UNIVERSITÄT HEIDELBERG
ZUKUNFT
SEIT 1386

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

Vorbereitungsblatt

0.1 Lineares Modell, KQ-Schätzer, F-Tests, Konfidenzintervalle

Aufgabe 1.

Wir beobchten $Y_{1i} = a + \varepsilon_{1i}$ und $Y_{2i} = bx_i + \varepsilon_{2i}$ jeweils für $i = 1, ..., n$, wobei $\varepsilon_{1i}, \varepsilon_{2i} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$, x_i bekannte deterministische Werte und a, b Parameter seien. Wir setzen $\beta = (a, b)'$. Wir beobachten $n = 8$ Werte und erhalten folgende Kennwerte in der nachstehenden Notation: $\overline{Y_j} := \sum_{i=1}^n Y_{ji}$, $\overline{Y_j^2} := \sum_{i=1}^n Y_{ji}^2$, $\overline{xY_2} := \sum_{i=1}^n x_i Y_{2i}$, $\overline{x^2} := \sum_{i=1}^n x_i^2$.

$\overline{Y_1^2}$	$\overline{Y_1}$	$\overline{Y_1}^2$	$\overline{Y_2^2}$	$\overline{xY_2}$	$\overline{x^2}$	$\frac{\overline{xY_2}^2}{\overline{x^2}}$	$\frac{\overline{Y_1}^2}{n}$	$\frac{(\overline{Y_1}+\overline{xY_2})^2}{n+\overline{x^2}}$
3.5	4.6	21.3	104.2	148.3	214.7	102.4	2.7	105.0

(a) Geben Sie den Kleinste-Quadrate-Schätzer $\hat{\beta}$ an.

(b) Sei $x^* = 7$. Geben Sie ein 0.95-Konfidenzintervall für die Prognose von $\psi = bx^*$ an.

(c) Geben Sie einen F-Test für die Hypothesen $H_0 : a = b$ gegen $H_1 : a \neq b$ zum Niveau α an und ermitteln Sie die Testentscheidung.

Einige Quantile der F-Verteilung sind durch $F_{1,6,0.95} = 6.0$ und $F_{1,14,0.95} = 4.6$ gegeben.

Aufgabe 2.

Für eine Untersuchung nehmen Forscher an, dass das Alter Y von Bäumen linear von der Größe x abhängt, d.h. $Y = a + bx$. Die Forscher wollen testen, ob Eichen und Buchen in etwa gleich schnell wachsen. Dafür untersuchen sie $n = 8$ Eichen und $m = 12$ Buchen vorgegebener Größe und ermitteln das Alter.

(a) Modellieren Sie die Zusammenhänge durch das lineare Modell $Y_{ji} = a_j + b_j x_i + \varepsilon_{ji}$ mit $j = 1, 2$ und $i = 1, ..., n$ bzw. $i = 1, ..., m$. Geben Sie dann den KQ-Schätzer $\hat{\beta}$ für $\beta = (a_1, b_1, a_2, b_2)'$ in Termen von $\overline{x} := n^{-1} \sum_{i=1}^n x_i$, $\overline{Y_j} := n^{-1} \sum_{i=1}^n Y_{ji}$, $S_{xx} = n^{-1} \sum_{i=1}^n (x_i - \overline{x})^2$, $S_{xY_j} = n^{-1} \sum_{i=1}^n (x_i - \overline{x})(Y_{ji} - \overline{Y_j})$, $S_{Y_j Y_j} = n^{-1} \sum_{i=1}^n (Y_{ji} - \overline{Y_j})^2$ an.

(b) Geben Sie einen F-Test für die Hypothesen $H_0 : a_1 = a_2 \wedge b_1 = b_2$ gegen $H_1 : a_1 \neq a_2 \vee b_1 \neq b_2$ zum Niveau α an.

0.2 Suffizienz / Vollständigkeit, Lemma von Basu, UMVU-Schätzer

Aufgabe 3.

Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \theta^2)$ mit $\theta \in \Theta = (0, \infty)$. Es sei $\mathbb{P}_\theta := \mathbb{P}^{X_1, ..., X_n}$.

(a) Geben Sie eine vollständige und suffiziente Statistik für $\{\mathbb{P}_\theta : \theta \in \Theta\}$ an.

(b) Zeigen Sie, dass $V = \frac{X_1^2}{\sum_{i=1}^n X_i^2}$ verteilungsfrei (ancillary) bezüglich θ ist.

(c) Berechnen Sie $\mathbb{E}[V]$

Aufgabe 4.

Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \Gamma(k, b)$ mit Dichte $p_{k,b}(x) = \frac{1}{\Gamma(k)b^k} x^{k-1} e^{-\frac{x}{b}} \mathbb{1}_{(0,\infty)}(x)$, wobei $k, b > 0$. Es gilt $\mathbb{E}X_1 = kb$. Es seinen $M := n^{-1} \sum_{i=1}^n X_i$ und $U := \sqrt[n]{\prod_{i=1}^n X_i}$. Zeigen Sie:

(a) U/M ist verteilungsfrei (ancillary) bezüglich des Parameters b.

(b) M und U/M sind stochastisch unabhängig.

(c) Berechnen Sie $\mathbb{E}[U/M]$.

Aufgabe 5.

Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{vExp}(\lambda, \theta)$ mit $\theta \in \mathbb{R}$ und $\lambda > 0$, $\mathbb{P}_{\lambda,\theta} := \mathbb{P}^{X_1,...,X_n}$. Zeigen Sie, dass $T(X) = (\sum_{i=1}^n X_i, \min\{X_1, ..., X_n\})'$ suffizient für $\mathbb{P}_{\lambda,\theta:\theta\in\mathbb{R},\lambda>0}$ ist.
Hinweis: Die Dichte von $\text{vExp}(\lambda, \theta)$ *ist gegeben durch* $f_{\lambda,\theta}(x) = \lambda e^{-\lambda(x-\theta)} \mathbb{1}_{\{x \geq \theta\}(x)}$.

Aufgabe 6.

Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{Bin}(1, p)$ mit $p \in (0, 1)$. Berechnen Sie dem UMVU-Schtzer für $g(p) = p^2$ mit einem beliebigen $m \in \{0, ..., n\}$.

Aufgabe 7.

Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{Poi}(\lambda)$ mit $\lambda \in (0, \infty)$. Berechnen Sie einen UMVU-Schätzer für (i) $g(\lambda) = \lambda$ sowie (ii) $g(\lambda) = e^{-2\lambda}$.

0.3 Äquivariante und Bayes-Schätzer

Aufgabe 8.

Seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{vExp}(\lambda, \theta)$ mit $\theta \in \mathbb{R}$ und $\lambda > 0$. Wir setzen λ als bekannt voraus. Es sei $\mathbb{P}_\theta := \mathbb{P}^{X_1,...,X_n}$. Es ist ferner bekannt, dass $S(x) = \min\{X_1, ..., X_n\}$ eine suffiziente und vollständige Statistik für θ ist.

(a) Zeigen Sie, dass $S(X) \sim \text{vExp}(n\lambda, \theta)$.

(b) Zeigen Sie, dass $S(X)$ unabhängig von $Y = (X_1 - X_n, ..., X_{n-1} - X_n)'$ ist.

(c) Berechnen Sie jeweils einen Minimierer $v^*(y)$ von $\mathbb{E}_0[L(S(x), v^*(Y)) \mid Y = y]$ für $L(s, v) = |s - v|$ bzw. $L(s, v) = |s - v|^2$.

(d) Berechnen Sie jeweils einen MRE-Schätzer $S^*(X)$ für θ bzgl. absoluter und quadratischer Verlustfunktion.

Hinweis: Die Dichte von $\text{vExp}(\lambda, \theta)$ *ist gegeben durch* $f_{\lambda,\theta}(x) = \lambda e^{-\lambda(x-\theta)} \mathbb{1}_{\{x \geq \theta\}(x)}$.

Aufgabe 9.

Sei $\sigma^2 > 0$ fest. Gegeben seien $X_1, ..., X_n \overset{\text{iid}}{\sim} \text{t}\mathcal{N}(\mu, \sigma^2)$ mit Dichte $f_{\mu,\sigma}(x) = \frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathbb{1}_{\{x \geq \mu\}}(x)$. Berechnen Sie den Pitman-Schätzer für μ (bzgl. quadratischer Verlustfunktion).

Aufgabe 10.

Es seien $\mathbb{P}^{X_1|\Theta=\theta} = \mathrm{vExp}(\theta, \alpha)$ mit Dichte $f_{X_1|\Theta=\theta}(z) = \alpha e^{-\alpha(z-\theta)} \mathbb{1}_{\{z \geq \theta\}}(z)$ und $\mathbb{P}^{\Theta} = \mathrm{Exp}(\lambda)$ bei vorgegebenem, festen $\lambda > 0$. Berechnen Sie die A-Posteriori $\mathbb{P}^{\Theta|X}$ mit $X = (X_1, ..., X_n)'$, wobei $X_1, ..., X_n$ gegeben Θ unabhängig und identisch verteilt seien. Berechen Sie den Bayes-Schätzer bzgl. absoluter und quadratischer Verlustfunktion.

Aufgabe 11.

Die A-Priori-Verteilung sei gegeben durch $\Theta \sim \mathrm{Exp}(\lambda)$, $\lambda > 0$ fest. Die Likelihood der Beobachtung seien $X_1 \mid \Theta \sim \mathrm{Exp}(\theta)$ und $X = (X_1, ..., X_n)'$, wobei $X_1, ..., X_n$ gegeben Θ unabhängig und identisch verteilt sind. Berechen Sie den Bayes-Schätzer bzgl. quadratischer Verlustfunktion.

Hinweis: $\int_0^\infty x^m e^{-xc} \, \mathrm{d} = m! c^{-(m+1)}$, $m \in \mathbb{N}$ *und* > 0

0.4 Bedingte Tests

Aufgabe 12 (Vergleich zweier Pareto-Verteilungen).

Gegeben seien unabhängige Zufallsvariablen $X_1, ..., X_n$, $Y_1, ..., Y_m$, wobei $X_i \sim \mathrm{Pareto}(p_1)$, $(i = 1, ..., n)$ und $Y_j \sim \mathrm{Pareto}(p_2)$ $(j = 1, ..., m)$. Die Dichte einer $\mathrm{Pareto}(p)$-Verteilung ist gegeben durch

$$f_p(x) = p x^{-(p+1)} \, \mathbb{1}_{[1,\infty)}(x).$$

Wir wollen die Hypothesen $H_0 : p_1 \leq p_2$ gegen $H_1 : p_1 > p_2$ testen. Gehen Sie dazu wie folgt vor:

(a) Zeigen Sie, dass mit $\theta := p_1 - p_2$ und $\eta := p_1 + 1$ die Verteilungsfamilie $\mathbb{P}_{\theta,\eta} := \mathbb{P}^{X_1,...X_n,Y_1,...,Y_m}$ eine strikt 2-parametrische Exponentialfamilie in $(\theta, \eta) \in \mathbb{R} \times (0, \infty)$ ist und bestimmen Sie die zugehörigen Statistiken $U, T : \mathbb{R}^{n+m} \to \mathbb{R}$.

(b) Geben Sie einen gleichmäßig besten unverfälschten Test $\phi^*(x, y)$ für die Hypothesen H_0 gegen H_1 zum Niveau $\alpha \in (0, 1)$ mit Bestimmungsgleichung für die im Test auftretende Konstante $c(t)$ an.

(c) Zeigen Sie, dass unter $\mathbb{P}_{0,\eta}$ die Statistiken

$$V(x, y) := \frac{\sum_{j=1}^m \log(y_j)}{\sum_{i=1}^n \log(x_i) + \sum_{j=1}^m \log(y_j)}, \qquad T(x, y) := -\sum_{i=1}^n \log(x_i) - \sum_{j=1}^m \log(y_j)$$

unabhängig sind.

(d) Zeigen Sie, dass der gleichmäßig beste Test aus (b) äquivalent umgeformt werden kann zu

$$\phi^*(x, y) = \begin{cases} 1, & V(x, y) > c^*, \\ 0, & V(x, y) \leq c^*, \end{cases}$$

wobei c^* das $(1 - \alpha)$-Quantil der Verteilung einer Zufallsvariable $\frac{Z_2}{Z_1 + Z_2}$, Z_1, Z_2 unabhängig und $Z_1 \sim \Gamma(n, 1)$, $Z_2 \sim \Gamma(m, 1)$, ist.

Hinweis: Nutzen Sie ohne Beweis die Faltungseigenschaft der Gamma-Verteilungen: Für zwei unabhängige Zufallsvariablen $W_1 \sim \Gamma(p_1, \beta)$, $W_2 \sim \Gamma(p_2, \beta)$ *gilt* $W_1 + W_2 \sim \Gamma(p_1 + p_2, \beta)$. *Außerdem gilt für* $W_1 \sim \mathrm{Pareto}(p) : \log(W_1) \sim \Gamma(1, p)$.

Aufgabe 13 (Testen mit störenden Parametern).

Gegeben seien unabhängige Zufallsvariablen $X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$. Wir wollen die Hypothesen $H_0 : \sigma \leq \sigma_0$ gegen $H_1 : \sigma > \sigma_0$ testen, wobei β unbekannt ist und

$p_0 \in (0,\infty)$ fest vorgegeben. Zeigen Sie, dass ein gleichmäßig bester unverfälschter Test $\phi^*(x)$ für die Hypothesen H_0 gegen H_1 zum Niveau $\alpha \in (0,1)$ gegeben ist durch

$$\phi^*(x) = \begin{cases} 1, & V(x) > \sigma_0^2 \chi^2_{n,1-\alpha}, \\ 0, & V(x) \le \sigma_0^2 \chi^2_{n,1-\alpha} \end{cases}, \qquad V(x) := \sum_{i=1}^n (x_i - \overline{x}_n)^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n}\Big(\sum_{i=1}^n x_i\Big)^2,$$

wobei $\chi^2_{n,1-\alpha}$ das $(1-\alpha)$-Quantil der χ^2_n-Verteilung ist.
Hinweis: Nutzen Sie ohne Beweis, dass für $X_1, ..., X_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_0^2)$, $X = (X_1, ..., X_n)'$ gilt: $\frac{V(X)}{\sigma_0^2} \sim \chi^2_{n-1}$.

0.5 Nichtparametrik

Aufgabe 14.
Sei $X_1, ..., X_n$ eine iid Stichprobe reeller Zufallsvariablen mit Dichte f bzgl. des Lebesguemaßes, wo bei f Lipschitz-stetig mi Träger $\subseteq [0,1]$ und $f(0) > 0$ ist. Sei ferner $K : \mathbb{R} \to \mathbb{R}$ eine Funktion mit Träger $\subseteq [-1,1]$ und $\int K(x)\,\mathrm{d}x = 1$. Sei $c \in (0,1)$ und $h = h_n > 0$ eine Nullfolge. Es bezeichne $\hat{f}_h$ den Kerndichteschätzer für f.

(a) Zeigen Sie, dass $\mathbb{E}[\hat{f}_h(ch)] - f(ch) \to \kappa \neq 0$.

(b) Wir betrachten nun eine erweiterte Stichprobe $Y_1, ..., Y_n$ mit $Y_i := -X_i$. Zeigen Sie, dass für
$$f_h^*(x) := \frac{1}{nh}\sum_{i=1}^n K\Big(\frac{X_i - x}{h}\Big) + \frac{1}{nh}\sum_{i=1}^n K\Big(\frac{Y_i - x}{h}\Big)$$
gerade $\mathbb{E}[f_h^*(ch)] - f(ch) \to 0$ gilt.

Aufgabe 15.
Sei $X_1, ..., X_n$ eine Stichprobe reeller Zufallsvariablen mit stetiger Dichte f bzgl. des Lebesguemaßes und $h = h_n > 0$ eine Nullfolge. Sei K eine Kernfunktion mit kompakten Träger. Sei $\hat{f}_h$ der zugehörige Kerndichteschätzer. Zeigen Sie, dass für zwei Punkte $x_1, x_2 \in \mathbb{R}$ mit $x_1 \neq x_2$ gerade $\operatorname{Cov}(\hat{f}_h(x_1), \hat{f}_h(x_2)) \to 0$ gilt.

Aufgabe 16.
Seien $X_1, ..., X_n$ iid reellwertige Zufallsvariablen mit stetiger und beschränkter Dichte f. Seien $\varepsilon_1, ..., \varepsilon_n$ iid unabhängig der X_i's mit $\mathbb{E}[\varepsilon_i] = 0$ und $\operatorname{Var}(\varepsilon_i) = \sigma^2 > 0$. Wir beobachten (X_i, Y_i), wobei $Y_i = m(X_i) + \varepsilon_i$ für eine stetige, beschränkte Funktion $m : \mathbb{R} \to \mathbb{R}$. Wir definieren

$$\hat{m}_h(x) := \frac{\sum_{i=1}^n Y_i K\Big(\frac{X_i - x}{h}\Big)}{\sum_{i=1}^n K\Big(\frac{X_i - x}{h}\Big)}$$

für eine beschränkte Funktion $K : \mathbb{R} \to \mathbb{R}$ mit kompaktem Träger, $\int K(x)\,\mathrm{d}x = 1$ und einer Bandbreite $h = h_n > 0$. Sei $x \in \mathbb{R}$ fest mit $f(x) > 0$. Zeigen Sie: Gilt $h \to 0$ und $nh \to \infty$, so folgt $\hat{m}_h(x) \xrightarrow{\mathbb{P}} m(x)$.
Hinweis: Sie wissen bereits, dass $\hat{f}_h(x) \xrightarrow{\mathbb{P}} f(x)$ gilt.

Hinweis:
Beachten Sie, dass das Kapitel über bedingte Tests aus der Vorlesung klausurrelevant sein kann. Es können Aufgaben ähnlich wie im Abschnitt 0.4 thematisiert werden.
Wir wünschen Ihnen viel Erfolg bei der Vorbereitung!

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