Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 1. Übungsblatt

1. Beweisen Sie für Entscheidungsregeln $\rho$ basierend auf einem statistischen Experiment $(\mathcal{X}, \mathcal{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ mit Verlustfunktion $l$:

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

   (b) Ist $\rho$ zulässig mit konstanter Risikofunktion, so ist $\rho$ minimax.

   (c) Ist $\rho$ eine Bayesregel (bzgl. $\pi$) und eindeutig in dem Sinn, dass jede andere Bayesregel (bzgl. $\pi$) die gleiche Risikofunktion besitzt, so ist $\rho$ zulässig.

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

2. Es sei $X_1, \dots, X_n$ eine mathematische $N(\vartheta, 1)$-verteilte Stichprobe mit $\vartheta \in \mathbb{R}$ unbekannt. Mit $R(\vartheta, \hat\vartheta) = \mathbb{E}_\vartheta[(\hat\vartheta - \vartheta)^2]$ werde das quadratische Risiko bezeichnet.

   (a) Beweisen Sie mit dominierter Konvergenz, dass die Abbildung $\vartheta \mapsto R(\vartheta, \hat\vartheta)$ für jeden Schätzer $\hat\vartheta$ mit $R(\vartheta, \hat\vartheta) < \infty$ für alle $\vartheta \in \mathbb{R}$ stetig ist.

   (b) Folgern Sie mit Aufgabe 1(d), dass der Bayesschätzer

   $$\hat\vartheta_{a,\sigma} = \frac{n\sigma^2}{1+n\sigma^2}\bar X + \frac{1}{1+n\sigma^2}a$$

   bezüglich der a priori-Verteilung $\pi = N(a, \sigma^2)$, $a \in \mathbb{R}$, $\sigma > 0$, zulässig ist.

3. *Truncated normal means.* Es sei $X_1, \dots, X_n$ eine mathematische $N(\vartheta, 1)$-verteilte Stichprobe sowie $\Theta_1 = [\vartheta_0, \infty)$ und $\Theta_2 = [-m, m]$ Parametermengen für $\vartheta_0 \in \mathbb{R}$, $m > 0$. $\bar X = \frac{1}{n}\sum_{i=1}^n X_i$ bezeichne das Stichprobenmittel. Lesen Sie Example 5.2.9 auf Seite 327 in Lehmann/Casella `https://link.springer.com/book/10.1007/b98854` und beweisen Sie vollständig:

   (a) In beiden Fällen $\vartheta \in \Theta_1$ und $\vartheta \in \Theta_2$ ist $\bar X$ nicht zulässig.

   (b) Im Fall $\vartheta \in \Theta_1$ ist $\bar X$ minimax, im Fall $\vartheta \in \Theta_2$ ist $\bar X$ nicht minimax.

4. Beweisen Sie das *Lemma von Stein* (1972):
   Eine reellwertige Zufallsvariable $X$ ist $N(\mu, \sigma^2)$-verteilt genau dann, wenn

   $$\mathbb{E}[(X-\mu)f(X)] = \sigma^2\,\mathbb{E}[f'(X)]$$

   für alle absolut-stetigen (schwach differenzierbaren) Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit $\mathbb{E}[|f'(X)|] < \infty$ gilt.

   *Anleitung für eine Richtung*: Für die $N(0,1)$-Verteilungsfunktion $\Phi$ und $z \in \mathbb{R}$ betrachte

   $$f_z(x) = \begin{cases} \sqrt{2\pi}e^{x^2/2}\Phi(x)(1-\Phi(z)), & \text{für } x \leqslant z, \\ \sqrt{2\pi}e^{x^2/2}\Phi(z)(1-\Phi(x)), & \text{für } x > z. \end{cases}$$

   Dann folgt aus der Identität für $f_z$ und $\mu = 0$, $\sigma = 1$, dass $X$ $N(0,1)$-verteilt ist. Betrachte im Anschluss die Identität für $Y = (X-\mu)/\sigma$.

   *Hinweis:* Dies ist Ausgangspunkt der mächtigen *Steinschen Methode*.

---

Abgabe vor der Vorlesung am **Donnerstag, dem 5.5.22**.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 2. Übungsblatt

1. Gegeben sei $X \sim N(\mu, \sigma^2 E_d)$ mit $\sigma > 0$ bekannt und $\mu \in \mathbb{R}^d$ unbekannt.

   (a) Zeigen Sie: Soll in einem statistischen Experiment $g(\vartheta) \in \mathbb{R}^d$ durch $\hat{g}$ geschätzt werden, so gilt die *Bias-Varianz-Zerlegung*:

   $$\mathbb{E}_\vartheta[|\hat{g} - g(\vartheta)|^2] = |\,\mathbb{E}_\vartheta[\hat{g}] - g(\vartheta)|^2 + \mathbb{E}_\vartheta[|\hat{g} - \mathbb{E}_\vartheta[\hat{g}]|^2]$$

   (b) Bestimmen Sie die Bias-Varianz-Zerlegung für $\hat{\mu}_\alpha = \alpha X$, $\alpha \in \mathbb{R}$, und zeigen Sie, dass $\alpha_{\text{Orakel}} := 1 - \frac{\sigma^2 d}{|\mu|^2 + \sigma^2 d}$ das quadratische Risiko minimiert, falls $\mu$ der wahre Parameter ist. Erklären Sie das Verhalten für $d \to \infty$ bei konstantem $|\mu|$.

   (c) Weisen Sie nach, dass $|X|^2$ ein erwartungstreuer Schätzer von $|\mu|^2 + \sigma^2 d$ ist, und setzen Sie $\hat{\alpha} := 1 - \frac{\sigma^2 d}{|X|^2}$. Schließen Sie durch Berechnen von $\operatorname{Var}(|X|^2)$, dass $\forall \varepsilon, R > 0\ \exists K > 0:$

   $$\mathbb{P}_\mu\left(\left|\frac{|X|^2}{\sigma^2 d} - \frac{|\mu|^2 + \sigma^2 d}{\sigma^2 d}\right| \geqslant \frac{K}{\sqrt{d}}\right) \leqslant \varepsilon, \quad \forall d \geqslant 1\ \forall \mu \in \mathbb{R}^d \text{ mit } |\mu| \leqslant R.$$

   Folgern Sie, dass $|\hat{\alpha} - \alpha_{\text{Orakel}}| = \mathcal{O}_P(d^{-1/2})$ für $d \to \infty$ und $|\mu| \leqslant R$ gilt, das heißt $\forall \varepsilon, R > 0\ \exists K' > 0:$

   $$\mathbb{P}_\mu(|\hat{\alpha} - \alpha_{\text{Orakel}}| > K'\, d^{-1/2}) \leqslant \varepsilon, \quad \forall d \geqslant 1\ \forall \mu \in \mathbb{R}^d \text{ mit } |\mu| \leqslant R.$$

2. Es sei $X_1, \ldots, X_n$ eine $N(\mu, E_d)$-verteilte mathematische Stichprobe. Der James-Stein-Schätzer mit positivem Gewicht ist definiert als $\hat{\mu}_{JS+} = \big(1 - \frac{d-2}{n|\overline{X}|^2}\big)_+ \overline{X}$. Beweisen Sie für alle $d \geqslant 3$ und $\mu \in \mathbb{R}^d$ schrittweise folgenden Risikovergleich mit dem klassischen James-Stein-Schätzer:

   $$\mathbb{E}_\mu[|\hat{\mu}_{JS+} - \mu|^2] < \mathbb{E}_\mu[|\hat{\mu}_{JS} - \mu|^2].$$

   (a) Die Abschätzung ist korrekt für $\mu = 0$.

   (b) Die Abschätzung folgt aus der Ungleichung $\mathbb{E}_\mu[\mu_i \overline{X}_i |G| \mathbf{1}_{\{G \leqslant 0\}}] > 0$ für $G = 1 - \frac{d-2}{n|\overline{X}|^2}$ und alle $i = 1, \ldots, d$ mit $\mu_i \neq 0$.

   (c) Für $a > 0$ und $\mu_i \neq 0$ gilt $\mathbb{E}_\mu[\mu_i \overline{X}_i \,|\, (\overline{X}_i)^2 = a^2] = a\mu_i \tanh(na\mu_i) > 0$. Dies ergibt die Ungleichung in (b) durch Einfügen einer auf $((\overline{X}_1)^2, \ldots, (\overline{X}_d)^2)$ bedingten Erwartung.

3. Es sei $X_1, \dots, X_n$ eine $N(\mu, 1)$-verteilte mathematische Stichprobe mit $\mu \in \mathbb{R}$ unbekannt.

   (a) Geben Sie das zugehörige statistische Experiment auf $\mathcal{X} = \mathbb{R}^n$ an und zeigen Sie, dass es vom Produktmaß $N(0,1)^{\otimes n}$ dominiert wird.

   (b) Bestimmen Sie die Likelihoodfunktion für das dominierende Maß in (a). Welcher Wert $\mu \in \mathbb{R}$ maximiert die Likelihoodfunktion zu gegebenem $x \in \mathbb{R}^n$ (dies ist der Maximum-Likelihood-Schätzer bei Beobachtung $X = x$)?

4. Beweisen oder widerlegen Sie die Aussage, dass folgende Verteilungen Exponentialfamilien bilden. Bestimmen Sie gegebenenfalls den natürlichen Parameterraum.

   (a) Multinomialverteilung $(M(p_0, \dots, p_s; n))_{0<p_i<1, \sum p_i = 1}$;

   (b) Poissonverteilung $(\mathrm{Poiss}(\lambda))_{\lambda>0}$;

   (c) Gleichmäßige Verteilung $(U([0,\vartheta]))_{\vartheta>0}$;

   (d) Gammaverteilung $(\Gamma(a,b))_{a,b>0}$.

---

Abgabe vor der Vorlesung am **Donnerstag, dem 12.5.22**.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 3. Übungsblatt

1. Ein Physiker untersucht die Radioaktivität bei zwei verschiedenen Präparaten. Die unabhängig gemessene Zahl der Zerfälle in einer Zeiteinheit bei Präparat 1 sei $X_1, \dots, X_{m_1}$ ($m_1$ Messungen), bei Präparat 2 $Y_1, \dots, Y_{m_2}$ ($m_2$ Messungen). Geben Sie eine vernünftige Regel an, um zu entscheiden, welches Präparat stärker radioaktiv ist. Begründen Sie dazu, weshalb die Annahme einer Poissonverteilung gerechtfertigt ist, und geben Sie ein Suffizienzargument.

2. Es sei $X_1, \dots, X_n$ eine gleichmäßig auf den Rechtecken $R_{x,r} = [x_1, x_1 + r_1] \times [x_2, x_2 + r_2] \subseteq \mathbb{R}^2$ verteilte mathematische Stichprobe mit $x \in \mathbb{R}^2$ und $r \in (\mathbb{R}^+)^2$ unbekannt. Bestimmen Sie eine suffiziente Statistik $T(X_1, \dots, X_n) \in \mathbb{R}^4$ und beschreiben Sie diese auch geometrisch.
*Zusatzaufgabe:* Können Sie den Flächeninhalt von $R_{x,r}$ erwartungstreu schätzen?

3. Es sei $(X_1, \dots, X_n)$ eine Poiss($\lambda$)-verteilte mathematische Stichprobe mit $\lambda > 0$ unbekannt.
   (a) Zeigen Sie, dass das Stichprobenmittel $\bar{X}$ eine suffiziente Statistik ist.
   (b) Finden Sie mittels des Satzes von Rao-Blackwell einen erwartungstreuen Schätzer von $\lambda$, der kleinere Varianz als $\hat{\lambda}_1 = X_1$ besitzt.
   (c) Finden Sie mittels des Satzes von Rao-Blackwell einen erwartungstreuen Schätzer von $\vartheta = e^{-\lambda}$, der kleinere Varianz als $\hat{\vartheta}_1 = \mathbf{1}(X_1 = 0)$ besitzt.

4. Eine suffiziente Statistik $T^*$ heißt *minimalsuffizient*, wenn es zu jeder suffizienten Statistik $T$ eine messbare Funktion $h$ gibt, so dass $T^* = h(T)$ $\mathbb{P}_\vartheta$-f.s. für alle $\vartheta \in \Theta$ gilt. Beweisen Sie, dass jede $\mathbb{R}^d$-wertige, suffiziente und vollständige Statistik minimalsuffizient ist, sofern eine minimalsuffiziente Statistik überhaupt existiert. Gilt die Umkehrung für $\mathbb{R}^d$-wertige Statistiken?
*Hinweis:* Man kann zeigen, dass minimalsuffiziente Statistiken für dominierte Experimente auf separablen Messräumen (wie $(\mathbb{R}^d, \mathscr{B}_{\mathbb{R}^d})$) stets existieren.

---

Abgabe vor der Vorlesung am Donnerstag, dem 19.5.22.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 4. Übungsblatt

1. Erklären Sie im Detail den Beweis der Cramér-Rao-Ungleichung unter „klassischen Bedingungen“ in Theorem 3.3 von J. Shao, Mathematical Statistics. Wieso muss dort eine Bedingung (3.3) an den Schätzer gestellt werden und bei uns nicht?

2. Zeigen Sie im Detail, dass die *score*-Funktion und die Fisher-Information bei Hellinger-differenzierbaren Modellen unabhängig vom dominierenden Maß $\mu$ sind.

3. Betrachten Sie für eine Lebesguedichte $f : \mathbb{R} \to [0, \infty)$ das *Lokationsmodell* einer mathematischen Stichprobe $X_1, \ldots, X_n$, die gemäß $f(\bullet - \vartheta)$ mit $\vartheta \in \mathbb{R}$ unbekannt verteilt ist. Zeigen Sie, dass diese Verteilungsfamilie $L^2$-differenzierbar ist, wenn $f \in C^1(\mathbb{R})$, $f > 0$ und $\int f'(x)^2 f(x)^{-1} dx < \infty$ gilt, und bestimmen Sie die Fisher-Information $I_n(\vartheta)$.
   Können Sie die Bedingungen so abschwächen, dass auch die $L^2$-Differenzierbarkeit im Fall der Laplace-Verteilung $f(x) = \frac{1}{2} e^{-|x|}$ folgt?

4. Bestimmen Sie die Fisher-Informationsmatrix für eine $N(\mu, \sigma^2)$-verteilte mathematische Stichprobe $X_1, \ldots X_N$ mit unbekannten Werten $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ sowie für $X \sim \mathrm{Bin}(n, p)$ mit $p \in (0, 1)$ unbekannt und $n$ bekannt. Finden Sie jeweils einen erwartungstreuen Schätzer für $\mu$ und für $p$, der die Cramér-Rao-Schranke erreicht. Bestimmen Sie einen erwartungstreuen Schätzer für $\sigma^2$ bei $N \geqslant 2$ Beobachtungen, der zumindest asymptotisch für $N \to \infty$ die Cramér-Rao-Schranke erreicht (bei Reskalierung in $N$).

---

Abgabe vor der Vorlesung am Montag, dem 30.5.22.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## Korrektur und Lösung zu Aufgabe 4.3

Betrachten Sie für eine Lebesguedichte $f : \mathbb{R} \to [0, \infty)$ das *Lokationsmodell* einer mathematischen Stichprobe $X_1, \ldots, X_n$, die gemäß $f(\bullet - \vartheta)$ mit $\vartheta \in \mathbb{R}$ unbekannt verteilt ist. Zeigen Sie, dass diese Verteilungsfamilie $L^2(\mathbb{P}_{\vartheta_0})$-differenzierbar ist, wenn $f \in C^1(\mathbb{R})$, $f > 0$ und $\int \sup_{|h|\leqslant\varepsilon} \frac{f'(x+h)^2+f(x+h)^2}{f(x)} dx < \infty$ für ein $\varepsilon > 0$ gilt.

Die Likelihoodfunktion bezüglich $\mathbb{P}_{\vartheta_0}$ ist

$$L_{\vartheta_0}(\vartheta, x) = \prod_{i=1}^{n} \frac{f(x_i - \vartheta)}{f(x_i - \vartheta_0)}, \quad x \in \mathbb{R}^n, \, \vartheta \in \mathbb{R}\,.$$

Es gilt (klassische Differenzierbarkeit in $\vartheta$ für jedes $x \in \mathbb{R}^n$)

$$\dot{L}_{\vartheta_0}(\vartheta, x) = \sum_{i=1}^{n} \frac{-f'(x_i - \vartheta)}{f(x_i - \vartheta)} L_{\vartheta_0}(\vartheta, x),$$

und für $\vartheta > \vartheta_0$

$$\frac{L_{\vartheta_0}(\vartheta) - L_{\vartheta_0}(\vartheta_0) - \dot{L}_{\vartheta_0}(\vartheta_0)(\vartheta - \vartheta_0)}{\vartheta - \vartheta_0} = \frac{1}{\vartheta - \vartheta_0} \int_{\vartheta_0}^{\vartheta} (\dot{L}_{\vartheta_0}(\vartheta') - \dot{L}_{\vartheta_0}(\vartheta_0)) d\vartheta'$$

sowie mit vertauschten Integralgrenzen analog für $\vartheta < \vartheta_0$. Wegen stetiger Differenzierbarkeit von $f$ gilt $\frac{1}{\vartheta-\vartheta_0} \int_{\vartheta_0}^{\vartheta} (\dot{L}_{\vartheta_0}(\vartheta', x) - \dot{L}_{\vartheta_0}(\vartheta_0, x)) d\vartheta' \to 0$ für $\vartheta \to \vartheta_0$ und jedes $x \in \mathbb{R}^n$. Mit dominierter Konvergenz folgt also die $L^2(\mathbb{P}_{\vartheta_0})$-Differenzierbarkeit

$$\Big\| \frac{L_{\vartheta_0}(\vartheta) - L_{\vartheta_0}(\vartheta_0) - \dot{L}_{\vartheta_0}(\vartheta_0)(\vartheta - \vartheta_0)}{\vartheta - \vartheta_0} \Big\|_{L^2(\mathbb{P}_{\vartheta_0})} \to 0 \text{ für } \vartheta \to \vartheta_0,$$

wobei eine integrierbare Majorante für alle $|\vartheta - \vartheta_0| < \varepsilon$ gegeben ist durch

$$\sup_{|h|\leqslant\varepsilon} (\dot{L}_{\vartheta_0}(\vartheta_0 + h) - \dot{L}_{\vartheta_0}(\vartheta_0))^2 \leqslant 4n \sum_{i=1}^{n} \sup_{|h|\leqslant\varepsilon} \frac{f'(X_i - \vartheta_0 - h)^2}{f(X_i - \vartheta_0 - h)^2} L_{\vartheta_0}(\vartheta_0 + h, X)^2,$$

wobei

$$\mathbb{E}_{\mathbb{P}_{\vartheta_0}} \Big[ \sup_{|h|\leqslant\varepsilon} \frac{f'(X_i - \vartheta_0 - h)^2}{f(X_i - \vartheta_0 - h)^2} L_{\vartheta_0}(\vartheta_0 + h, X)^2 \Big]$$
$$\leqslant \int \sup_{|h|\leqslant\varepsilon} \frac{f'(x_i - \vartheta_0 - h)^2}{f(x_i - \vartheta_0)^2} f(x_i - \vartheta_0)\, dx_i \prod_{j\neq i} \int \sup_{|h|\leqslant\varepsilon} \frac{f(x_j - \vartheta_0 - h)^2}{f(x_j - \vartheta_0)^2} f(x_j - \vartheta_0)\, dx_j < \infty.$$

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 5. Übungsblatt

1. Für zwei Wahrscheinlichkeitsmaße $\mathbb{P}, \mathbb{Q}$ auf demselben Raum $(\mathcal{X}, \mathscr{F})$ bezeichnet
$$H^2(\mathbb{P}, \mathbb{Q}) := \int_{\mathcal{X}} \Big(\sqrt{p(x)} - \sqrt{q(x)}\Big)^2 \mu(dx)$$
den quadrierten *Hellinger-Abstand*, wobei $\mu$ ein Maß ist, das $\mathbb{P}$ und $\mathbb{Q}$ dominiert (z.B. $\mu = \mathbb{P} + \mathbb{Q}$), und $p, q$ die entsprechenden $\mu$-Dichten bezeichnet. Zeigen Sie:

   (a) Der Hellingerabstand $H$ ist unabhängig vom dominierenden Maß $\mu$ und definiert eine Metrik.

   (b) Es gilt $H^2(\mathbb{P}, \mathbb{Q}) = 2 - 2\int_{\mathcal{X}} \sqrt{p(x)q(x)}\,\mu(dx) \in [0, 2]$.

   (c) Für Produktmaße $\mathbb{P} = \otimes_{i=1}^n \mathbb{P}_i$, $\mathbb{Q} = \otimes_{i=1}^n \mathbb{Q}_i$ gilt
$$H^2(\mathbb{P}, \mathbb{Q}) = 2\Big(1 - \prod_{i=1}^n \Big(1 - \frac{H^2(\mathbb{P}_i, \mathbb{Q}_i)}{2}\Big)\Big) \leqslant \sum_{i=1}^n H^2(\mathbb{P}_i, \mathbb{Q}_i).$$

   (d) Es gilt $H^2(\mathbb{P}, \mathbb{Q}) \leqslant \int_{\mathcal{X}} |p(x) - q(x)|\,\mu(dx) = 2\|\mathbb{P} - \mathbb{Q}\|_{TV}$ mit dem *Totalvariationsabstand* $\|\mathbb{P} - \mathbb{Q}\|_{TV} := \sup_{A \in \mathscr{F}} |\mathbb{P}(A) - \mathbb{Q}(A)|$ (eine der *Le-Cam-Ungleichungen*).

2. Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ ein statistisches Modell mit $\Theta \subseteq \mathbb{R}^k$, $k \geqslant 2$, das die Voraussetzungen an die Cramér-Rao-Ungleichung erfüllt mit strikt positiv-definiter Fisher-Informationsmatrix $I(\vartheta)$.

   (a) Zeigen Sie, dass stets $(I(\vartheta))_{11}(I(\vartheta)^{-1})_{11} \geqslant 1$ gilt.
   *Tipp:* Schreiben Sie den ersten Einheitsvektor $e_1 = \sum_i \langle e_1, v_i\rangle v_i$ in einer Orthonormalbasis $(v_1, \ldots, v_k)$ von Eigenvektoren von $I(\vartheta)$.

   (b) Schließen Sie, dass die Cramér-Rao-Schranke für $g(\vartheta) = g(\vartheta_1, \ldots, \vartheta_n) = \vartheta_1$ stets mindestens $1/(I(\vartheta))_{11}$ beträgt. In welchen Fällen gilt Gleichheit?

   (c) Interpretieren Sie dieses Ergebnis statistisch, indem Sie mit dem Modell vergleichen, wo nur $\vartheta_1$ unbekannt ist, während $\vartheta_2, \ldots, \vartheta_k$ bekannt sind.

3. Es sei $(X,Y)^\top$ bivariat normalverteilt mit $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, $\mathrm{Var}(X) = \mathrm{Var}(Y) = \sigma^2$ und $\mathrm{Cov}(X,Y) = \sigma^2\rho$, wobei $\vartheta = (\sigma^2, \rho) \in \mathbb{R}^+ \times (-1,1)$ der unbekannte Parameter sei.

   (a) Weisen Sie nach, dass für $\hat{\sigma}^2 = \frac{1}{2}(X^2+Y^2)$ gilt $\mathbb{E}[\hat{\sigma}^2] = \sigma^2$ und $\mathrm{Var}(\hat{\sigma}^2) = (1+\rho^2)\sigma^4$.

   (b) Nehmen Sie nun an, dass $\rho$ bekannt ist, und konstruieren Sie einen erwartungstreuen Schätzer $\tilde{\sigma}^2_\rho$ mit $\mathrm{Var}(\tilde{\sigma}^2_\rho) = \sigma^4 < \mathrm{Var}(\hat{\sigma}^2)$ im Fall $\rho \neq 0$.
   *Tipps:* Diagonalisieren Sie die Kovarianzmatrix $\Sigma$ von $(X,Y)^\top$ und betrachten Sie die Beobachtung in der Eigenbasis von $\Sigma$ ('whitening', 'decorrelation'). Alternativ bestimmen Sie direkt einen Maximum-Likelihood-Schätzer.

   (c) Weisen Sie nach, dass die Varianz von $\tilde{\sigma}^2_\rho$ die Cramér-Rao-Schranke bei bekanntem $\rho$ erreicht.

   (d*) Bestimmen Sie die Cramér-Rao-Schranke für $\sigma^2$ im Fall von unbekanntem $\rho$ (Einsatz von Mathematik-Software bei Angabe von Zwischenergebnissen gestattet).

4. Zeigen Sie, dass im Lokationsmodell mit Produktdichte $f(x) = \prod_{i=1}^n f_1(x_i)$ der Pitman-Schätzer $\hat{\vartheta}$ für die angegebenen Dichten folgende Form besitzt:

   (a) $f_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$, $\hat{\vartheta} = \bar{X}$;

   (b) $f_1(x) = a^{-1}\mathbf{1}_{[-\frac{a}{2},\frac{a}{2}]}(x)$ für $a > 0$, $\hat{\vartheta} = \frac{X_{(1)}+X_{(n)}}{2}$;

   (c) $f_1(x) = \lambda e^{-\lambda(x+1)}\mathbf{1}_{[-1,\infty)}(x)$ für $\lambda > 0$, $\hat{\vartheta} = X_{(1)} + 1 - \frac{1}{n\lambda}$.

---

Abgabe vor der Vorlesung am Donnerstag, dem 9.6.22.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 6. Übungsblatt

1. Zeigen Sie für die Kullback-Leibler-Divergenz:

   (a) Sie ist additiv für Produktmaße:

   $$\mathrm{KL}(\mathbb{P}_1 \otimes \mathbb{P}_2 \mid \mathbb{Q}_1 \otimes \mathbb{Q}_2) = \mathrm{KL}(\mathbb{P}_1 \mid \mathbb{Q}_1) + \mathrm{KL}(\mathbb{P}_2 \mid \mathbb{Q}_2).$$

   (b) Sie ist größer als der quadrierte Hellingerabstand:

   $$H^2(\mathbb{P}, \mathbb{Q}) \leqslant \mathrm{KL}(\mathbb{P} \mid \mathbb{Q}).$$

   *Tipp:* Verwende $\log(x) \geqslant -2(x^{-1/2} - 1)$, $x > 0$ (Beweis?).

2. Ein Teich enthält eine unbekannte Anzahl $\vartheta$ von Karpfen. Zur Schätzung von $\vartheta$ werden zunächst $w$ Fische gefangen, markiert und wieder freigelassen. Wenn sich die markierten Fische wieder gut verteilt haben, werden $n$ Fische gefangen, von denen $x$ markiert sind.
   Modellieren Sie die Schätzung des Fischbestandes $\vartheta$ durch ein statistisches Experiment mit hypergeometrischen Verteilungen und bestimmen Sie einen MLE, diskutieren Sie dabei den Fall $x = 0$ separat.

3. Sei $X_1, \dots, X_n$ eine mathematische Stichprobe bezüglich der Lebesguedichte

   $$f_\vartheta(x) = \frac{1-\vartheta}{\varphi(\vartheta)} \left(1 - \frac{|x-\vartheta|}{\varphi(\vartheta)}\right)_+ + \frac{\vartheta}{2} \mathbf{1}_{[-1,1]}(x), \quad x \in \mathbb{R},$$

   wobei $\vartheta \in [0,1)$ und $\varphi : [0,1] \to [0,1]$ eine stetige, fallende Funktion mit $\varphi(0) = 1$ und $0 < \varphi(\vartheta) \leqslant 1 - \vartheta$ für $\vartheta \in (0,1)$ ist. Setze weiterhin $f_1(x) = \frac{1}{2}\mathbf{1}_{[-1,1]}(x)$. Ziel ist es, für geeignetes $\varphi$ zu sehen, dass für alle $\vartheta \in \Theta = [0,1]$ jeder MLE fast sicher gegen Eins konvergiert und insbesondere inkonsistent ist. Zeigen Sie:

   (a) Es existiert ein Maximum-Likelihood-Schätzer $\hat{\vartheta}_n$.

   (b) Für $\vartheta < 1$ ist $f_\vartheta(x) < 1/\varphi(\vartheta) + 1/2$ und daraus folgt, dass für die Loglikelihoodfunktion $\ell_n$ bei $n$ Beobachtungen und für jedes $\alpha < 1$

   $$\max_{0 \leqslant \vartheta \leqslant \alpha} \frac{\ell_n(\vartheta)}{n} \leqslant \log\left(\frac{1}{\varphi(\alpha)} + \frac{1}{2}\right) < \infty$$

   gilt. Um zu beweisen, dass $\lim_{n\to\infty} \hat{\vartheta}_n = 1$ f.s. für alle $\vartheta \in [0,1]$, reicht es $\max_{0\leqslant\vartheta\leqslant 1} \ell_n(\vartheta)/n \to \infty$ f.s. zu zeigen.

(c) Mit $X_{(n)} = \max\{X_1, \dots, X_n\}$ gilt f.s.

$$\max_{0\leqslant\vartheta\leqslant 1} \frac{\ell_n(\vartheta)}{n} \geqslant \frac{n-1}{n} \log\left(\frac{X_{(n)}}{2}\right) + \frac{1}{n} \log\left(\frac{1-X_{(n)}}{\varphi(X_{(n)})}\right).$$

(d) Aus dem Lemma von Borel-Cantelli folgt $n^{1/4}(1 - X_{(n)}) \to 0$ f.s. für $\vartheta = 0$ und auch für alle $\vartheta \in [0,1]$. Mit $\varphi(\vartheta) := (1-\vartheta)\exp(-(1-\vartheta)^{-4}+1)$ folgt $\liminf_{n\to\infty}(1/n)\log((1-X_{(n)})/\varphi(X_{(n)})) = \infty$ f.s. und damit die gewünschte Aussage.

4. Es sei $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$ ein statistisches Modell mit Likelihoodfunktion $L(\vartheta)$ und MLE $\hat{\vartheta}$ sowie $g : \Theta \to \Theta'$ eine injektive Funktion.

   (a) Zeigen Sie, dass $\hat{\vartheta}' := g(\hat{\vartheta})$ MLE ist für das reparametrisierte Modell mit Verteilungen $(\mathbb{P}_{g^{-1}(\vartheta')})_{\vartheta'\in g(\Theta)}$ (*Plug-in-Prinzip*)

   (b) Betrachten Sie nun $(\mathbb{R}^n, \mathfrak{B}_{\mathbb{R}^n}, (N(\mu,1)^{\otimes n})_{\mu\in\mathbb{R}})$ mit MLE $\hat{\mu}_n = \bar{X}$. Zeigen Sie für $g(\mu) = \mu^{2k+1}$, $k \in \mathbb{N}$, dass $\widehat{\mu^{2k+1}} = (\hat{\mu}_n)^{2k+1}$ MLE, aber nicht erwartungstreu ist. Finden Sie durch Biaskorrektur von $\widehat{\mu^3}$ einen erwartungstreuen Schätzer $\widetilde{\mu^3}$ von $\mu^3$, der einen kleineren quadratischen Fehler als $\widehat{\mu^3}$ besitzt und UMVU ist. Vergleichen Sie asymptotisch für $n \to \infty$ die quadratischen Fehler von $\widehat{\mu^3}$, $\widetilde{\mu^3}$ und die entsprechende Cramér-Rao-Schranke.

---

Abgabe vor der Vorlesung am Donnerstag, dem 16.6.22.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 7. Übungsblatt

1. Beweise: die Bedingungen (A1), (A2) zur Konsistenz des M-Schätzers $\hat{\gamma}_n$ unter $\mathbb{P}^n_{\vartheta_0}$ können ersetzt werden durch die eine Bedingung ($d$ sei Metrik von $\Gamma$)
$$(A1') : \ \forall \varepsilon > 0 \ \inf_{d(\gamma, g(\vartheta_0)) \geqslant \varepsilon} K(\vartheta_0, \gamma) > K(\vartheta_0, g(\vartheta_0)).$$
Anleitung: Weise $|K(\vartheta_0, \hat{\gamma}_n) - K(\vartheta_0, g(\vartheta_0))| \leqslant 2 \sup_{\gamma \in \Gamma} |K_n(\gamma) - K(\vartheta_0, \gamma)|$ nach, betrachte $\mathbb{P}_{\vartheta_0}(d(\hat{\gamma}_n, g(\vartheta_0)) > \varepsilon)$ und verwende (A3).

2. Für eine $\mathbb{R}$-wertige mathematische Stichprobe $X_1, \dots, X_n$ mit $\mathbb{E}_\vartheta[X_i^2] < \infty$, sei $g(\vartheta) = \mathbb{E}_\vartheta[X_1]$, $\Gamma = \mathbb{R}$, $K_n(\gamma) = \frac{1}{n} \sum_{i=1}^n (X_i - \gamma)^2$, $K(\vartheta_0, \gamma) = \mathbb{E}_{\vartheta_0}[(X_1 - \gamma)^2]$.
   (a) Bestimmen Sie den zugehörigen M-Schätzer $\hat{\gamma}_n$.
   (b) Verwenden Sie Aufgabe 1 sowie eine geeignete Version von Lemma 3.33 im Skript von Martin Wahl, um die Konsistenz von $\hat{\gamma}_n$ nachzuweisen ($\Gamma$ ist nicht kompakt!).

3. (*freiwillig*) Bearbeiten Sie Aufgabe 2 diesmal mit der Kontrastfunktion $K(\vartheta_0, \gamma) = \mathbb{E}_{\vartheta_0}[|X_1 - \gamma|]$ und entsprechendem Kontrastprozess $K_n$.

4. Betrachten Sie für $\vartheta \in \Theta := C([0,1])$ das Regressionsmodell
$$Y_i = \vartheta(i/n) + \varepsilon_i, \quad i = 1, \dots, n,$$
mit $\varepsilon_i$ i.i.d., $\mathbb{E}[\varepsilon_i] = 0$, $\mathbb{E}[\varepsilon_i^4] < \infty$. Es bezeichne $\Gamma = \{\sum_{k=0}^d c_k x^k\} \subseteq \Theta$ die Menge aller Polynome vom Grad kleiner oder gleich $d$ sowie $g(\vartheta) = \Pi_\Gamma \vartheta$ die Orthogonalprojektion von $\Theta$ auf $\Gamma$ in $L^2([0,1])$.
   (a) Zeigen Sie, dass $K_n(\gamma) := \frac{1}{n} \sum_{i=1}^n (Y_i - \gamma(i/n))^2$, $\gamma \in \Gamma$, ein Kontrastprozess ist zur Kontrastfunktion
$$K(\vartheta_0, \gamma) = \|\vartheta_0 - \Pi_\Gamma \vartheta_0\|^2_{L^2([0,1])} + \|\Pi_\Gamma \vartheta_0 - \gamma\|^2_{L^2([0,1])} + \mathbb{E}[\varepsilon_i^2].$$
   (b) Schließen Sie, dass der Kleinste-Quadrate-Schätzer der Polynomregression $\hat{\gamma}_n = \operatorname{argmin}_{\gamma \in \Gamma} \sum_{i=1}^n (Y_i - \gamma(i/n))^2$ M-Schätzer ist und gegen die Orthogonalprojektion $g(\vartheta_0)$ $\mathbb{P}^n_{\vartheta_0}$-stochastisch konvergiert.
   *Hinweis:* Betrachten Sie zunächst eine kompakte Umgebung $\Gamma_0 \subseteq \Gamma$ von $g(\vartheta_0)$ und zeigen Sie dann $\mathbb{P}^n_{\vartheta_0}(\hat{\gamma}_n \notin \Gamma_0) \to 0$. Alternativ kann wie in Aufgaben 1-2 argumentiert werden.
   (c*) Weisen Sie $\sqrt{n}(\hat{\gamma}_n - g(\vartheta_0)) \xrightarrow{d} N(0, V^{-1} U V^{-1})$ unter $\mathbb{P}^n_{\vartheta_0}$ nach mit $U = 2\,\mathbb{E}[\varepsilon_i^2] V$, $V = 2(\int_0^1 x^k x^l dx)_{0 \leqslant k,l \leqslant d}$ (in der Basis der Monome).

---

Abgabe vor der Vorlesung am Donnerstag, dem 23.6.22.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 8. Übungsblatt

1. Es sei $X_1, \dots, X_n$ eine $N(\vartheta, 1)$-verteilte Stichprobe mit $\vartheta \in [0,1]$ unbekannt.
   (a) Bestimmen Sie den MLE $\hat{\vartheta}_n$.
   (b) Weisen Sie für $\vartheta_0 \in (0,1)$ die Bedingungen des Satzes zur MLE-Asymptotik nach und folgern Sie $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \xrightarrow{d} N(0,1)$ für $n \to \infty$ unter $N(\vartheta_0, 1)^{\otimes n}$.
   (c) Untersuchen Sie die Asymptotik von $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)$ im Fall $\vartheta_0 \in \{0,1\}$.

2. *Supereffizienz.* Für eine $N(\mu, 1)$-verteilte Stichprobe $X_1, \dots, X_n$ sei
$$\hat{\mu}_n := \begin{cases} \bar{X}, & \text{falls } |\bar{X}| > n^{-1/4} \\ a\bar{X}, & \text{falls } |\bar{X}| \leqslant n^{-1/4} \end{cases}$$
der sogenannte *Hodges-Schätzer* mit $|a| < 1$. Zeigen Sie
$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{d} N(0, v(\mu))$$
mit $v(\mu) = 1$ für $\mu \neq 0$ und $v(0) = a^2 < 1$, während die Fisher-Information bei einer Beobachtung $I(\mu) = 1$ für alle $\mu \in \mathbb{R}$ beträgt.

3. Beweisen Sie: Jeder beste Test im binären statistischen Modell für $H_0 : \vartheta = 0$ gegen $H_1 : \vartheta = 1$ besitzt $\mathbb{P}_0$-fast sicher und $\mathbb{P}_1$-fast sicher die Form eines Neyman-Pearson-Tests.
*Tipp:* Untersuchen Sie, wann Gleichheit im Beweis des Neyman-Person-Lemmas gilt.

4. Beweisen Sie das verallgemeinerte Neyman-Pearson-Lemma:
Es seien $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta \in \Theta})$ mit $\Theta = \{0,1\}$ ein binäres statistisches Modell, $p_0, p_1$ die entsprechenden Dichten und $T \in L^1(\mathbb{P}_0)$ eine reellwertige Statistik. Ein Test der Form
$$\varphi(x) = \begin{cases} 1, & \text{falls } p_1(x) > kp_0(x) + lT(x)p_0(x) \\ 0, & \text{falls } p_1(x) < kp_0(x) + lT(x)p_0(x) \\ \gamma, & \text{falls } p_1(x) = kp_0(x) + lT(x)p_0(x) \end{cases}$$
mit $k, l \in \mathbb{R}$ und $\gamma \in [0,1]$, der für $\alpha \in [0,1]$ die Nebenbedingungen
$$\mathbb{E}_0[\varphi] = \alpha \quad \text{und} \quad \mathbb{E}_0[T\varphi] = \alpha\, \mathbb{E}_0[T]$$
erfüllt, maximiert die Güte $\mathbb{E}_1[\varphi]$ in der Menge aller Tests, die diese Nebenbedingungen erfüllen.

---

Abgabe vor der Vorlesung am Donnerstag, dem 30.6.22.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 9. Übungsblatt

1. Betrachten Sie das Testproblem $H_0 : p = 1/2$ gegen $H_1 : p \in (0,1) \setminus \{1/2\}$ bei Beobachtung von $X \sim \mathrm{Bin}(n,p)$.
   (a) Geben Sie für $n = 10$ zum Niveau $\alpha = 0.05$ einen gleichmäßig besten unverfälschten Test an.
   (b) Beweisen Sie allgemein, dass es keinen gleichmäßig besten Test unter allen Tests vom Niveau $\alpha$ gibt, indem Sie die Gütefunktion der besten einseitigen Tests studieren.

2. Basierend auf einer mathematischen Stichprobe $X_1 \dots, X_n \sim N(0, \sigma^2)$ mit $\sigma > 0$ unbekannt, soll $H_0 : \sigma = \sigma_0$ gegen $H_1 : \sigma \neq \sigma_0$ für gegebenes $\sigma_0 > 0$ getestet werden. Bestimmen Sie die Bedingungen für einen gleichmäßig besten unverfälschten Test zum Niveau $\alpha \in (0,1)$.
   *Freiwillig:* Bestimmen Sie diesen Test numerisch im Fall $\alpha = 0.05$ und $\sigma_0 = 1$ für verschiedene Werte von $n \in \mathbb{N}$. Sind die Ablehnbereiche des Tests symmetrisch um $\sigma_0$?

3. Es seien $X_1, \dots, X_{n_1} \sim N(\mu_1, \sigma_1^2)$ und $Y_1, \dots, Y_{n_2} \sim N(\mu_2, \sigma_2^2)$ zwei unabhängige mathematische Stichproben mit $\mu_1, \mu_2 \in \mathbb{R}$ und $\sigma_1, \sigma_2 > 0$ unbekannt. Zeigen Sie, dass der Likelihood-Quotienten-Test für $H_0 : \sigma_1 = \sigma_2$ gegen $H_1 : \sigma_1 \neq \sigma_2$ bei beliebigen $\mu_1, \mu_2$ sich durch die Statistik $s_1^2/s_2^2$ ausdrücken lässt mit
$$s_1^2 := \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2, \quad s_2^2 := \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2.$$
   Begründen Sie, weshalb die Verteilung von $s_1^2/s_2^2$ unter $H_0$ nicht von den konkreten Werten von $\mu_1, \mu_2, \sigma_1$ abhängt. Wie könnte man daher einen Test zum Niveau $\alpha$ durch einfache Simulation gewinnen?

4. Ein Zufallszahlengenerator erzeugt bei 1 000 Aufrufen die Ziffern $0, 1, \dots, 9$ mit folgenden Häufigkeiten:

| Ziffer | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Anzahl | 103 | 91 | 105 | 93 | 89 | 110 | 108 | 105 | 90 | 106 |

   Testen Sie mittels $\chi^2$-Test die Nullhypothese, dass eine Multinomialverteilung mit Zifferwahrscheinlichkeiten $1/10$ vorliegt, zum asymptotischen Niveau $\alpha = 0.05$.

---

Abgabe vor der Vorlesung am Donnerstag, dem 7.7.22.

Vorlesung *Mathematische Statistik*
Sommersemester 2022
Humboldt-Universität zu Berlin

Prof. Dr. Markus Reiß

## 10. Übungsblatt

1. Zeigen Sie, dass für $\mathbb{Q}_n = N(\mu_n, 1)$ und $\mathbb{P}_n = N(0,1)$ Kontiguität $\mathbb{Q}_n \triangleleft \mathbb{P}_n$ vorliegt, falls $(\mu_n) \subseteq \mathbb{R}$ eine beschränkte Folge ist. Gibt es auch eine unbeschränkte Folge $(\mu_n)$, so dass $\mathbb{Q}_n \triangleleft \mathbb{P}_n$ gilt?

2. Es sei $v : (-1,1) \to H$ eine Hilbertraum-wertige Abbildung mit $\|v(t)\| = 1$ für alle $t \in (-1,1)$, die bei $t = 0$ differenzierbar sei mit Ableitung $\dot{v}(0)$. Zeigen Sie:

   (a) Es gilt $0 = 2\langle v(0), v(t) - v(0)\rangle + \|v(t) - v(0)\|^2$ für alle $t \in (-1,1)$, und dies impliziert $\langle \dot{v}(0), v(0)\rangle = 0$ sowie

   $$\langle v(t), v(0)\rangle = 1 - \tfrac{1}{2}\|\dot{v}(0)\|^2 t^2 + o(t^2) \text{ für } t \to 0.$$

   Insbesondere ist $f(t) := \langle v(t), v(0)\rangle$ zweimal differenzierbar bei $t = 0$ mit $f''(0) = -\|\dot{v}(0)\|^2$.

   (b) Ist $L(\vartheta) = \frac{d\mathbb{P}_\vartheta}{d\mu}$ die Likelihoodfunktion in einem bei $\vartheta_0 \in \operatorname{int}(\Theta) \subseteq \mathbb{R}$ Hellinger-differenzierbaren Modell $(\mathcal{X}, \mathscr{F}, (\mathbb{P}_\vartheta)_{\vartheta\in\Theta})$, so folgt mit $v(t) = \sqrt{L(\vartheta_0 + t)} \in L^2(\mu)$ und $\|\dot{v}(0)\|^2_{L^2(\mu)} = \frac{1}{4}I(\vartheta_0)$ für den quadrierten Hellingerabstand:

   $$\tfrac{d}{d\vartheta} H^2(\mathbb{P}_{\vartheta_0}, \mathbb{P}_\vartheta)|_{\vartheta=\vartheta_0} = 0, \quad \tfrac{d^2}{d\vartheta^2} H^2(\mathbb{P}_{\vartheta_0}, \mathbb{P}_\vartheta)|_{\vartheta=\vartheta_0} = \tfrac{1}{2} I(\vartheta_0).$$

   Die Fisher-Information gibt also die Krümmung im quadrierten Hellingerabstand an.

3. Erklären Sie folgende Begriffe:

   Statistisches Modell, Verlustfunktion, Risiko, zulässig, minimax, Bayes-optimal, erwartungstreu, Likelihoodfunktion, suffizient, vollständig, Fisher-Information, Hellinger-differenzierbar, Momentenschätzer, Delta-Methode, MLE, M-Schätzer, beobachtete Fisher-Information, Neyman-Pearson-Test, unverfälschter Test, Likelihood-Quotienten-Test, LAN, Kontiguität

4. Geben Sie jeweils Beispiele und ggf. Gegenbeispiele zu den Objekten in 3. Was können wir über diese aussagen?

---

Abgabe von Aufgaben 1und 2 vor der Vorlesung am Donnerstag, dem 14.7.22.