Statistik IV

Modul P8: Grundlagen der Statistik II
Vorlesung P8.1: Wahrscheinlichkeitstheorie und Inferenz II

Prof. Dr. Torsten Hothorn
Institut für Statistik
Ludwig–Maximilians–Universität München
LaTeX-Satz von cand. stat. Andreas Böck

7. Mai 2009

Kapitel 1

Grundlagen des statistischen Schließens

1.1 Grundbegriffe und Problemstellung

Definition 1.1 (Beobachtung, Stichprobe)
Ist $\widehat{\omega} \in \Omega_1$ ein beobachtetes Elementereignis, so heißt $x = X(\widehat{\omega}) = (X_1(\widehat{\omega}), \ldots, X_n(\widehat{\omega}))$ eine Stichprobe vom Stichprobenumfang n mit Beobachtungen $x_1 = X_1(\widehat{\omega}), \ldots, x_n = X_n(\widehat{\omega})$.

Definition 1.2 (Verteilungsannahme)
Sei $\mathbb{P}_{X,I}$ eine Menge von Verteilungen, die für die Zufallsvariable X im Prinzip in Betracht kommen:

$$\mathbb{P}_{X,I} = \{\mathbb{P}_{X,i} \mid i \in I\}\,.$$

Dann heißt $\mathbb{P}_{X,I}$ Verteilungsannahme.

Definition 1.3 (Parameter, Parameterraum)
Sei $\Theta \subseteq \mathbb{R}^k$ und $g : \Theta \to \mathbb{P}_{X,I}$ bijektiv mit $\mathbb{P}_{X,\vartheta} = g(\vartheta)$ für $\vartheta \in \Theta$. Dann heißt ϑ Parameter der Verteilung $\mathbb{P}_{X,\vartheta}$, g heißt Parametrisierung von $\mathbb{P}_{X,I}$ $\quad (=: \mathbb{P}_{X,\Theta})$ und $\mathbb{P}_{X,\Theta}$ heißt k-parametrige Verteilungsannahme mit Parameterraum Θ.

Definition 1.4 (Stichprobenraum, statistischer Raum)
$(\mathcal{X}, \sigma(\mathcal{X}), \mathbb{P}_{X,\Theta})$ heißt statistischer Raum, $(\mathcal{X}, \sigma(\mathcal{X}))$ heißt Stichprobenraum.

1.2 Schätzer und ihre Eigenschaften

Definition 1.5 (Stichprobenfunktion, Statistik)
Eine meßbare Funktion $t : \mathcal{X} \to \mathbb{R}^k$ heißt Stichprobenfunktion oder Statistik.

Definition 1.6 (Schätzen, Schätzfunktion, Schätzwert)
Sei $\Theta \subseteq \mathbb{R}^k$ und $t : \mathcal{X} \to \mathbb{R}^k$ eine Statistik. Dann heißt t Schätzer oder Schätzfunktion für $\vartheta \in \Theta$ und $\widehat{\vartheta} = t(x)$ heißt Schätzwert für $\vartheta \in \Theta$.

Definition 1.7 (parametrische Funktion)
Eine Abbildung $g : \Theta \to \mathbb{R}^l$ heißt parametrische Funktion.

Definition 1.8
Sei $\Theta \subseteq \mathbb{R}^k, g : \mathbb{R}^k \to \mathbb{R}^l$ eine parametrische Funktion und $t : \mathcal{X} \to \mathbb{R}^l$ eine Statistik. Dann ist t ein Schätzer für $g(\vartheta)$ und $\widehat{g(\vartheta)} = t(x)$ dessen Schätzwert.

Definition 1.9 (Erwartungstreue, Verzerrung)
Sei $T = t \circ X$ und t ein Schätzer für $g(\vartheta)$. Dann heißt

$$b_g(\vartheta) = \mathbb{E}_\vartheta(T) - g(\vartheta)$$

die Verzerrung (engl. bias) von t, wobei $\mathbb{E}_\vartheta(T) = \int t \; d\mathbb{P}_{X,\vartheta}$. Falls $b_g(\vartheta) = 0$, also $\mathbb{E}_\vartheta(T) = g(\vartheta)$, so heißt t erwartungstreu oder unverzerrt (engl. unbiased) für $g(\vartheta)$.

Definition 1.10 (Schätzverfahren)
Für $n \in \mathbb{N}$ sei jeweils $t_n : \mathbb{R}^n \to \mathbb{R}^l$ eine Schätzfunktion für $g(\vartheta)$. Dann heißt die Menge $\{t_n \mid n \in \mathbb{N}\}$ Schätzverfahren für $g(\vartheta)$.

Definition 1.11 (asymptotisch erwartungstreu, Konsistenz)
Sei $\underline{X_n} = (X_1, \ldots, X_n)$, eine Folge von Stichproben, $\mathbb{P}_{\underline{X}_n,\Theta}$ die Verteilungsannahme, $\vartheta \in \Theta$ und g eine parametrische Funktion, sowie $\{t_n \mid n \in \mathbb{N}\}$ ein Schätzverfahren. Dann heißt die Folge von Zufallsvariablen $T_n = t_n \circ \underline{X}_n, n \in \mathbb{N}$

- asymptotisch erwartungstreu für $g(\vartheta)$:⇔
  $\lim_{n\to\infty} \mathbb{E}_\vartheta(T_n) = g(\vartheta) \quad \forall \vartheta \in \Theta$
- schwach konsistent für $g(\vartheta)$:⇔
  $T_n \xrightarrow{\mathbb{P}} g(\vartheta)$
- stark konsistent für $g(\vartheta)$:⇔
  $T_n \xrightarrow{\text{f.s.}} g(\vartheta)$

Satz 1.12
Sei $t : \mathbb{R}^k \to \mathbb{R}$ ein Schätzer für $\vartheta \in \Theta \subset \mathbb{R}$. Dann gilt:

$$\mathbb{P}_\vartheta(|t(x) - \vartheta| < \epsilon) \geq 1 - \frac{1}{\epsilon^2} \mathbb{E}_\vartheta \left((t(X) - \vartheta)^2 \right)$$

Der Ausdruck $\mathbb{E}_\vartheta((t(X) - \vartheta)^2) := \text{MSE}(t, \vartheta)$ heißt mittlerer quadratischer Fehler (mean squared error).

Definition 1.13
Seien $t_1 : \mathbb{R}^n \to \mathbb{R}$ und $t_2 : \mathbb{R}^n \to \mathbb{R}$ Schätzer für $\theta \in \Theta \subset \mathbb{R}$. Dann heißt t_1 mindestens so gut wie t_2, wenn gilt

$$\text{MSE}(t_1, \vartheta) \leq \text{MSE}(t_2, \vartheta) \quad \forall \vartheta \in \Theta$$

Satz 1.14 (Bias-Varianz-Zerlegung)

$$\text{MSE}(t, \vartheta) = \mathbb{V}_\vartheta(t(X)) + (\underbrace{\mathbb{E}_\vartheta\,(t(X)) - \vartheta}_{\text{bias } b(t,\vartheta)})^2$$

1.3 Exponentialfamilien

Definition 1.15 (Exponentialfamilie)
Eine Verteilungsfamilie $\mathbb{P}_{X,\Theta} = \{f(x; \vartheta) \mid \vartheta \in \Theta\}$ (d.h., die Dichten existieren) heißt k-parametrige Exponentialfamilie wenn sich die Dichten $f(x; \vartheta)$ bzgl. (des oder eines geeigneten dominierenden Maßes) ν auf die folgende Gestalt bringen lassen:

$$f(x, \vartheta) = c(\vartheta)\, b(x)\, \exp\left(\sum_{j=1}^{k} \gamma_j(\vartheta)\, t_j(x) \right)$$

mit

$$\left.\begin{array}{rcl} t_j & : & \mathcal{X} \to \mathbb{R} \qquad t = (t_1, \dots, t_k) \\ b & : & \mathcal{X} \to \mathbb{R}^+ \end{array}\right\} \quad \text{meßbar}$$

und

$$\begin{array}{rcl} \gamma_j & : & \Theta \to \mathbb{R} \qquad \gamma = (\gamma_1, \dots, \gamma_k) \\ c & : & \Theta \to \mathbb{R}^+ \end{array}$$

$\gamma = \gamma(\vartheta)$ heißt natürlicher Parameter dieser Exponentialfamilie mit natürlichem Parameterraum $\gamma(\Theta) = \{\gamma(\vartheta) \mid \vartheta \in \Theta\} \subseteq \mathbb{R}^k$.

Satz 1.16
Folgende Verteilungsfamilien sind Exponentialfamilien:

a) $\{Exp(\lambda) \mid \lambda > 0\}$ Exponentialverteilung

b) $\{B(1, p) \mid p \in [0, 1]\}$ Bernoulliverteilung

c) $\{B(n, p) \mid p \in [0, 1]\}$ Binomialverteilung

d) $\{P(\lambda) \mid \lambda > 0\}$ Poissonverteilung

e) $\left\{N(\mu, \sigma_0^2) \mid \mu \in \mathbb{R}\right\}$ Normalverteilung

f) $\left\{N(\mu_0, \sigma^2) \mid \sigma^2 > 0\right\}$ Normalverteilung

Satz 1.17
Sei $\mathbb{P}_{X,\vartheta} = f(\cdot;\vartheta) \odot \nu$ Maß mit Dichte f bzgl. ν (in der Regel $\nu = \lambda$ oder $\nu = \mu_z$). Sei nun $\tilde{\nu}(B) := \int_B b(x)\, d\nu(x) \quad (B \subseteq \mathcal{X})$ (also $\tilde{\nu} = b \odot \nu$) , dann gilt

$$\mathbb{P}_{X,\vartheta} = \tilde{f} \odot \tilde{\nu}$$

wobei

$$\tilde{f}(x, \vartheta) = c(\vartheta)\, \exp(\gamma(\vartheta)^\top t(x))$$

bzgl. mit $\gamma_0(\vartheta) = \log(c(\vartheta))$

$$\tilde{f}(x, \vartheta) = \exp(\gamma_0(\vartheta) + \gamma(\vartheta)^\top t(x))$$

Definition 1.18
Eine Exponentialfamilie heißt strikt k-parametrig, wenn ihre Dichten in der Form von Def. 1.15 oder Satz 1.17 darstellbar sind, und zusätzlich die Funktionen $t_0 \equiv 1, t_1, \dots, t_k$ sowie $\gamma_0 \equiv 1, \gamma_1, \dots, \gamma_k$ linear unabhängig sind (außer auf Nullmengen).

Satz 1.19
Sei $X_1 : \Omega_1 \to \mathbb{R}$ und $\mathbb{P}_{X_1,\Theta}$ eine k-parametrige Exponentialfamilie mit Funktionen c, b, γ und t. Dann gilt: $\mathbb{P}_{X,\Theta}$ mit $X = (X_1, \dots, X_n)$ u.i.v. ist ebenfalls k-parametrige Exponentialfamilie mit Funktionen

$$\begin{aligned} c_n(\vartheta) &= c(\vartheta)^n \\ b_n(x_1, \dots, x_n) &= \prod_{i=1}^n b(x_i) \\ \gamma_n(\vartheta) &= \gamma(\vartheta) \\ t_n(x_1, \dots, x_n) &= \sum_{i=1}^n (\, t_1(x_i), \dots, t_k(x_i)\,) \end{aligned}$$

Definition 1.20 (natürliche Parametrisierung)
Sei $\gamma = \gamma(\vartheta)$ und $\tilde{c}(\gamma)$ derart, daß gilt:

$$\tilde{c}(\gamma) = c(\vartheta).$$

Dann heißt die k-parametrische Exponentialfamilie in kanonischer Form dargestellt mit Dichte

$$\tilde{f}(x;\gamma) = \tilde{c}(\gamma)\ \exp(\gamma^\top t(x)).$$

Satz 1.21
Der natürliche Parameterraum ist konvex.

Satz 1.22
Sei $\mathbb{P}_{X,\Theta}$ eine k-parametrige Exponentialfamilie in γ und $T = t(x)$. Dann bildet $\mathbb{P}_{T,\Theta}$ ebenfalls eine k-parametrige Exponentialfamilie in γ und der Identität.

Satz 1.23
Sei $\mathbb{P}_{X,\Theta}$ eine k-parametrige Exponentialfamilie in γ und $T = t(X)$. Dann besitzt die Statistik T Momente beliebig hoher Ordnung, diese sind als Funktion von γ beliebig oft nach γ differenzierbar; die Ableitungen dürfen durch Differentiation unter dem betreffenden Integral gebildet werden.

1.4 Invariante Verteilungsfamilien

Definition 1.24 (Invarianz)
Sei $X = (X_1, \ldots, X_n)$ eine Stichprobe und $\mathbb{P}_{X,\Theta} = \{\mathbb{P}_{X,\vartheta} \mid \vartheta \in \Theta\}$ die Familie von Verteilungen von X. Sei G eine Klasse von Transformationen des Stichprobenraumes

$$g : \mathcal{X} \to \mathcal{X},$$

wobei g bijektiv und g, g^{-1} meßbar sind. Dann ist die Verteilung der ZV $g(X)$

$$\mathbb{P}_{g(X),\vartheta}(B) = \mathbb{P}_{X,\vartheta}(g^{-1}(B)) \quad \forall\, B \in \sigma(\mathcal{X})$$

und $\mathbb{P}_{g(X),\Theta}$ die zugehörige Verteilungsannahme. Diese heißt invariant gegenüber der Klasse G von Transformationen, falls

$$\mathbb{P}_{g(X),\Theta} = \mathbb{P}_{X,\Theta} \quad \forall\, g \in G.$$

1.5 Suffizienz

Definition 1.25 (Suffiziente Statistik)
Sei $X = (X_1, \ldots, X_n)$ eine Stichprobe und $\mathbb{P}_{X,\Theta} = \{\mathbb{P}_{X,\vartheta} \mid \vartheta \in \Theta\}$ eine Verteilungsannahme sowie $t : \mathcal{X} \to \mathbb{R}^k$ eine Statistik. Dann heißt $T = t \circ X$ suffizient für $\mathbb{P}_{X,\Theta}$ (oder einfach für ϑ), wenn die bedingte Verteilung

$$\mathbb{P}_{X|T=t(x)}$$

definiert durch

$$\mathbb{P}_{X|T=t(x)}(B) = \frac{\mathbb{P}(X^{-1}(B) \cap (t \circ X)^{-1}(t(x)))}{\mathbb{P}((t \circ X)^{-1}(t(x)))}$$

unabhängig von ϑ ist (deren Existenz vorausgesetzt; dies ist mit $\mathcal{X} = \mathbb{R}^n$ aber der Fall).

Definition 1.26 (bedingter Erwartungswert)
Seien X und Y Zufallsvariablen mit Randverteilungen $\mathbb{P}_X = f_X \odot \nu_X$ und $\mathbb{P}_Y = f_Y \odot \nu_Y$ sowie bedingte Verteilungen $\mathbb{P}_{X|Y=y} = f_{X,Y=y} \odot \nu_X$. Dann ist der bedingte Erwartungswert von $g(X)$ gegeben $Y = y$ als Funktion von y definiert durch

$$\begin{aligned}\mathbb{E}(g(X)|Y=y) &= \int g(X)\; d\mathbb{P}_{X|Y=y} \\ &= \int g(X) f_{X|Y=y}(x|y)\; d\nu_X\end{aligned}$$

Satz 1.27 (Satz vom iterierten Erwartungswert)

$$\mathbb{E}_Y(\mathbb{E}_{X|Y}(X|Y)) = \mathbb{E}_X(X)$$

Satz 1.28 (Faktorisierungssatz von Fisher und Neyman)
Sei $X = (X_1, \ldots, X_n)$ eine Stichprobe und $\mathbb{P}_{X,\Theta} = \{f(x;y) \mid \vartheta \in \Theta\}$ eine Verteilungsannahme, so gilt für eine Staistik $T = t \circ X$ $(t : \mathbb{R}^n \to \mathbb{R}^k)$

$$\begin{aligned}&T \text{ ist suffizient für } \vartheta \Leftrightarrow \\ &\exists \begin{array}{llll} g: & \Theta & \times & \mathbb{R}^k \to \mathbb{R} \\ h: & & & \mathbb{R}^n \to \mathbb{R} \end{array}\Bigg\}\begin{array}{c}\text{meßbar}\\ \text{so daß}\end{array} \\ &f(x;\vartheta) = g(\vartheta, t(x))\; h(x).\end{aligned}$$

Satz 1.29
Sei $\mathbb{P}_{X,\Theta}$ eine k-parametrige Exponentialfamilie mit Dichten

$$\begin{aligned}& f_X(x;\vartheta) = c(\vartheta)\, b(x) \exp(\gamma(\vartheta)^\top t(x)) \\ \Rightarrow\quad & t \odot X \text{ ist suffizient für } \vartheta \in \Theta.\end{aligned}$$

1.6 Die Fisher-Information

Definition 1.30 (Fisher Regularität)
Sei $\mathbb{P}_{X,\Theta} = \{f(x;\vartheta) \mid \vartheta \in \Theta \subseteq \mathbb{R}\}$ mit den folgenden Regularitätseigenschaften

R1) Die Dichten $f(x;\vartheta)$ existieren (d.h. $\mathbb{P}_{X,\vartheta} = f \odot \nu$).

R2) Θ ist offenes Intervall.

R3) Der Träger $C := \{x \mid f(x;\vartheta) > 0\}$ ist unabhängig von ϑ.

R4)

$$\begin{aligned}\int \frac{\partial f(x;\vartheta)}{\partial\vartheta} d\nu &= \frac{\partial}{\partial\vartheta}\int f(x;\vartheta) d\nu = 0 \\ \int \frac{\partial^2 f(x;\vartheta)}{\partial^2\vartheta} d\nu &= \frac{\partial}{\partial\vartheta}\int \frac{\partial f(x;\vartheta)}{\partial\vartheta} d\nu = 0 \\ \forall x \in C.\end{aligned}$$

Dann heißt die Verteilungsfamilie $\mathbb{P}_{X,\Theta}$ Fisher-regulär.

Definition 1.31 (Scorefunktion)

$$\mathrm{S}(\vartheta;x) = \frac{\partial \log f(x;\vartheta)}{\partial\vartheta} = \frac{\frac{\partial f(x;\vartheta)}{\partial\vartheta}}{f(x;\vartheta)}$$

heißt Scorefunktion der Beobachtung x und die Zufallsvariable $\mathrm{S}(\vartheta;X)$ heißt Scorefunktion der Stichprobe X.

Satz 1.32

$$\mathbb{E}_\vartheta \mathrm{S}(\vartheta; X) = 0$$
$$\mathbb{V}_\vartheta \mathrm{S}(\vartheta; X) = -\mathbb{E}_\vartheta \frac{\partial}{\partial\vartheta} \mathrm{S}(\vartheta; X)$$

Definition 1.33 (Fisher-Information)
Sei $\mathbb{P}_{X,\Theta} = \{f(x;\vartheta) \mid \vartheta \in \Theta\}$ Fisher-regulär und $x \in C$ (dem Träger von f). Dann heißt

$$\mathrm{I}(\vartheta; x) := -\frac{\partial}{\partial\vartheta} \mathrm{S}(\vartheta; x)$$

die Fisher-Information der Beobachtung x über den Parameter ϑ. Der Erwartungswert der Zufallsvariable $\mathrm{I}(\vartheta; X)$ heißt Erwartete Fisher-Information $\mathrm{I}(\vartheta)$ der Stichprobe X:

$$\begin{aligned} \mathrm{I}_X(\vartheta) &= \mathbb{E}_\vartheta \mathrm{I}(\vartheta; X) \\ &= \mathbb{E}_\vartheta - \frac{\partial}{\partial\vartheta} \mathrm{S}(\vartheta; X) = \mathbb{V}_\vartheta \mathrm{S}(\vartheta; X) \end{aligned}$$

Satz 1.34
Sei $X = (X_1, \ldots, X_n)$ u.i.v. und $\mathrm{I}_{X_i}(\vartheta)$ die erwartete Fisher-Information von X_i. Dann gilt:

$$\mathrm{I}_X(\vartheta) = \sum_{i=1}^{n} \mathrm{I}_{X_i}(\vartheta)$$

Satz 1.35 (Cramèr-Rao-Ungleichung)
Sei $\mathbb{P}_{X,\Theta}$ Fisher regulär und $t : X \to \mathbb{R}$ ein erwartungstreuer Schätzer für $\vartheta \in \Theta$. Falls für die ZV $T = t \circ X$ gilt

$$\mathrm{Cov}(T, \mathrm{S}(\vartheta; X)) = 1$$

so folgt

$$\mathbb{V}_\vartheta(T) \geq \frac{1}{\mathrm{I}_X}(\vartheta).$$

Kapitel 2

Punktschätzungen

2.1 Maximum-Likelihood-Schätzung

Definition 2.1 (Likelihood)
Sei $X = (X_1, \ldots, X_n)$ mit Verteilungsannahme $\mathbb{P}_{X,\Theta} = \{f(x;\vartheta) \mid \vartheta \in \Theta\}$ (d.h., die Dichten existieren), dann heißt die Funktion

$$\begin{aligned} \mathcal{L} &: \Theta \times \mathcal{X} \to \mathbb{R}^+ \\ \vartheta &\mapsto \mathcal{L}(\vartheta; x) = f(x;\vartheta) \end{aligned}$$

die Likelihoodfunktion für ein festes $x \in \mathcal{X}(= \mathbb{R}^n$ üblicherweise).

Definition 2.2 (Maximum- Likelihood- Schätzer)
Sei $X = (X_1, \ldots, X_n)$ und $\mathbb{P}_{X,\Theta}$ die zugehörige Verteilungsannahme (siehe Def. 2.1). Dann heißt der Schätzer

$$\begin{aligned} t_{\mathrm{ML}} &: \mathbb{R}^n \to \mathbb{R}^k \\ x &\mapsto t_{\mathrm{ML}}(x) = \widehat{\vartheta}_{\mathrm{ML}} := \operatorname*{argmax}_{\vartheta\in\Theta} \mathcal{L}(\vartheta; X) \end{aligned}$$

der Maximum-Likelihood-Schätzer für $\vartheta \in \Theta \subseteq \mathbb{R}^k$.

Definition 2.3 (Log- Likelihood)

$$\ell(\vartheta; x) = \log \mathcal{L}(\vartheta; x) = \log f(x;\vartheta)$$

heißt Log- Likelihoodfunktion.

Satz 2.4 (Konsistenz des ML-Schätzers)
Sei $X = (X_1, \ldots, X_n)$ mit R1-R4 und C1,C2. Dann existiert eine Folge von Statistiken $\widehat{\vartheta}_n = t_{\mathrm{ML}}(X)$, definiert über $\mathrm{S}(\widehat{\vartheta}_n; x = (x_1, \ldots, x_n) = 0)$ und $\mathrm{I}(\widehat{\vartheta}_n, (x_, \ldots, x_n)) > 0$ (also potentiell lokale Maxima der Likelihood) so daß gilt:

$$\widehat{\vartheta}_n \xrightarrow{\mathbb{P}} \vartheta \qquad (X \sim \mathbb{P}_{X,\vartheta})$$

($\widehat{\vartheta}_n$ ist schwach konsistent).

Korollar 2.5
Es gelten die Annahmen von Satz 2.4; wenn $\mathrm{S}(\widehat{\vartheta}_n; x) = 0$ <u>eindeutige</u> Nullstelle ist, so gilt:

i) $\widehat{\vartheta}_n \xrightarrow{\mathbb{P}} \vartheta$

i)) $\widehat{\vartheta}_n \overset{\text{f.s.}}{\to} \widehat{\vartheta}_{\text{ML}}$
(d.h. die Lösung der Scoregleichung führt zum Maximum-Likelihood-Schätzer mit Wahrscheinlichkeit 1 für $n \to \infty$).

Satz 2.6 (Invarianz des ML-Schätzers)
Seien die Bedingungen von Satz 2.4 erfüllt und $g(\vartheta)$ eine stetige parametrische Funktion. Dann gilt: $\widehat{g(\vartheta)}_n = g(\widehat{\vartheta}_n)$ ist ein konsistenter Schätzer für $g(\vartheta)$.

Satz 2.7 (Verteilung der Scorefunktion)
Sei $X = (X_1, \dots, X_n)$ u.i.v. und $\mathrm{S}(\vartheta; X)$ die zugehörige Scorefunktion. Dann gilt:

$$\frac{1}{\sqrt{n}} \mathrm{S}(\vartheta; X) \overset{D}{\to} N(0, \mathrm{I}_{X_1}(\vartheta))$$

Satz 2.8 (Quadratische Approximation der Log-Likelihood)

$$\ell(\vartheta; x) \approx \ell(\widehat{\vartheta}_{\text{ML}}; x) - \frac{1}{2} \mathrm{I}(\widehat{\vartheta}_{\text{ML}})(\vartheta - \widehat{\vartheta}_{\text{ML}})^2$$

Satz 2.9 (asymptotische Normalität des ML-Schätzers)
Sei $X = (X_1, \dots, X_n)$ u.i.v. mit Verteilungsannahme $\mathbb{P}_{X,\Theta} = \{f(x; \vartheta) \mid \vartheta \in \Theta\}$ (Fisher-regulär). Weiterhin gelte:
C1, C2 sowie

C3) $\forall x \in C$ (Träger von f) ist $f(x; \vartheta)$ mindestens dreimal stetig differenzierbar.

C3) Für den wahren Parameterwert ϑ_0 existiert eine Funktion $M_{\vartheta_0}(x)$ und eine Konstante $c(\vartheta_0)$ so daß

$$\left| \frac{\partial^3 \ell(\vartheta; x)}{\partial^3 \vartheta} \right| \leq M_{\vartheta_0}(x) \quad \forall x \in C, \ (\vartheta - \vartheta_0) < x(\vartheta_0)$$

sowie

$$\mathbb{E}_{\vartheta_0}(M_\vartheta(X)) < \infty.$$

Dann gilt für jede konsistente Folge von ZV $\widehat{\vartheta}_k = t_{\text{Score}}(x)$ (siehe Satz 2.4)

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) \overset{D}{\to} N(0, \frac{1}{\mathrm{I}_{X_1}(\vartheta_0)})$$

wobei $0 < \mathrm{I}_{X_1}(\vartheta_0) < \infty$

Satz 2.10
Sei $\widehat{\vartheta}_n$ Schätzer für ϑ und g eine diffbare parametrische Funktion. Dann gilt

$$\sqrt{n} \left(g(\widehat{\vartheta}_n) - g(\vartheta) \right) \overset{D}{\to} N \left(0, \left(\frac{\partial g(\vartheta)}{\partial \vartheta} \right)^2 \mathrm{I}_{X_1}^{-1}(\vartheta) \right)$$

Satz 2.11
Unter den Annahmen von Satz 2.9 gilt ebenso

$$\begin{aligned}
\sqrt{n}(\widehat{\vartheta}_n - \vartheta) &\overset{D}{\to} N\left(0, \mathrm{I}_{X_1}(\widehat{\vartheta}_n)^{-1}\right) \\
\sqrt{n}(\widehat{\vartheta}_n - \vartheta) &\overset{D}{\to} N\left(0, \mathrm{I}(\vartheta, X)^{-1}\right) \\
\sqrt{n}(\widehat{\vartheta}_n - \vartheta) &\overset{D}{\to} N\left(0, \mathrm{I}(\widehat{\vartheta}_n, X)^{-1}\right)
\end{aligned}$$

2.2 M-Schätzung

Definition 2.12 (M-Schätzer)
Sei $X = (X_1, \dots, X_n)$ u.i.v. mit $X_i \sim \mathrm{F}$; $X_i \in \mathbb{R}^p$. Sei $\psi : \mathbb{R}^p \times \mathbb{R}^k \to \mathbb{R}$ eine Funktion, so daß für den wahren Parameter ϑ gilt:

$$\mathbb{E}\ \psi(X_1, \vartheta) = \int \psi(x, \vartheta)\ d\mathrm{F}(x) = 0$$

Dann heißt $\widehat{\vartheta}$ mit

$$\sum_{i=1}^{n} \psi\left(X_i, \widehat{\vartheta}\right) = 0$$

M-Schätzer.

Satz 2.13
Sei $X = (X_1, \dots, X_n)$; $\psi(x, \vartheta)$ monoton in ϑ und in ϑ stetig. Dann gilt für die Lösung der Schätzgleichung $\widehat{\vartheta}_n$:

$$\sum_{i=1}^{n} \psi(x_i, \vartheta) \stackrel{!}{=} 0$$

$\widehat{\vartheta}_n \to \vartheta$.

Satz 2.14 (asymptotische Normalität des M-Schätzers)

$$\sqrt{n}\left(\widehat{\vartheta}_n - \vartheta_0\right) \xrightarrow{D} N(0, V(\vartheta_0))$$

Satz 2.15 (Sandwich Schätzer)

$$\widehat{V}(\widehat{\vartheta}_n) = \widehat{A}(\widehat{\vartheta}_n)^{-1} \widehat{B}(\widehat{\vartheta}_n) \widehat{A}(\widehat{\vartheta}_n)^{-1^\top} \xrightarrow{\mathbb{P}} V(\vartheta_0)$$

mit

$$\begin{aligned} \widehat{A}(\widehat{\vartheta}_n) &= \frac{1}{n} \sum_{i=1}^{n} -\frac{\partial \psi(X_i, \widehat{\vartheta}_n)}{\partial \widehat{\vartheta}_n} \\ \widehat{B}(\widehat{\vartheta}_n) &= \frac{1}{n} \sum_{i=1}^{n} \psi(X_i, \widehat{\vartheta}_n) \psi(X_i, \widehat{\vartheta}_n)^\top \end{aligned}$$

Kapitel 3

Inferenzkonzepte

3.1 Parametertests

Definition 3.1 (Testproblem)
Sei $\vartheta \in \mathbb{R}^k$ Parameter und $\Theta \subset \mathbb{R}^k$ der zugehörige Parameterraum. Dann heißt

$$H_0 : \vartheta \in \Theta_0 \quad \text{vs.} \quad H_1 : \vartheta \in \Theta_1$$

mit $\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset \quad \Theta_0 \neq \emptyset, \ \Theta_1 \neq \emptyset$
Statistisches Testproblem. H_0 heißt (Null)hypothese, H_1 Alternative.

Definition 3.2 (einfache, einseitige und zweiseitige Hypothesen)
$h_0 : \ \vartheta \in \Theta_0 \quad \text{vs.} \quad H_1 : \ \vartheta \in \Theta_1$

$|\Theta_0| = 1 \ \Leftrightarrow$**:** „einfache Hypothese“, sonst: „zusammengesetzte Hypothese“

$|\Theta_1| = 1 \ \Leftrightarrow$**:** „einfache Alternative“, sonst: „zusammengesetze Alternative“

Bei skalaren Parametern unterscheidet man noch

einseitiges Testproblem	$H_0 : \ \vartheta \leq \vartheta_0$	vs.	$H_1 : \ \vartheta > \vartheta_0$	„rechtsseitige Alternative“
	$H_0 : \ \vartheta \geq \vartheta_0$	vs.	$H_1 : \ \vartheta < \vartheta_0$	„linksseitige Altervative“
zweiseitiges Testproblem	$H_0 : \ \vartheta = \vartheta_0$	vs.	$H_1 : \ \vartheta \neq \vartheta_0$	

Definition 3.3 (statistischer Test)

$$\phi : \mathcal{X} \to \{0,1\} \quad \text{(meßbar)}$$

heißt nichtrandomisierter Test. ($\mathcal{X}$ Stichprobenraum)

$$\phi_R : \mathcal{X} \to [0,1] \quad \text{(meßbar)}$$

heißt randomisierter Test.

Definition 3.4 (kritischer Bereich)

$$K := \{x \in \mathcal{X} \mid \varphi(x) = 1\}$$

heißt kritischer Bereich oder Ablehnbereich.
$\overline{K}$ heißt Annahmebereich.

Definition 3.5 (Fehlerarten)

$$\begin{aligned} \vartheta \in \Theta_0 \ , \ \phi(x) = 1 \ \ldots \ \ & \underline{\text{Fehler erster Art}} \\ \vartheta \in \Theta_1 \ , \ \phi(x) = 0 \ \ldots \ \ & \underline{\text{Fehler zweiter Art}} \end{aligned}$$

Definition 3.6 (Gütefunktion)

$$\begin{aligned} \beta_\phi(\vartheta) \ &:= \ \mathbb{P}_\vartheta(\phi(X) = 1) \\ &= \ \mathbb{E}_\vartheta \phi(X) = \int \phi(x) \ d\mathbb{P}_{X,\vartheta}(x) \end{aligned}$$

heißt Gütefunktion oder power des Tests ϕ. Diese Funktion beschreibt die Wahrscheinlichkeit, die Nullhypothese $H_0 : \vartheta \in \Theta_0$ abzulehnen für eine Stichprobe $X = (X_1, \ldots X_n)$ aus der Verteilung $\mathbb{P}_{X,\vartheta}$.

Definition 3.7 (Niveau-α-Test)
Ein Test ϕ mit der Eigenschaft

$$\beta_\phi(\vartheta) \leq \alpha \quad \forall\, \vartheta \in \Theta_0$$

heißt <u>Niveau-α-Test</u>. D.h., die Wahrscheinlichkeit für den Fehler erster Art ist maximal $\alpha \in (0, 1)$.

Definition 3.8 (Macht, Power)

$$\beta_\phi(\vartheta) \quad \forall\, \vartheta \in \Theta_1$$

heißt Macht oder Powerfunktion des Tests.

Definition 3.9 (Bester Test)
Seien ϕ und ϕ^* Niveau-α-Test für das TP H_0 vs. H_1. Dann heißt ϕ mindestens so gut wie ϕ^* wenn gilt:

$$\beta_\phi(\vartheta) \geq \beta_{\phi^*}(\vartheta) \quad \forall\, \vartheta \in \Theta_1.$$

Gilt diese Beziehung für alle beliebigen Niveau-α-Tests ϕ^*, so heißt ϕ <u>gleichmäßig bester</u> Test (uniformly most powerful, UMP).

Definition 3.10 (Unverfälschter Test)
Ein Niveau-α-Test ϕ heißt <u>unverfälscht</u>, wenn gilt

$$\beta_\phi(\vartheta_0) \leq \beta_\phi(\vartheta_1) \quad \forall\, \vartheta_0 \in \Theta_0 \text{ und } \vartheta_1 \in \Theta_1$$

Ein Niveau-α-Test ϕ heißt gleichmäßig bester unverfälschter Test, wenn er gleichmäßig bester Test unter allen unverfälschter Niveau-α-Test ist.

Definition 3.11 (Teststatistik)
Für Tests der Form

$$\phi(x) = \begin{cases} 1 & T(X) > k \\ 0 & T(X) \leq k \end{cases}$$

heißt die Statistik $T : \mathcal{X} \to \mathbb{R}$ <u>Teststatistik</u> und deren Verteilung unter den Bedingungen der Nullhypothese $\mathbb{P}_{\Theta_0}(T(X) \leq k)$ <u>Prüfverteilung</u>.

Definition 3.12 (Konfidenzintervall)
Die Funktion $C : \mathcal{X} \to \mathbb{R}^2, \quad x \mapsto C(X) = [\underline{C}(X), \overline{C}(X)]$ heißt Konfidenzintervall für ϑ zum Niveau $1 - \alpha$ falls gilt

$$\mathbb{P}_\vartheta(\vartheta \in C(X)) = \mathbb{P}_\vartheta(\underline{C}(X) \leq \vartheta \leq \overline{C}(X)) \geq 1 - \alpha \quad \forall\, \vartheta \in \Theta.$$

Definition 3.13 (Gütekriterien für Konfidenzintervalle)
Ein Konfidenzintervall $C(X)$ heißt unverzerrt (unbiased), wenn gilt

$$\mathbb{P}_\vartheta(\vartheta^* \in C(X)) \leq \mathbb{P}_\vartheta(\vartheta \in C(X)) \qquad \forall \vartheta^k \neq \vartheta$$

Ein Konfidenzintervall $C(X)$ ist besser als ein Konfidenzintervall $\tilde{C}(X)$ (beide zum Niveau $1-\alpha$) wenn

$$\mathbb{E}_\vartheta(\overline{C}(X) - \underline{C}(X)) < \mathbb{E}_\vartheta(\overline{\tilde{C}}(X) - \underline{\tilde{C}}(X))$$

d.h., wenn die erwartete Länge kleiner ist.

3.2 Wald-, Likelihood-Quotienten- und Score-Test

Satz 3.14 (Wald-Test)
Sei $X = (X_1, \dots, X_n)$ u.i.v. und $\widehat{\vartheta}_n$ ein Schätzer für ϑ mit der Eigenschaft

$$\sqrt{n}\left(\widehat{\vartheta}_n - \vartheta\right) \xrightarrow{D} N\left(0, \mathrm{I}_{X_1}^{-1}(\vartheta)\right)$$

Dann ist der Test

$$\phi_W(X) = \begin{cases} 1 & |W(X)| \geq u_{1-\alpha/2} \\ 0 & \text{sonst} \end{cases}$$

mit $W(X) = \sqrt{n}\frac{|\widehat{\vartheta}_n - \vartheta_0|}{\sqrt{\mathrm{I}_{X_1}(\vartheta_0)}}$ (Wald-Teststatistik) ein approximativer Niveau-α-Test für das Testproblem

$$H_0:\ \vartheta = \vartheta_0 \quad \text{vs.} \quad H_1:\ \vartheta \neq \vartheta_0$$

Definition 3.15 (Likelihood-Quotienten-Test)
Ein Test basierend auf der Teststatistik

$$\mathrm{LQ}(X) = \frac{\mathcal{L}(\widehat{\vartheta}_n; X)}{\max\limits_{\vartheta \in \Theta_0} \mathcal{L}(\vartheta; X)}$$

heißt Likelihood-Quotienten-Test für

$$H_0;\ \vartheta \in \Theta_0 \quad \text{vs.} \quad H_1:\ \vartheta \in \Theta_1$$

Satz 3.16 (Verteilung von LQ)

$$2\log\left(\mathrm{LQ}(X)\right) \xrightarrow{D} \chi_1^2$$

Definition 3.17 (Score-Test, LM-Test)
Ein Test basierend auf der Teststatistik

$$\mathrm{LM}(X) = \frac{\mathrm{S}(\vartheta_0, X)}{\sqrt{n \cdot \mathrm{I}_{X_1}(\vartheta_0)}}$$

heißt Score-Test oder Lagrange-Multiplier-Test für $H_0:\ \vartheta = \vartheta_0 \quad \text{vs.} \quad H_1:\ \vartheta \neq \vartheta_0$.

Satz 3.18 (Verteilung von LM)

$$\mathrm{LM}(X) \xrightarrow{D} N(0,1)$$

Satz 3.19 (asymptotische Äquivalenz von W und LM)

$$\mathrm{LM}(X) - W(X) \xrightarrow{\mathbb{P}} 0$$

3.3 Bayes-Inferenz

Definition 3.20 (Priori-Verteilung)
Sei $\vartheta \sim \mathbb{P}_\vartheta$ mit Dichte $f_\vartheta(\vartheta)$ eine ZV. $\mathbb{P}_\vartheta$ heißt Priori-Verteilung.

Definition 3.21 (Posteriori-Verteilung)
Sei $X \sim \mathbb{P}_{X|\vartheta}$ eine ZV mit auf $\vartheta = \vartheta$ bedingter Verteilung und $X = x$ eine Realisation von X. Dann heisst die Dichte der bedingten Verteilung von ϑ gegeben $X = x$

$$f_{\vartheta|X=x}(\vartheta|X=x) = \frac{f_{X|\vartheta}(x|\vartheta) \cdot f_\vartheta(\vartheta)}{\int f_{X|\vartheta}(x|\vartheta)\, d\nu(\vartheta)}$$

Posteriori-Dichte und die bedingte Verteilung $\vartheta|X = x$ Posteriori-Verteilung.

Definition 3.22 (Posteriori-Punktschätzer)

Posteriori-Erwartungswert:

$$\mathbb{E}(\vartheta|X=x) = \int \vartheta\, f_{\vartheta|X}(\vartheta|X=x)\, d\nu(\vartheta)$$

Posteriori-Modus:

$$\mathrm{Mod}(\vartheta|X=x) = \underset{\vartheta}{\operatorname{argmax}}\, f_{\vartheta|X}(\vartheta|X=x)$$

Posteriori-Median

$$\mathrm{Median}(\vartheta|X=x)$$

Definition 3.23 (Kredibilitätsintervall)
Ein Intervall $[\underline{b}, \overline{b}]$ mit

$$\int_{[\underline{b},\overline{b}]} f_{\vartheta|x}(\vartheta|x)\; d\nu(\vartheta) = 1-\alpha$$

heißt (1-α)-Kredibilitätsintervall. Gilt weiterhin

$$f_{\vartheta|x}(\vartheta|X=x) \geq f_{\vartheta|x}(\vartheta^*|X=x)$$

$$\forall \vartheta \in [\underline{b}, \overline{b}] \text{ und } \vartheta^* \notin [\underline{b}, \overline{b}]$$

so heißt $[\underline{b}, \overline{b}]$
„highest posterior density intervall" HPD- Intervall.

Satz 3.24

$$f_\vartheta(\vartheta) = \text{const} \;\Rightarrow$$

$$\mathrm{Mod}(\vartheta|X=x) = \widehat{\vartheta}_{\mathrm{ML}}$$

Satz 3.25
Sei $\vartheta \in \Theta$, Θ höchstens abzählbar, also $\Theta = \{\vartheta_1, \vartheta_2, \ldots\}$. Sei ϑ_t der wahre Parameter (also $X \sim \mathbb{P}_{X,\vartheta_t}$). Dann gilt

$$\lim_{n\to\infty} f(\vartheta_t \mid X=x) = 1 \quad \text{und}$$

$$\lim_{n\to\infty} f(\vartheta_i \mid X=x) = 0 \quad \forall i \neq t$$

Literaturverzeichnis

Leonhard Held. *Methoden der statistischen Inferenz. Likelihood und Bayes.* Spektrum Akademischer Verlag, Heidelberg, 2008.

Erich Lehmann. *Elements of Large-Sample Theory.* Springer, Berlin, 2001.

Bernhard Rüger. *Test- und Schätztheorie (Band I: Grundlagen).* Oldenbourg, München, 1999.