UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386

Statistik I
Prof. Dr. Rainer Dahlhaus
Wintersemester 2020/2021

5. Übungsblatt

Aufgabe 17	Aufgabe 18	Aufgabe 19	Aufgabe 20	Summe:

Übungsgruppe: Tutor(in):

Namen:

Aufgabe 17 (Frisch-Waugh-Theorem, 4 = 1 + 1.5 + 1.5 Punkte).
Wir betrachten das lineare Modell

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

mit $\beta_i \in \mathbb{R}^{k_i}$, $X_i \in M(n \times k_i, \mathbb{R})$ und $k_i > 0$ für $i = 1, 2$, wobei $n \geq k_1 + k_2$. Ferner habe die Matrix $X := (X_1, X_2)$ vollen Rang $\text{Rang}(X) = k_1 + k_2$. Die Komponenten ε_j vom Zufallsvektor ε seien iid verteilt mit $\mathbb{E}\varepsilon_j = 0$ und $\text{Var}(\varepsilon_j) < \infty$.
In dieser Aufgabe wollen wir den Kleinste-Quadrate-Schätzer $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$ von $\beta = (\beta_1, \beta_2)'$ untersuchen. Dazu sei $P_{\mathcal{H}}$ die Orthogonalprojektion auf $\mathcal{H} := \{X_1 v : v \in \mathbb{R}^{k_1}\}$, $\tilde{X}_2 := P_{\mathcal{H}_\perp}(X_2)$ und $\tilde{Y} = P_{\mathcal{H}_\perp}(Y)$, wobei $P_{\mathcal{H}_\perp} = I_n - P_{\mathcal{H}}$. Zeigen Sie folgende Aussagen:

(a) Die Matrix $(\tilde{X}_2'\tilde{X}_2)$ ist invertierbar.

(b) Es gilt $\hat{\beta}_2 = (\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'Y$.

(c) Der Schätzer $\hat{\beta}_2$ ändert sich nicht, wenn Y durch $\tilde{Y}$ ersetzt wird, d.h. $\hat{\beta}_2 = (\tilde{X}_2'\tilde{X}_2)^{-1}\tilde{X}_2'\tilde{Y}$. Interpretieren Sie dieses Ergebnis.

Aufgabe 18 (Lineares Modell mit Störung, 3 Punkte).
Wir betrachten das lineare Modell

$$Y_i = Z_i'\beta + \varepsilon_i, \qquad i = 1, ..., n,$$

mit $\beta \in \mathbb{R}^p$, wobei die Annahmen (A1)-(A3) aus Kapitel 4 gelten. Die Variablen Z_i sind durch einen Messfehler η_i verrauscht beobachtbar. Das heißt, dass wir statt Z_i nun

$$X_i = Z_i + \eta_i.$$

messen. Die η_i seien unabhängig identisch verteilte Zufallsvariablen mit Werten in $\mathbb{R}^p$ und haben ein bekanntes positiv definites Moment $\mathbb{E}[\eta_1\eta_1'] = \Sigma_\eta$, $\mathbb{E}[Z_i\eta_i'] = 0$, $\mathbb{E}[\varepsilon_i\eta_i] = 0$.
Untersuchen Sie, ob der Kleinste-Quadrate-Schätzer $\hat{\beta} = \text{argmin}_{b \in \mathbb{R}^p} n^{-1}\sum_{i=1}^n (Y_i - X_i b)^2$ auf seine Konsistenz. Sollte $\hat{\beta}$ nicht konsistent sein, so geben Sie an, gegen welche Größe er stochastisch konvergiert.

Aufgabe 19 (Beispiele für suffiziente Statistiken, 4 = 1 + 1 + 1 + 1 Punkte).

Für eine Verteilung $\mathbb{Q}$ bezeichnet $\mathbb{Q}^n := \bigotimes_{i=1}^n \mathbb{Q}_\theta$ die gemeinsame Verteilung von n unabhängigen Zufallsvariablen, die gemäß $\mathbb{Q}$ verteilt sind. Sei $x = (x_1, ..., x_n) \in \mathbb{R}^n$. Nutzen Sie das Faktorisierungskriterium von Neyman, um zu zeigen, dass die angegebenen Statistiken suffizient für die jeweiligen Verteilungsfamilien sind:

(a) $T(x) = \overline{x}_n := \frac{1}{n}\sum_{i=1}^n x_i$ für $\{\mathrm{Bin}(1,\theta)^n : \theta \in (0,1)\}$,

(b) $T(x) = (\min_{i=1,...,n}\{x_i\}, \max_{i=1,...,n}\{x_i\})$ für $\{U[\theta_1,\theta_2]^n : \theta = (\theta_1,\theta_2) \in \mathbb{R}^2, \theta_1 < \theta_2\}$,

(c) $T(x) = (\frac{1}{n}\sum_{i=1}^n \log(x_i), \min_{i=1,...,n}\{x_i\})$ für $\{\mathrm{Pareto}(\theta_1,\theta_2)^n : \theta = (\theta_1,\theta_2) \in (0,\infty)^2\}$,
Hinweis: Die Pareto-Verteilung $\mathrm{Pareto}(\alpha, x_0)$ *hat Dichte* $f_{\alpha,x_0}(x) = \frac{\alpha x_0^\alpha}{x^{\alpha+1}}\mathbb{1}_{\{x>x_0\}}$ *bzgl. des Lebesgue-Maßes auf* $\mathbb{R}$.

(d) $T(x) = (\frac{1}{n}\sum_{i=1}^n x_i, \min_{i=1,...,n}\{x_i\})$ für $\mathcal{P} = \{\mathrm{vExp}(\theta_1,\theta_2)^n : \theta = (\theta_1,\theta_2) \in (0,\infty)\times\mathbb{R}\}$.
Hinweis: Die Dichte der verschobenen Exponentialverteilung $\mathrm{vExp}(\lambda, x_0)$ *ist gegeben über* $f_{\lambda,x_0}(x) = \lambda\ exp(-\lambda(x-x_0))\mathbb{1}_{\{x\geq x_0\}}$ *bzgl. des Lebesgue-Maßes auf* $\mathbb{R}$.

Aufgabe 20 (Minimalsuffizienz, 5 = 1.5 + 1.5 + 1 + 1 Punkte).

In dieser Aufgabe lernen wir, dass die Eigenschaft „Minimalsuffizienz“ eine Statistik auszeichnet, die eine möglichst geringe Dimension hat, aber immer noch alle Informationen aus den Beobachtungen über die Parameter der Verteilungsfamilie entält. In der Vorlesung nächsten Dienstag verkörpern „vollständige und suffizienet“ Statistiken dieselben Ideen.
Eine suffiziente Statistik $T^* : \mathcal{X} \to \mathcal{Y}$ ist minimalsuffizient für eine Verteilungsfamilie $\mathcal{P}$ auf $(\mathcal{X}, \mathcal{B})$, falls für jede weitere suffiziente Statistik $T : \mathcal{X} \to \mathcal{Y}$ eine messbare Funktion $g : \mathcal{Y} \to \mathcal{X}$ existiert, so dass $T^* = g \circ T$ $\mathbb{P}$-f.s. für alle $\mathbb{P} \in \mathcal{P}$.
Sei nun $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ eine Famlie von Verteilungen auf $(\mathcal{X}, \mathcal{B})$ mit Dichten p_θ bzgl. eines σ-endlichen Maßes μ. Ferner gebe es eine Menge $K \subseteq \mathcal{X}$ sodass $p_\theta(x) = 0 \iff x \in K$ (der Träger der Dichte p_θ ist unabhängig von θ). Zeigen Sie:

(a) Eine Statistik T ist suffizient für $\mathcal{P}$ genau dann, wenn für feste $\theta, \theta' \in \Theta$ eine messbare Abbildung $f_{\theta,\theta'} : \mathcal{Y} \to \mathbb{R}_{\geq 0}$ existiert, sodass μ-f.s. $\frac{p_\theta}{p_{\theta'}}(x) = f_{\theta,\theta'}(T(x))$.

(b) Sei $k \in \mathbb{N}$ beliebig. Ist $\Theta = \{0, ..., k\}$, so ist
$$T(x) = \left(\frac{p_1(x)}{p_0(x)}, ..., \frac{p_k(x)}{p_0(x)}\right)$$
minimalsuffizient für $\mathcal{P}$.

(c) Sei nun Θ wieder beliebig. Sei $\mathcal{P}_0 \subseteq \mathcal{P}$, T minimalsuffizient für $\mathcal{P}_0$ und suffizient für $\mathcal{P}$. Beweisen Sie, dass dann T minimalsuffizient für $\mathcal{P}$ ist.
Hinweis: Da die Dichten alle den gleichen Träger besitzen, ist $\mathbb{P}_{\theta_0}$*-f.s. für ein* $\theta_0 \in \Theta$ *äquivalent zu* $\mathbb{P}_\theta$*-f.s. für jedes* $\theta \in \Theta$.

(d) $T(x) = \overline{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$ ist minimalsuffizient für $\{\mathcal{N}(\mu, 1)^n : \mu \in \mathbb{R}\}$.

Abgabe:
In Zweiergruppen, bis spätestens Donnerstag, den **10.12.2020, 11:00 Uhr**.
Bitte geben Sie Ihre Lösungsvorschläge im PDF-Format ab. Es reicht, wenn einer der beiden Personen das Dokument hochlädt.

Homepage der Vorlesung:
`https://ssp.math.uni-heidelberg.de/stat-ws2020/index.html`