{JS} := \Big(1 - \frac{d-2}{n|\bar{Y}|^2}\Big)\bar{Y}$$

mit $\bar{Y} := \frac{1}{n}\sum_{i=1}^n Y_i$, dass

$$\mathbb{E}_\mu[|\hat{\mu}_{JS} - \mu|^2] = \frac{d}{n} - \mathbb{E}_\mu\Big[\frac{(d-2)^2}{n^2|\bar{Y}|^2}\Big] < \frac{d}{n} = \mathbb{E}_\mu[|\bar{Y} - \mu|^2].$$

Insbesondere ist $\bar{Y}$ bei quadratischem Risiko kein zulässiger Schätzer von μ im Fall $d \geqslant 3$!

2.19 Satz. *Es sei $d \geqslant 3$ und $Y_1, \dots, Y_n$ eine $N(\mu, E_d)$-verteilte mathematische Stichprobe mit $\mu \in \mathbb{R}^d$ unbekannt. Dann ist der James-Stein-Schätzer mit positivem Gewicht*

$$\hat{\mu}_{JS+} := \Big(1 - \frac{d-2}{n|\bar{Y}|^2}\Big)_+\bar{Y}, \qquad x_+ := \max(x, 0)$$

bei quadratischem Risiko besser als der James-Stein-Schätzer $\hat{\mu}_{JS}$.

2.4 Ergänzungen

2.20 Definition. *Zu vorgegebener Verlustfunktion l heißt eine Entscheidungsregel ρ <u>unverzerrt</u>, falls*

$$\forall \vartheta, \vartheta' \in \Theta: \ \mathbb{E}_\vartheta[l(\vartheta', \rho)] \geqslant \mathbb{E}_\vartheta[l(\vartheta, \rho)] =: R(\vartheta, \rho).$$

2.21 Lemma. *Es seien $g: \Theta \to A \subseteq \mathbb{R}$ und $l(\vartheta, \rho) = (\rho - g(\vartheta))^2$ der quadratische Verlust. Dann ist eine Entscheidungsregel (ein Schätzer von $g(\vartheta)$) $\hat{g}: \mathcal{X} \to A$ mit $\mathbb{E}_\vartheta[\hat{g}^2] < \infty$ und $\mathbb{E}_\vartheta[\hat{g}] \in g(\Theta)$ für alle $\vartheta \in \Theta$ genau dann unverzerrt, wenn sie <u>erwartungstreu</u> ist, d.h. $\mathbb{E}_\vartheta[\hat{g}] = g(\vartheta)$ für alle $\vartheta \in \Theta$ gilt.*

2.22 Lemma. *Es sei $\Theta = \Theta_0 \dot\cup \Theta_1$, $A = [0,1]$. Für den Verlust $l(\vartheta, a) = l_0 a \mathbf{1}_{\Theta_0}(\vartheta) + l_1(1-a)\mathbf{1}_{\Theta_1}(\vartheta)$ ist eine Entscheidungsregel ρ (ein randomisierter Test von $H_0: \vartheta \in \Theta_0$ gegen $H_1: \vartheta \in \Theta_1$) genau dann unverzerrt, wenn sie zum Niveau $\alpha := \frac{l_1}{l_0 + l_1}$ <u>unverfälscht</u> ist, d.h.*

$$\forall \vartheta \in \Theta_0: \ \mathbb{E}_\vartheta[\rho] \leqslant \alpha, \quad \forall \vartheta \in \Theta_1: \ \mathbb{E}_\vartheta[\rho] \geqslant \alpha.$$

2.23 Definition. *Ein <u>Entscheidungskern</u> oder <u>randomisierte Entscheidungsregel</u> $\rho: \mathcal{X} \times \mathscr{A} \to [0,1]$ ist eine reguläre bedingte Wahrscheinlichkeit auf dem Aktionsraum $(A, \mathscr{A})$ mit der Interpretation, dass bei Vorliegen der Beobachtung x gemäß $\rho(x, \bullet)$ eine Entscheidung zufällig ausgewählt wird. Das zugehörige Risiko ist*

$$R(\vartheta, \rho) := \mathbb{E}_\vartheta\Big[\int_A l(\vartheta, a)\,\rho(da)\Big] = \int_{\mathcal{X}}\int_A l(\vartheta, a)\rho(x, da)\,\mathbb{P}_\vartheta(dx).$$

2.24 Lemma. *Es sei $A \subseteq \mathbb{R}^d$ konvex sowie $l(\vartheta, a)$ eine im zweiten Argument konvexe Verlustfunktion. Dann gibt es zu jeder randomisierten Entscheidungsregel eine deterministische Entscheidungsregel, deren Risiko nicht größer ist.*

3 Dominierte Experimente und Suffizienz

3.1 Dominierte Experimente

3.1 Definition. *Ein statistisches Experiment $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ heißt dominiert (von μ), falls es ein σ-endliches Maß μ auf $\mathscr{F}$ gibt, so dass $\mathbb{P}_\vartheta$ absolutstetig bezüglich μ ist ($\mathbb{P}_\vartheta \ll \mu$) für alle $\vartheta \in \Theta$. Die durch ϑ parametrisierte Radon-Nikodym-Dichte*

$$L(\vartheta, x) := \frac{d\,\mathbb{P}_\vartheta}{d\mu}(x), \quad \vartheta \in \Theta,\, x \in \mathcal{X},$$

heißt auch Likelihoodfunktion, wobei diese meist als durch x parametrisierte Funktion in ϑ aufgefasst wird.

3.2 Satz. *Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein dominiertes Experiment. Dann gibt es ein Wahrscheinlichkeitsmaß $\mathbb{Q}$ der Form $\mathbb{Q} = \sum_{i=1}^\infty c_i\,\mathbb{P}_{\vartheta_i}$ mit $c_i \geqslant 0$, $\sum_i c_i = 1$, $\vartheta_i \in \Theta$, so dass $\mathbb{P}_\vartheta \ll \mathbb{Q}$ für alle $\vartheta \in \Theta$ gilt.*

3.3 Satz. *Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein dominiertes Experiment mit produktmessbarer Likelihoodfunktion $L(\vartheta, x)$. Zu vorgegebener a priori-Verteilung π hat die a posteriori-Verteilung von $\bar{\vartheta}$ gegeben $X = x$ folgende Dichte bezüglich π:*

$$Z_x^\pi(\vartheta) = \frac{L(\vartheta, x)}{\int_\Theta L(\vartheta', x)\,\pi(d\vartheta')}\mathbf{1}_{\{\int L(\vartheta',x)\pi(d\vartheta')>0\}}, \quad \vartheta \in \Theta \qquad \text{(Bayesformel).}$$

3.2 Exponentialfamilien

3.4 Definition. *Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein von μ dominiertes Experiment. Dann heißt $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ Exponentialfamilie (in $\eta(\vartheta)$ und T), wenn $k \in \mathbb{N}$, $\eta : \Theta \to \mathbb{R}^k$, $C : \Theta \to \mathbb{R}^+$, $T : \mathcal{X} \to \mathbb{R}^k$ messbar und $h : \mathcal{X} \to \mathbb{R}^+$ messbar existieren, so dass*

$$\frac{d\,\mathbb{P}_\vartheta}{d\mu}(x) = C(\vartheta)h(x)\exp(\langle \eta(\vartheta), T(x)\rangle_{\mathbb{R}^k}), \quad x \in \mathcal{X},\ \vartheta \in \Theta.$$

T wird natürliche suffiziente Statistik von $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ genannt. Sind $\eta_1, \ldots, \eta_k$ linear unabhängige Funktionen und gilt für alle $\vartheta \in \Theta$ die Implikation

$$\lambda_0 + \lambda_1 T_1 + \cdots + \lambda_k T_k = 0 \ \mathbb{P}_\vartheta\text{-f.s.} \Rightarrow \lambda_0 = \lambda_1 = \cdots = \lambda_k = 0$$

($1, T_1, \ldots, T_k$ sind $\mathbb{P}_\vartheta$-f.s. linear unabhängig), so heißt die Exponentialfamilie k-parametrisch.

3.5 Definition. *Bildet $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ eine Exponentialfamilie (mit obiger Notation), so heißt*

$$\mathscr{Z} := \Big\{ u \in \mathbb{R}^k \;\Big|\; \int_{\mathcal{X}} e^{\langle u, T(x)\rangle} h(x)\mu(dx) \in (0, \infty) \Big\}$$

ihr natürlicher Parameterraum. Die entsprechend mit $u \in \mathscr{Z}$ parametrisierte Familie wird natürliche Exponentialfamilie in T genannt.

3.6 Lemma. *Bildet $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ eine (k-parametrische) Exponentialfamilie in $\eta(\vartheta)$ und $T(x)$, so bilden auch die Produktmaße $(\mathbb{P}_\vartheta^{\otimes n})_{\vartheta\in\Theta}$ eine (k-parametrische) Exponentialfamilie in $\eta(\vartheta)$ und $\sum_{i=1}^n T(x_i)$ mit*

$$\frac{d\,\mathbb{P}_\vartheta^{\otimes n}}{d\mu^{\otimes n}}(x) = C(\vartheta)^n\Big(\prod_{i=1}^n h(x_i)\Big)\exp(\langle\eta(\vartheta),\textstyle\sum_{i=1}^n T(x_i)\rangle_{\mathbb{R}^k}),\quad x\in\mathcal{X}^n,\ \vartheta\in\Theta.$$

3.7 Satz. *Es sei $(\mathbb{P}_\vartheta)_{\vartheta\in\mathscr{Z}}$ eine Exponentialfamilie mit natürlichem Parameterraum $\mathscr{Z}\subseteq\mathbb{R}^k$ und Darstellung*

$$\frac{d\,\mathbb{P}_\vartheta}{d\mu}(x) = C(\vartheta)h(x)\exp(\langle\vartheta,T(x)\rangle) = h(x)\exp(\langle\vartheta,T(x)\rangle - A(\vartheta)),$$

wobei $A(\vartheta)=\log\big(\int h(x)\exp(\langle\vartheta,T(x)\rangle)\mu(dx)\big)$. Ist $\tilde\vartheta$ ein innerer Punkt von $\mathscr{Z}$, so ist die erzeugende Funktion $\psi_{\tilde\vartheta}(s)=\mathbb{E}_{\tilde\vartheta}[e^{\langle T,s\rangle}]$ in einer Umgebung der Null wohldefiniert und beliebig oft differenzierbar. Es gilt $\psi_{\tilde\vartheta}(s)=\exp(A(\tilde\vartheta+s)-A(\tilde\vartheta))$ für alle s mit $\tilde\vartheta+s\in\mathscr{Z}$.

Für $i,j=1,\dots,k$ folgt $\mathbb{E}_{\tilde\vartheta}[T_i]=\frac{dA}{d\vartheta_i}(\tilde\vartheta)$ und $\operatorname{Cov}_{\tilde\vartheta}(T_i,T_j)=\frac{d^2A}{d\vartheta_i d\vartheta_j}(\tilde\vartheta)$.

3.3 Suffizienz

3.8 Definition. *Eine $(S,\mathscr{S})$-wertige Statistik T auf $(\mathcal{X},\mathscr{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ heißt suffizient (für $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$), falls für jedes $\vartheta\in\Theta$ die reguläre bedingte Wahrscheinlichkeit von $\mathbb{P}_\vartheta$ gegeben T (existiert und) nicht von ϑ abhängt, d.h.*

$$\exists k\ \forall\vartheta\in\Theta,\ B\in\mathscr{F}:\ k(T,B)=\mathbb{P}_\vartheta(B\,|\,T):=\mathbb{E}_\vartheta[\mathbf{1}_B\,|\,T]\quad \mathbb{P}_\vartheta\text{-f.s.}$$

Statt $k(t,B)$ schreiben wir $\mathbb{P}_\bullet(B\,|\,T=t)$ bzw. $\mathbb{E}_\bullet[\mathbf{1}_B\,|\,T=t]$.

3.9 Satz (Faktorisierungskriterium von Neyman). *Es sei $(\mathcal{X},\mathscr{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein von μ dominiertes Experiment mit Likelihoodfunktion L sowie T eine $(S,\mathscr{S})$-wertige Statistik. Dann ist T genau dann suffizient, wenn eine messbare Funktion $h:\mathcal{X}\to\mathbb{R}^+$ existiert, so dass für alle $\vartheta\in\Theta$ eine messbare Funktion $g_\vartheta:S\to\mathbb{R}^+$ existiert mit*

$$L(\vartheta,x)=g_\vartheta(T(x))h(x)\quad\text{für }\mu\text{-f.a. }x\in\mathcal{X}.$$

3.10 Korollar. *Die natürliche suffiziente Statistik einer Exponentialfamilie ist in der Tat suffizient.*

3.11 Satz (Rao-Blackwell). *Es sei $(\mathcal{X},\mathscr{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Experiment, $A\subseteq\mathbb{R}^k$ konvex und $l(\vartheta,a)$ eine im zweiten Argument konvexe Verlustfunktion. Ist T eine für $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ suffiziente Statistik, so gilt für jede Entscheidungsregel ρ die Risikoabschätzung*

$$\forall\vartheta\in\Theta:\ R(\vartheta,\tilde\rho)\leqslant R(\vartheta,\rho)\qquad\text{mit }\tilde\rho:=\mathbb{E}_\bullet[\rho\,|\,T].$$

3.12 Satz. *Es sei $(\mathcal{X},\mathscr{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Experiment und T eine suffiziente Statistik. Dann gibt es zu jedem randomisierten Test φ einen randomisierten Test $\tilde\varphi$, der nur von T abhängt und dieselbe Gütefunktion besitzt, nämlich $\tilde\varphi=\mathbb{E}_\bullet[\varphi\,|\,T]$.*

4 Testtheorie

4.1 Neyman-Pearson-Theorie

4.1 Definition. *Es sei* $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ *ein statistisches Experiment mit Zerlegung* $\Theta = \Theta_0 \dot\cup \Theta_1$. *Jede messbare Funktion* $\varphi : \mathcal{X} \to [0,1]$ *heißt (randomisierter) Test.* φ *besitzt Niveau* $\alpha \in [0,1]$, *falls* $\mathbb{E}_\vartheta[\varphi] \leqslant \alpha$ *für alle* $\vartheta \in \Theta_0$ *gilt. Die Abbildung* $\vartheta \mapsto \mathbb{E}_\vartheta[\varphi]$ *heißt Gütefunktion von* φ. *Ein Test* φ *der Hypothese* $H_0 : \vartheta \in \Theta_0$ *gegen die Alternative* $H_1 : \vartheta \in \Theta_1$ *ist ein gleichmäßig bester Test zum Niveau* α, *falls* φ *Niveau* α *besitzt sowie für alle anderen Tests* φ' *vom Niveau* α *gilt*

$$\forall \vartheta \in \Theta_1 : \ \mathbb{E}_\vartheta[\varphi] \geqslant \mathbb{E}_\vartheta[\varphi'].$$

φ *heißt gleichmäßig bester unverfälschter Test zum Niveau* α, *falls* φ *unverfälscht zum Niveau* α *ist sowie für alle anderen unverfälschten Tests* φ' *zum Niveau* α *obige Ungleichung gilt.*

4.2 Definition. *Es sei* $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ *ein (binäres) statistisches Experiment mit* $\Theta = \{0,1\}$. *Bezeichnet* p_i, $i = 1,2$, *die Dichte von* $\mathbb{P}_i$ *bezüglich* $\mathbb{P}_0 + \mathbb{P}_1$, *so heißt ein Test der Form*

$$\varphi(x) = \begin{cases} 1, & \text{falls } p_1(x) > kp_0(x) \\ 0, & \text{falls } p_1(x) < kp_0(x) \\ \gamma(x), & \text{falls } p_1(x) = kp_0(x) \end{cases}$$

mit $k \in \mathbb{R}^+$ *und* $\gamma(x) \in [0,1]$ *Neyman-Pearson-Test.*

4.3 Satz (Neyman-Pearson-Lemma)**.**

(a) *Jeder Neyman-Pearson-Test* φ *ist ein (gleichmäßig) bester Test für* $H_0 : \vartheta = 0$ *gegen* $H_1 : \vartheta = 1$ *zum Niveau* $\mathbb{E}_0[\varphi]$.

(b) *Für jedes vorgegebene* $\alpha \in (0,1)$ *gibt es einen Neyman-Pearson-Test zum Niveau* α *mit* $\gamma(x) = \gamma \in [0,1]$ *konstant.*

4.4 Definition. *Es seien* $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ *ein dominiertes Experiment mit* $\Theta \subseteq \mathbb{R}$ *und Likelihoodfunktion* $L(\vartheta, x)$ *sowie* T *eine reellwertige Statistik. Dann hat die Familie* $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ *monotonen Dichtequotienten (oder monotonen Likelihoodquotienten) in* T, *falls*

(a) $\vartheta \neq \vartheta' \Rightarrow \mathbb{P}_\vartheta \neq \mathbb{P}_{\vartheta'}$;

(b) *Für alle* $\vartheta < \vartheta'$ *gibt es eine monoton wachsende Funktion* $h(\bullet, \vartheta, \vartheta') : \mathbb{R} \to \mathbb{R}^+ \cup \{+\infty\}$ *mit (Konvention* $a/0 := +\infty$ *für* $a > 0$*)*

$$\frac{L(\vartheta, x)}{L(\vartheta', x)} = h(T(x), \vartheta, \vartheta') \quad \text{für } (\mathbb{P}_\vartheta + \mathbb{P}_{\vartheta'})\text{-f.a. } x \in \mathcal{X}.$$

4.5 Satz. *Ist* $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ *mit* $\Theta \subseteq \mathbb{R}$ *eine einparametrische Exponentialfamilie in* $\eta(\vartheta)$ *und* T, *so hat sie monotonen Dichtequotienten, sofern* η *streng monoton wächst.*

4.6 Satz. *Die Familie* $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$, $\Theta \subseteq \mathbb{R}$, *besitze monotonen Dichtequotienten in* T. *Für* $\alpha \in (0,1)$ *und* $\vartheta_0 \in \Theta$ *gilt dann:*

(a) *Unter allen Tests* φ *für das* <u>*einseitige Testproblem*</u> $H_0 : \vartheta \leqslant \vartheta_0$ *gegen* $H_1 : \vartheta > \vartheta_0$ *mit der Eigenschaft* $\mathbb{E}_{\vartheta_0}[\varphi] = \alpha$ *gibt es einen Test* φ^*, *der die Fehlerwahrscheinlichkeiten erster und zweiter Art gleichmäßig minimiert, nämlich*
$$\varphi^*(x) = \begin{cases} 1, & \textit{falls } T(x) > k, \\ 0, & \textit{falls } T(x) < k, \\ \gamma, & \textit{falls } T(x) = k, \end{cases}$$
wobei $k \in \mathbb{R}$, $\gamma \in [0,1]$ *gemäß* $\mathbb{E}_{\vartheta_0}[\varphi^*] = \alpha$ *bestimmt werden.*

(b) *Dieser Test* φ^* *ist gleichmäßig bester Test zum Niveau* α *für* $H_0 : \vartheta \leqslant \vartheta_0$ *gegen* $H_1 : \vartheta > \vartheta_0$.

(c) *Für alle* $\vartheta < \vartheta'$ *gilt* $\mathbb{E}_\vartheta[\varphi^*] \leqslant \mathbb{E}_{\vartheta'}[\varphi^*]$, *wobei in den Fällen* $\mathbb{E}_\vartheta[\varphi^*] \in (0,1)$ *und* $\mathbb{E}_{\vartheta'}[\varphi^*] \in (0,1)$ *sogar die strikte Ungleichung gilt.*

4.7 Satz (Verallgemeinertes NP-Lemma)**.** *Es seien* $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ *eine Exponentialfamilie in* $\eta(\vartheta)$ *und* T, L *die zugehörige Likelihoodfunktion sowie* $\vartheta_0, \vartheta_1 \in \Theta$ *zwei Parameter. Erfüllt ein Test für* $H_0 : \vartheta = \vartheta_0$ *gegen* $H_1 : \vartheta = \vartheta_1$ *der Form*
$$\varphi(x) = \begin{cases} 1, & \textit{falls } L(\vartheta_1, x) > kL(\vartheta_0, x) + lT(x)L(\vartheta_0, x) \\ 0, & \textit{falls } L(\vartheta_1, x) < kL(\vartheta_0, x) + lT(x)L(\vartheta_0, x) \\ \gamma, & \textit{falls } L(\vartheta_1, x) = kL(\vartheta_0, x) + lT(x)L(\vartheta_0, x) \end{cases}$$
mit $k, l \in \mathbb{R}^+$ *und* $\gamma \in [0,1]$ *die Nebenbedingungen*
$$\mathbb{E}_{\vartheta_0}[\varphi] = \alpha \quad \textit{und} \quad \mathbb{E}_{\vartheta_0}[T\varphi] = \alpha\,\mathbb{E}_{\vartheta_0}[T],$$
so maximiert er die Güte $\mathbb{E}_{\vartheta_1}[\varphi]$ *in der Menge aller Tests, die diese Nebenbedingungen erfüllen.*

4.8 Satz. $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ *sei eine einparametrische Exponentialfamilie in* $\eta(\vartheta)$ *und* T. $\Theta \subseteq \mathbb{R}$ *sei offen,* $\vartheta_0 \in \Theta$ *und* $\eta \in C^1(\Theta)$ *sei streng monoton (wachsend oder fallend) mit* $\eta'(\vartheta_0) \neq 0$. *Für* $\alpha \in (0,1)$, $c_1 < c_2$ *und* $\gamma_1, \gamma_2 \in [0,1]$ *erfülle der Test*
$$\varphi^*(x) = \begin{cases} 1, & \textit{falls } T(x) < c_1 \textit{ oder } T(x) > c_2 \\ 0, & \textit{falls } T(x) \in (c_1, c_2) \\ \gamma_i, & \textit{falls } T(x) = c_i,\ i = 1, 2 \end{cases}$$
die Nebenbedingungen
$$\mathbb{E}_{\vartheta_0}[\varphi^*] = \alpha \quad \textit{und} \quad \mathbb{E}_{\vartheta_0}[T\varphi^*] = \alpha\,\mathbb{E}_{\vartheta_0}[T].$$
Dann ist φ^* *gleichmäßig bester unverfälschter Test zum Niveau* α *für* $H_0 : \vartheta = \vartheta_0$ *gegen* $H_1 : \vartheta \neq \vartheta_0$.

4.2 Bedingte Tests

4.9 Definition. *Eine $(S,\mathscr{S})$-wertige Statistik T auf $(\mathcal{X},\mathscr{F},(\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ heißt vollständig (bezüglich Θ), falls für alle messbaren Funktionen $f : S \to \mathbb{R}$ gilt*

$$\forall\vartheta\in\Theta:\ \mathbb{E}_\vartheta[f(T)] = 0 \text{ (und existiert) } \Rightarrow \forall\vartheta\in\Theta:\ \mathbb{P}_\vartheta(f(T)=0)=1.$$

4.10 Definition. *Es sei $\Theta' \subseteq \Theta$. Dann heißt ein Test φ α-ähnlich auf Θ', wenn $\mathbb{E}_\vartheta[\varphi] = \alpha$ für alle $\vartheta \in \Theta'$ gilt.*

4.11 Satz. *Ist T eine bezüglich Θ' vollständige und suffiziente Statistik und ist φ ein auf Θ' α-ähnlicher Test, so gilt $\mathbb{E}_\bullet[\varphi\,|\,T] = \alpha$ $\mathbb{P}_\vartheta$-f.s. für alle $\vartheta\in\Theta'$.*

4.12 Satz. *Es sei $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ eine k-parametrische natürliche Exponentialfamilie in T. Enthält $\Theta'\subseteq\Theta$ eine offene Menge im $\mathbb{R}^k$, so ist T suffizient und vollständig bezüglich Θ'.*

4.13 Satz. *Gegeben sei die natürliche Exponentialfamilie*

$$\frac{d\,\mathbb{P}_\vartheta}{d\mu}(x) = C(\vartheta)h(x)\exp\Big(\vartheta^0 U(x) + \sum_{i=1}^k \vartheta^i T_i(x)\Big), \quad x\in\mathcal{X},\, \vartheta\in\Theta,$$

sowie $\alpha\in(0,1)$ und ein Punkt ϑ_0 im Innern von Θ. Dann ist

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } U(x) < K(T(x)) \\ 0, & \text{falls } U(x) > K(T(x)) \\ \gamma(T(x)), & \text{falls } U(x) = K(T(x)) \end{cases}$$

mit $K(t)\in\mathbb{R}$, $\gamma(t)\in[0,1]$ derart, dass $\mathbb{E}_{\vartheta_0}[\varphi^*\,|\,T] = \mathbb{E}_{\vartheta_0^0}[\varphi^*\,|\,T] = \alpha$ $\mathbb{P}_{\vartheta_0}$-f.s., ein gleichmäßig bester unverfälschter Test zum Niveau α von $H_0 : \vartheta^0 \leqslant \vartheta_0^0$ gegen $H_1 : \vartheta^0 > \vartheta_0^0$ (d.h. $\Theta_0 = \{\vartheta\in\Theta\,|\,\vartheta^0\leqslant\vartheta_0^0\}$, $\Theta_1 = \{\vartheta\in\Theta\,|\,\vartheta^0 > \vartheta_0^0\}$).

4.14 Satz. *Es liege die Situation des vorigen Satzes vor. Dann ist*

$$\varphi^*(x) = \begin{cases} 1, & \text{falls } U(x) < K_1(T(x)) \text{ oder } U(x) > K_2(T(x)) \\ 0, & \text{falls } U(x) \in (K_1(T(x)), K_2(T(x))) \\ \gamma_i(T(x)), & \text{falls } U(x) = K_i(T(x)),\ i=1,2, \end{cases}$$

mit $K_i(t)\in\mathbb{R}$, $\gamma_i(t)\in[0,1]$ derart, dass

$$\mathbb{E}_{\vartheta_0^0}[\varphi^*\,|\,T] = \alpha \text{ und } \mathbb{E}_{\vartheta_0^0}[U\varphi^*\,|\,T] = \alpha\,\mathbb{E}_{\vartheta_0^0}[U\,|\,T] \quad \mathbb{P}_{\vartheta_0}\text{-f.s.}$$

ein gleichmäßig bester unverfälschter Test zum Niveau α von $H_0 : \vartheta^0 = \vartheta_0^0$ gegen $H_1 : \vartheta^0 \neq \vartheta_0^0$.

4.3 Tests im Normalverteilungsmodell

4.15 Satz. *Es sei $X_1, \dots, X_n$ eine $N(\mu, \sigma^2)$-verteilte mathematische Stichprobe mit $\mu \in \mathbb{R}$ und $\sigma > 0$ unbekannt. Für $\sigma_0 > 0$ ist ein gleichmäßig bester unverfälschter Test von $H_0 : \sigma \leqslant \sigma_0$ gegen $H_1 : \sigma > \sigma_0$ zum Niveau $\alpha \in (0,1)$ gegeben durch*

$$\varphi^*(X_1, \dots, X_n) = \begin{cases} 1, & \text{falls } \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 > K_\alpha \\ 0, & \text{falls } \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \bar{X})^2 \leqslant K_\alpha \end{cases}$$

mit dem α-Fraktil K_α der $\chi^2(n-1)$-Verteilung:

$$\int_{K_\alpha}^\infty \frac{2^{-(n-1)/2}}{\Gamma((n-1)/2)} z^{(n-1)/2-1} e^{-z/2} dz = \alpha.$$

4.16 Lemma. *Sind $Z_1, \dots, Z_n$ unabhängig $N(0, \sigma^2)$-verteilt sowie $f : \mathbb{R}^n \to \mathbb{R}$ messbar mit $f(cx) = f(x)$ für alle $c > 0$, $x \in \mathbb{R}^n$, so ist $f(Z_1, \dots, Z_n)$ unabhängig von $\sum_{i=1}^n Z_i^2$. Insbesondere sind jeweils $\frac{\bar{Z}}{\sqrt{\sum_i (Z_i - \bar{Z})^2}}$ und $\frac{\bar{Z}}{\sqrt{\sum_i Z_i^2}}$ unabhängig von $\sum_{i=1}^n Z_i^2$.*

4.17 Satz. *Es sei $X_1, \dots, X_n$ eine $N(\mu, \sigma^2)$-verteilte mathematische Stichprobe mit $\mu \in \mathbb{R}$ und $\sigma > 0$ unbekannt. Ein gleichmäßig bester unverfälschter Test von $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$ zum Niveau $\alpha \in (0,1)$ ist gegeben durch den <u>zweiseitigen t-Test</u>*

$$\varphi^*(X) = \mathbf{1}_{\{|t(X)| > K_{\alpha/2}\}}, \quad t(X) := \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}},$$

mit dem $\alpha/2$-Fraktil $K_{\alpha/2}$ der $t(n-1)$-Verteilung :

$$\int_{K_{\alpha/2}}^\infty \frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma((n-1)/2)} \Big(1 + \frac{z^2}{n-1}\Big)^{-n/2} dz = \alpha/2.$$

4.18 Satz. *Es werden zwei unabhängige mathematische Stichproben $X_1, \dots, X_m \sim N(\mu, \sigma^2)$ und $Y_1, \dots, Y_n \sim N(\nu, \sigma^2)$ beobachtet mit $\mu, \nu \in \mathbb{R}$ und $\sigma > 0$ unbekannt. Ein gleichmäßig bester unverfälschter Test von $H_0 : \mu = \nu$ gegen $H_1 : \mu \neq \nu$ zum Niveau $\alpha \in (0,1)$ ist gegeben durch*

$$\varphi^*(X, Y) = \mathbf{1}_{\{|t(X,Y)| > K_{\alpha/2}\}},$$

$$\text{mit} \qquad t(X, Y) := \frac{(\frac{1}{m} + \frac{1}{n})^{-1/2} (\bar{Y} - \bar{X})}{\sqrt{(\sum_{i=1}^m (X_i - \bar{X})^2 + \sum_{j=1}^n (Y_j - \bar{Y})^2)/(m+n-2)}}$$

und dem $\alpha/2$-Fraktil $K_{\alpha/2}$ der $t(m+n-2)$-Verteilung.

4.19 Satz. *Es werden zwei unabhängige mathematische Stichproben $X_1, \dots, X_m \sim N(\mu, \sigma^2)$ und $Y_1, \dots, Y_n \sim N(\nu, \tau^2)$ beobachtet mit $\mu, \nu \in \mathbb{R}$ und $\sigma, \tau > 0$ unbekannt. Für $c_0 > 0$ ist ein gleichmäßig bester unverfälschter*

Test von $H_0 : \tau^2 \leqslant c_0\sigma^2$ gegen $H_1 : \tau^2 > c_0\sigma^2$ zum Niveau $\alpha \in (0,1)$ gegeben durch

$$\varphi^*(X,Y) = \mathbf{1}_{\{c_0^{-1}V(X,Y)>K_\alpha\}},$$
$$\textit{mit} \qquad V(X,Y) := \frac{\sum_{j=1}^n (Y_j - \bar{Y})^2/(n-1)}{\sum_{i=1}^m (X_i - \bar{X})^2/(m-1)}$$

und dem α-Fraktil K_α der $F(n-1,m-1)$-Verteilung:

$$\int_{K_\alpha}^\infty \frac{\Gamma((m+n-2)/2)(\frac{n-1}{m-1})^{(n-1)/2}}{\Gamma((m-1)/2)\Gamma((n-1)/2)} \frac{z^{(n-3)/2}}{(1+\frac{n-1}{m-1}z)^{(m+n-2)/2}}\,dz = \alpha.$$

5 Schätztheorie

5.1 Momentenschätzer

5.1 Definition. *Es seien $(\mathcal{X}^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n})_{\vartheta\in\Theta})$ ein statistisches (Produkt-)Experiment mit $\mathcal{X} \subseteq \mathbb{R}$, $\mathscr{F} \subseteq \mathfrak{B}_\mathbb{R}$ und $g(\vartheta)$ mit $g : \Theta \to \mathbb{R}^p$ ein abgeleiteter Parameter. Ferner sei $\psi = (\psi_1, \ldots, \psi_q) : \mathbb{R} \to \mathbb{R}^q$ derart, dass*

$$\varphi(\vartheta) := \mathbb{E}_\vartheta[\psi] = (\mathbb{E}_\vartheta[\psi_j])_{j=1,\ldots,q}$$

existiert. Gibt es nun eine Borel-messbare Funktion $G : \varphi(\Theta) \to g(\Theta)$ mit $G \circ \varphi = g$ und liegt $\frac{1}{n}\sum_{i=1}^n \psi(x_i)$ in $\varphi(\Theta)$ für alle $x_1, \ldots, x_n \in \mathcal{X}$, so heißt $G(\frac{1}{n}\sum_{i=1}^n \psi(x_i))$ <u>Momentenschätzer</u> für $g(\vartheta)$ mit Momentenfunktionen $\psi_1, \ldots, \psi_q$.

5.2 Lemma. *Existiert für hinreichend großes n der Momentenschätzer $\hat{g}_n = G(\frac{1}{n}\sum_{i=1}^n \psi(x_i))$ und ist G stetig, so ist $\hat{g}_n$ (stark) konsistent, d.h. $\lim_{n\to\infty} \hat{g}_n = g(\vartheta)$ $\mathbb{P}_\vartheta$-f.s.*

5.3 Satz (Δ-Methode)**.** *Es seien (X_n) eine Folge von Zufallsvektoren im $\mathbb{R}^k$, $\sigma_n > 0$, $\sigma_n \to 0$, $\vartheta_0 \in \mathbb{R}^k$ sowie $\Sigma \in \mathbb{R}^{k\times k}$ positiv definit und es gelte*

$$\sigma_n^{-1}(X_n - \vartheta_0) \xrightarrow{d} N(0,\Sigma).$$

Ist $f : \mathbb{R}^k \to \mathbb{R}$ in einer Umgebung von ϑ_0 stetig differenzierbar mit $(\nabla f(\vartheta_0))^\top \Sigma\, \nabla f(\vartheta_0) > 0$, so folgt

$$\sigma_n^{-1}(f(X_n) - f(\vartheta_0)) \xrightarrow{d} N(0, (\nabla f(\vartheta_0))^\top \Sigma\, \nabla f(\vartheta_0)).$$

5.4 Satz. *Es seien $\vartheta_0 \in \Theta$, $g : \Theta \to \mathbb{R}$ und für hinreichend großes n existiere der Momentenschätzer $\hat{g}_n = G(\frac{1}{n}\sum_{i=1}^n \psi(x_i))$ mit Momentenfunktionen $\psi_j \in L^2(\mathbb{P}_{\vartheta_0})$, $j = 1, \ldots, q$. Setze $\Sigma(\vartheta_0) := (\mathrm{Cov}_{\vartheta_0}(\psi_i, \psi_j))_{i,j=1,\ldots,q}$. Sofern G in einer Umgebung von $\varphi(\vartheta_0)$ stetig differenzierbar ist mit $\sigma^2 := (\nabla G(\varphi(\vartheta_0)))^\top \Sigma(\vartheta_0)\, \nabla G(\varphi(\vartheta_0)) > 0$, ist $\hat{g}_n$ unter $\mathbb{P}_{\vartheta_0}^{\otimes n}$ asymptotisch normalverteilt mit Rate $n^{-1/2}$ und asymptotischer Varianz σ^2:*

$$\sqrt{n}(\hat{g}_n - g(\vartheta_0)) \xrightarrow{d} N(0,\sigma^2).$$

5.2 Maximum-Likelihood- und M-Schätzer

5.5 Definition. *Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein von μ dominiertes Experiment mit Likelihoodfunktion $L(\vartheta, x)$. Eine Statistik $\hat{\vartheta} : \mathcal{X} \to \Theta$ (Θ trage eine σ-Algebra $\mathscr{F}_\Theta$) heißt Maximum-Likelihood-Schätzer (MLE) von ϑ, falls $L(\hat{\vartheta}(x), x) = \sup_{\vartheta\in\Theta} L(\vartheta, x)$ für $\mathbb{P}_\vartheta$-fast alle $x \in \mathcal{X}$ und alle $\vartheta \in \Theta$ gilt.*

Mit $\ell(\vartheta, x) := \log L(\vartheta, x)$ wird die Loglikelihood-Funktion bezeichnet.

5.6 Lemma. *Für eine natürliche Exponentialfamilie $(\mathbb{P}_\vartheta)_{\vartheta\in\Theta}$ in $T(x)$ ist der MLE $\hat{\vartheta}$ implizit gegeben durch die Momentengleichung $\mathbb{E}_{\hat{\vartheta}}[T] = T(x)$, vorausgesetzt der MLE existiert und liegt im Innern int(Θ) von Θ.*

5.7 Definition. *Es sei $(\mathcal{X}_n, \mathscr{F}_n, (\mathbb{P}^n_\vartheta)_{\vartheta\in\Theta})_{n\geqslant 1}$ eine Folge statistischer Experimente. Eine Funktion $K : \Theta \times \Theta \to \mathbb{R} \cup \{+\infty\}$ heißt Kontrastfunktion, falls $\vartheta \mapsto K(\vartheta_0, \vartheta)$ ein eindeutiges Minimum bei ϑ_0 hat für alle $\vartheta_0 \in \Theta$. Eine Folge $K_n : \Theta \times \mathcal{X}_n \to \mathbb{R} \cup \{+\infty\}$ heißt zugehöriger Kontrastprozess (oder bloß Kontrast), falls folgende Bedingungen gelten:*

(a) $K_n(\vartheta, \bullet)$ ist $\mathscr{F}_n$-messbar für alle $\vartheta \in \Theta$;

(b) $\forall \vartheta, \vartheta_0 \in \Theta : \; K_n(\vartheta) \xrightarrow{\mathbb{P}^n_{\vartheta_0}} K(\vartheta_0, \vartheta)$ für $n \to \infty$.

Ein zugehöriger M-Schätzer (oder Minimum-Kontrast-Schätzer) ist gegeben durch $\hat{\vartheta}_n(x_n) := \operatorname{argmin}_{\vartheta\in\Theta} K_n(\vartheta, x_n)$ (sofern existent; nicht notwendigerweise eindeutig).

5.8 Satz. *Es sei $(K_n)_{n\geqslant 1}$ ein Kontrastprozess zur Kontrastfunktion K. Dann ist der zugehörige M-Schätzer $\hat{\vartheta}_n$ konsistent für $\vartheta_0 \in \Theta$ unter folgenden Bedingungen:*

(A1) *Θ ist ein kompakter Raum;*

(A2) *$\vartheta \mapsto K(\vartheta_0, \vartheta)$ ist stetig und $\vartheta \mapsto K_n(\vartheta)$ ist $\mathbb{P}^n_{\vartheta_0}$-f.s. stetig;*

(A3) $\sup_{\vartheta\in\Theta} |K_n(\vartheta) - K(\vartheta_0, \vartheta)| \xrightarrow{\mathbb{P}^n_{\vartheta_0}} 0.$

5.9 Satz. *Es mögen die Annahmen (A1)-(A3) sowie $\Theta \subseteq \mathbb{R}^k$ und $\vartheta_0 \in$ int(Θ) gelten. Der Kontrastprozess K_n sei zweimal stetig differenzierbar in einer Umgebung von ϑ_0 ($\mathbb{P}^n_{\vartheta_0}$-f.s.), so dass mit*

$$U_n(\vartheta) := \nabla_\vartheta K_n(\vartheta) \text{ (\underline{Score})}, \qquad V_n(\vartheta) := \nabla^2_\vartheta K_n(\vartheta)$$

folgende Konvergenzen unter $\mathbb{P}^n_{\vartheta_0}$ gelten:

(a) $\sqrt{n} U_n(\vartheta_0) \xrightarrow{d} N(0, I(\vartheta_0))$ mit $I(\vartheta_0) \in \mathbb{R}^{k\times k}$ positiv definit.

(b) Aus $\vartheta_n \xrightarrow{\mathbb{P}^n_{\vartheta_0}} \vartheta_0$ folgt $V_n(\vartheta_n) \xrightarrow{\mathbb{P}^n_{\vartheta_0}} V(\vartheta_0)$ mit $V(\vartheta_0) \in \mathbb{R}^{k\times k}$ regulär.

Dann ist der M-Schätzer $\hat{\vartheta}_n$ asymptotisch normalverteilt. Genauer gilt unter $\mathbb{P}^n_{\vartheta_0}$:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \xrightarrow{d} N(0, V(\vartheta_0)^{-1} I(\vartheta_0) V(\vartheta_0)^{-1}).$$

5.10 Satz. *Ist $\Theta \subseteq \mathbb{R}^k$ kompakt, $(X_n(\vartheta),\, \vartheta \in \Theta)_{n\geqslant 1}$ eine Folge stetiger Prozesse mit $X_n(\vartheta) \xrightarrow{\mathbb{P}} X(\vartheta)$ für alle $\vartheta \in \Theta$ und stetigem Grenzprozess $(X(\vartheta),\, \vartheta \in \Theta)$, so gilt $\max_{\vartheta\in\Theta}|X_n(\vartheta) - X(\vartheta)| \xrightarrow{\mathbb{P}} 0$ genau dann, wenn (X_n) straff ist, also wenn*

$$\forall \varepsilon, \eta > 0\, \exists \delta > 0 : \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{|\vartheta_1-\vartheta_2|<\delta} |X_n(\vartheta_1) - X_n(\vartheta_2)| \geqslant \varepsilon\Big) \leqslant \eta.$$

5.11 Satz. *Es seien $(\mathcal{X}^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n})_{\vartheta\in\Theta})_{n\geqslant 1}$ eine Folge dominierter Produktexperimente mit eindimensionaler Loglikelihoodfunktion $\ell(\vartheta, x) = \log(\frac{d\mathbb{P}_\vartheta}{d\mu}(x))$. Es gelte:*

(a) $\Theta \subseteq \mathbb{R}^k$ ist kompakt und ϑ_0 liegt im Innern $int(\Theta)$ von Θ.

(b) $\vartheta \mapsto \ell(\vartheta, x) = \log(L(\vartheta, x))$ ist zweimal stetig differenzierbar in einer Umgebung U von ϑ_0 für alle $x \in \mathcal{X}$.

(c) Für $i = 0, 1, 2$ gibt es $H_i \in L^1(\mathbb{P}_{\vartheta_0})$ mit $\sup_{\vartheta\in\Theta}|\ell(\vartheta, x)| \leqslant H_0(x)$ und $\sup_{\vartheta\in U}|\nabla_\vartheta^i \ell(\vartheta, x)| \leqslant H_i(x)$ für $i = 1, 2$, $x \in \mathcal{X}$.

(d) Die Fisher-Informationsmatrix $I(\vartheta_0) = \mathbb{E}_{\vartheta_0}[(\nabla_\vartheta \ell(\vartheta_0))(\nabla_\vartheta \ell(\vartheta_0))^\top]$ ist positiv definit.

Dann gilt für den MLE $\hat{\vartheta}_n$ unter $\mathbb{P}_{\vartheta_0}^{\otimes n}$

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \xrightarrow{d} N(0, I(\vartheta_0)^{-1}).$$

Ferner gilt die Formel $I(\vartheta_0) = -\,\mathbb{E}_{\vartheta_0}[\nabla_\vartheta^2 \ell(\vartheta_0)]$.

5.3 Effizienz

5.12 Definition. *Für $n \geqslant 1$ seien $\hat{\vartheta}_{n,1}$ und $\hat{\vartheta}_{n,2}$ Schätzer von ϑ definiert auf $(\mathcal{X}^n, \mathscr{F}^{\otimes n}, (\mathbb{P}_\vartheta^{\otimes n})_{\vartheta\in\Theta})$ mit $\Theta \subseteq \mathbb{R}^k$ sowie*

$$V_{n,i}(\vartheta)^{-1/2}(\hat{\vartheta}_{n,i} - \vartheta) \xrightarrow{d} N(0, E_k) \qquad \text{für alle } \vartheta \in \Theta,\ i = 1, 2$$

mit geeigneten symmetrisch, positiv-definiten Matrizen $V_{n,i}(\vartheta)$, E_k Einheitsmatrix. Dann heißt $(\hat{\vartheta}_{n,1})_{n\geqslant 1}$ asymptotisch effizienter als $(\hat{\vartheta}_{n,2})_{n\geqslant 1}$, falls $V_{n,1}(\vartheta) \leqslant V_{n,2}(\vartheta)$ (d.h. $V_{n,2}(\vartheta) - V_{n,1}(\vartheta)$ positiv semi-definit) für alle $\vartheta \in \Theta$, $n \geqslant 1$ gilt.

5.13 Satz (Cramér-Rao)**.** *Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ mit $\Theta \subseteq \mathbb{R}^k$ ein von μ dominiertes Experiment mit Likelihoodfunktion $L(\vartheta, x)$. Ferner sei $g : \Theta \to \mathbb{R}$ differenzierbar, $\hat{g}$ ein erwartungstreuer Schätzer von $g(\vartheta)$ sowie*

$$\nabla_\vartheta \int_{\mathcal{X}} h(x) L(\vartheta, x)\, \mu(dx) = \int_{\mathcal{X}} h(x) \nabla_\vartheta L(\vartheta, x)\, \mu(dx), \qquad \vartheta \in \Theta,$$

für $h(x) = 1$ und $h(x) = \hat{g}(x)$. Ist die Fisher-Informationsmatrix $I(\vartheta)$ positiv definit, so gilt folgende untere Schranke für das quadratische Risiko von $\hat{g}$:

$$\mathbb{E}_\vartheta[(\hat{g} - g(\vartheta))^2] = \operatorname{Var}_\vartheta(\hat{g}) \geqslant (\nabla_\vartheta g(\vartheta))^\top I(\vartheta)^{-1}\, \nabla_\vartheta g(\vartheta), \qquad \vartheta \in \Theta.$$

5.4 Nichtparametrische Dichteschätzung

5.14 Definition. *Eine Funktion $K : \mathbb{R} \to \mathbb{R}$ heißt Kern (oder Kernfunktion), falls $\int_{-\infty}^{\infty} K(x) = 1$ und $K \in L^2(\mathbb{R})$. Gilt*

$$\int_{-\infty}^{\infty} K(x)x^p dx = 0, \quad 1 \leqslant p \leqslant P,$$

sowie $\int |K(x)x^{P+1}|dx < \infty$, so besitzt der Kern K die Ordnung P. Für $h > 0$ setze $K_h(x) := h^{-1}K(h^{-1}x)$. Hierbei wird h als Bandweite bezeichnet.

5.15 Definition. *Für reellwertige Beobachtungen $X_1, \ldots, X_n$ bezeichnet*

$$\hat{f}_{h,n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i), \quad x \in \mathbb{R}$$

den Kerndichteschätzer zu gegebenem Kern K mit Bandweite $h > 0$.

5.16 Satz. *Es sei $X_1, \ldots, X_n$ eine mathematische Stichprobe gemäß einer Dichte f. Gilt $f \in C^s(\mathbb{R})$ und besitzt der Kern K die Ordnung $P \geqslant s - 1$, so gilt für das quadratische Risiko der Kerndichteschätzung*

$$\forall\, x_0 \in \mathbb{R} :\ \mathbb{E}_f[(\hat{f}_{h,n}(x_0) - f(x_0))^2] \leqslant C(K,s)\|f^{(s)}\|_\infty h^s + \|K\|_{L^2}^2 \|f\|_\infty (nh)^{-1},$$

wobei $C(K,s) > 0$ nur von K und s abhängt.

5.17 Korollar. *Setze für $s \geqslant 1$, $R > 0$*

$$\mathscr{D}(s,R) := \{f : \mathbb{R} \to \mathbb{R}^+ \mid f \in C^s(\mathbb{R}),\ \int f(x)dx = 1,\ \max(\|f\|_\infty, \|f^{(s)}\|_\infty) \leqslant R\}.$$

Dann erfüllt der Kerndichteschätzer mit einem Kern der Ordnung $P \geqslant s - 1$ und der Bandweite $h(n) = Cn^{-s/(2s+1)}$, $C > 0$ beliebig, asymptotisch:

$$\forall\, x_0 \in \mathbb{R} :\ \limsup_{n\to\infty} n^{2s/(2s+1)} \sup_{f \in \mathscr{D}(s,R)} \mathbb{E}_f[(\hat{f}_{n,h(n)}(x_0) - f(x_0))^2] < \infty.$$

Insbesondere ergeben sich die Konvergenzraten $n^{-2/3}$ (s=1), $n^{-4/5}$ (s=2) sowie als Grenzwert für $s \to \infty$ die parametrische Rate n^{-1} für das quadratische Risiko.