**Statistik 1**
Prof. Dr. Jan Johannes
Sergio Brenner Miguel
Sommersemester 2019

UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386

# 10. Übungsblatt

| Aufgabe 1 | Aufgabe 2 | Aufgabe 3 | Aufgabe 4 | Vorbereitung | $\Sigma$ |
|---|---|---|---|---|---|
| | | | | | |

**Aufgabe 1 (4 Punkte)**

Für festes $\lambda > 0$ sei $\mathbb{P}_\Theta := \{\mathbb{P}_\theta \,|\theta \in \Theta = \mathbb{R}\}$ die Familie geshifteter Exponentialverteilungen mit Dichten $\mathbb{f}_\theta(x) = \lambda \exp(-\lambda(x-\theta))\mathbb{1}_{(\theta,\infty)}(x), x \in \mathbb{R}$.

a) Zeigen Sie, dass $T(x) = \min_{i\in[\![1,n]\!]} x_i$ erschöpfend für $\mathbb{P}^n_\Theta = \{\mathbb{P}^n_\theta \,|\theta \in \Theta\}$ ist.

b) Zeigen Sie, dass $F^T_\theta(t) := \mathbb{P}^n_\theta[T \le t] = (1 - \exp(-\lambda n(t-\theta)))\mathbb{1}_{(\theta,\infty)}(t)$ ist und bestimmen sie die korrespondieren Dichte $\mathbb{f}^T_\theta$.

c) Verwenden Sie den Aufgabenteil b) um zu zeigen, dass $T(x) = \min_{i\in[\![1,n]\!]} x_i$ vollständig für $\mathbb{P}^n_\Theta = \{\mathbb{P}^n_\theta \,|\theta \in \Theta\}$ ist.

d) Ermitteln Sie den besten translations-äquivarianten Schätzer bezüglich

   i) der quadratischen Verlustfunktion $\nu(\theta, e) = (\theta - e)^2$ und

   ii) der absoluten Verlustfunktion $\nu(\theta, e) = |\theta - e|$

   unter Verwendung des Korollar 17.08 und Beispiel 17.10.

**Aufgabe 2 (4 Punkte)**

Es sei $\mathbb{P}^n_\Theta$ eine Lokations-Familie mit Likelihood-Funktion $L(\theta, x) = \prod_{i=1}^n \mathbb{f}(x_i - \theta), x \in \mathbb{R}^n, \theta \in \Theta = \mathbb{R}$, bezüglich einer Lebesguedichte $\mathbb{f} : \mathbb{R} \to [0,\infty)$. Sei $\theta_0 \in \mathbb{R}$ beliebig.

a) Zeigen Sie die Hellinger-Differenzierbarkeit der Verteilungsfamilie $\mathbb{P}^n_\Theta$ , wenn $\mathbb{f} \in \mathcal{C}^\infty(\mathbb{R}), \mathbb{f} > 0$ und $\int |\,\mathbb{f}'(x)|^2/\,\mathbb{f}(x)dx < \infty$ gilt und bestimmen Sie die zugehörige Fisher-Information.

b) Die Laplace-Verteilung, $\mathbb{f}(x) = \frac{1}{2}\exp(-|x|/2)$, erfüllt nicht das Kriterium $\mathbb{f} \in \mathcal{C}^\infty(\mathbb{R})$. Zeigen Sie, dass dennoch für $n = 1$ die Verteilungsfamilie Hellinger-differenzierbar ist mit Scorefunktion $\dot{\ell}(\theta_0, x) = \operatorname{sgn}(x - \theta_0)/\sigma$ wobei $\operatorname{sgn}(x)$ das Signum einer reellen Zahl $x$ bezeichnet.

**Aufgabe 3 (4 Punkte)**

Für zwei Wahrscheinlichkeitsmaße $\mathbb{P}$ und $\mathbb{Q}$ auf $(\mathcal{X}, \mathscr{X})$ definieren wir die Kullback-Leibler-Divergenz als

$$\mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q}) = \begin{cases} \int_{\mathcal{X}} \log\left(\frac{d\mathbb{P}}{d\mathbb{Q}}(x)\right) \mathbb{P}(dx), & \text{falls } \mathbb{P} \ll \mathbb{Q}, \\ +\infty, & \text{sonst.} \end{cases}$$

.

a) Verwenden Sie ohne Beweis, dass $\log(x) \leq x - 1$ für alle $x > 0$ gilt und zeigen Sie, dass

$$\mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q}) \geq 0 \text{ und } \mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q}) = 0 \Leftrightarrow \mathbb{P} = \mathbb{Q}\,.$$

b) Zeigen Sie, dass für $\mathbb{P} \ll \mathbb{Q}$ und $\mathbb{Q} \ll \mathbb{P}$ gilt

$$\mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q}) = -\,\mathbb{E}_{\mathbb{P}}\left[\log\left(\frac{d\,\mathbb{Q}}{d\,\mathbb{P}}\right)\right].$$

Geben Sie ein Paar $(\mathbb{P}, \mathbb{Q})$ von äquivalenten Wahrscheinlichkeitsmaßen an, derart, dass

$$\mathrm{KL}(\mathbb{P} \,|\, \mathbb{Q}) \neq \mathrm{KL}(\mathbb{Q} \,|\, \mathbb{P})$$

gilt.

c) Zeigen Sie, dass für Produktmaße die Kullback-Leibler-Divergenz additiv ist, das heißt

$$\mathrm{KL}\left(\bigotimes_{i=1}^n \mathbb{P}_i, \bigotimes_{i=1}^n \mathbb{Q}_i\right) = \sum_{i=1}^n \mathrm{KL}(\mathbb{P}_i, \mathbb{Q}_i).$$

*Hinweis:* Verwenden Sie ohne Beweis, dass für $\mathbb{Q} := \bigotimes_{i=1}^n \mathbb{Q}_i, \mathbb{P} := \bigotimes_{i=1}^n \mathbb{P}_i$ gilt $\mathbb{P} \ll \mathbb{Q} \Leftrightarrow \forall i \in [\![1, n]\!] : \mathbb{P}_i \ll \mathbb{Q}_i$ und $\frac{d\mathbb{P}}{d\mathbb{Q}}(x) = \prod_{i=1}^n \frac{d\mathbb{P}_i}{d\mathbb{Q}_i}(x_i)$.

d) Es sei $\mathbb{P}_\Theta$ eine natürliche Exponentialfamile und $\theta_0 \in \Theta \subset \mathbb{R}^k$ ein innerer Punkt. Zeigen Sie, dass $\mathrm{KL}(\mathbb{P}_{\theta_0} \,|\, \mathbb{P}_\theta) = A(\theta) - A(\theta_0) + \langle \dot{A}(\theta_0), \theta_0 - \theta \rangle$ gilt.

**Aufgabe 4 (4 Punkte)**

Ein Teich enthält eine unbekannte Anzahl $\theta$ an Karpfen. Zur Schätzung von $\theta$ werden zunächst $k$ Fische gefangen, markiert und wieder freigelassen. Wenn sich die markierten Fische wieder gut verteilt haben, werden $n$ Fische gefangen, von denen $x$ markiert sind.

a) Modellieren Sie die Schätzung des Fischbestandes $\theta$ durch ein statistisches Experiment mit hypergeometrischen Verteilungen und bestimmen Sie einen Maximum-Likelihood-Schätzer. Zeigen Sie, dass für den Fall $x = 0$ kein Maximum-Likelihood-Schätzer existiert.

Betrachten Sie nun die Familie $\mathbb{P}_\Theta = \{\mathbb{P}_\theta \,|\theta \in \mathbb{R}\}$ der geshifteten Laplace-Verteilungen mit Dichten $\mathbb{f}_\theta(x) = \frac{1}{2}\exp(-|x-\theta|), x \in \mathbb{R}$.

b) Zeigen Sie, dass der empirisiche Median einer Stichprobe $X \odot \mathbb{P}_\theta^n$ ein Maximum-Likelihood-Schätzer für $\theta$ ist.
*Hinweis:* Sei $X_1, \dots, X_n$ eine Stichprobe und $X_{(1)}, \dots, X_{(n)}$ die korrespondierende geordnete Stichprobe, sodass $X_{(i)} \le X_{(i+1)}$. Dann ist der empirische Median definiert als eine Wahl

$$X_{\text{med}} \in \begin{cases} \{X_{(\frac{n+1}{2})}\}, & n \text{ ungerade;} \\ [X_{(n/2)}, X_{(n/2+1)}], & n \text{ gerade.} \end{cases}$$

---

**Klausurvorbereitung:**
Die folgende Aufgabe dient zur Wiederholung des Vorlesungsstoffes und als Hilfsmittel zum gezielten Lernen auf die Klausur.

**Aufgabe 5 (Vorbereitung; 4 Bonuspunkte)**

Überprüfen Sie, ob die folgende Verteilungsfamilie eine Exponentialfamilie bildet und geben Sie gegebenfalls den natürlichen Parameterraum an.

a) Geometrische Verteilung $\{\text{Geom}_\theta : \theta \in (0,1)\}$ mit Dichte

$$\mathbb{p}_\theta(k) = \theta \cdot (1-\theta)^k, \quad k \in \mathbb{N}_0,$$

bezüglich des Zählmaßes auf $\mathbb{N}_0$.

Seien nun $X = (X_1, \dots, X_n)^t \odot \{\text{U}_{[0,\theta]}^n | \theta \in (0, \infty)\}$.

b) Zeigen Sie, dass $S(x_1, \dots, x_n) := \frac{x_1}{\max(x_1, \dots, x_n)}$ unwesentlich ist.

c) Zeigen Sie, dass $S(X_1, \dots, X_n)$ und $\max(X_1, \dots, X_n)$ stochastisch unabhängig sind.

d) Bestimmen Sie den Erwartungswert $\mathbb{E}[S(X_1, \dots, X_n)]$.

**Abgabe:**
In Zweiergruppen, bis spätestens Montag, den **01. Juli 2019, 11:15 Uhr**.
(Die Übungszettelkästen sind im 1. OG, INF 205, vor dem Dekanat.)

**Homepage:**
https://sip.math.uni-heidelberg.de/vl/st1-ss19/