UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

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

10. Übungsblatt

Aufgabe 37	Aufgabe 38	Aufgabe 39	Aufgabe 40	Summe:

Übungsgruppe: Tutor(in):

Namen:

Aufgabe 37 (Exkurs: Konstruktion von Kernen ℓ-ter Ordnung, 4 = 1 + 2 + 1 Bonuspunkte).

Sei $\ell \in \mathbb{N}$. Wir wollen in dieser Aufgabe Kerne mit höheren Ordnung konstruieren, d.h. $K : \mathbb{R} \to \mathbb{R}$ mit $\int K(x)\,\mathrm{d}x = 1$ und

$$\int x^j K^j(x)\,\mathrm{d}x = 0 \quad \text{für} \quad j = 1, ..., \ell.$$

(a) Zeigen Sie, dass ein Kern K der Ordnung 2 negative Werte annehmen muss.

Es sei ℓ ene vorgegebene Ordnung und $k : \mathbb{R} \to \mathbb{R}$ eine Funktion mit $\int k(x)\,\mathrm{d}x = 1$ und $\int x^{2\ell} k(x)\,\mathrm{d}x < \infty$. Für $j = 0, ..., 2\ell$ setzen wir $c_j = \int x^j k(x)\,\mathrm{d}x$. Wir definieren

$$K(x) = k(x) \cdot \det \begin{pmatrix} c_1 & c_2 & \dots & c_\ell & 1 \\ c_2 & c_3 & \dots & c_{\ell+1} & x \\ \vdots & & \ddots & & \vdots \\ c_{\ell+1} & c_{\ell+2} & \dots & c_{2\ell} & x^\ell \end{pmatrix} \Big/ \det \begin{pmatrix} c_1 & c_2 & \dots & c_\ell & c_0 \\ c_2 & c_3 & \dots & c_{\ell+1} & c_1 \\ \vdots & & \ddots & & \vdots \\ c_{\ell+1} & c_{\ell+2} & \dots & c_{2\ell} & c_\ell \end{pmatrix}$$

(b) Zeigen Sie, dass K ein Kern der Ordnung ℓ ist.

(c) Berechnen Sie einen Kern der Ordnung $\ell = 2$ mit $k(x) = \frac{1}{2}\mathbb{1}_{[-1,1]}(x)$.

Aufgabe 38 (Schätzung der Ableitungen von Dichten, 4 = 1.5 + 1.5 + 1 Punkte).

Sei $X_1, ..., X_n \in \mathbb{R}$ eine unabhängig und identisch verteilte Stichprobe mit Dichte $f \in \Sigma(\beta, L)$. Es gelte ferner $h > 0$ und K sei ein beschränkter Kern mit Träger $\subseteq [-1, 1]$, welche mit $l = \lfloor \beta \rfloor$, $s < \beta$ folgende zwei Eigenschaften erfüllt:

$$\int_{-1}^{1} u^j K(u)\,\mathrm{d}u = 0, \qquad j \in \{0, ..., l\} \setminus \{s\}, \tag{1}$$

$$\int_{-1}^{1} u^s K(u)\,\mathrm{d}u = s!. \tag{2}$$

Ein Kernschätzer für die s-te Ableitung $f^{(s)}$ ergibt sich dann durch

$$\hat{f}_{h,s}(x) := \frac{1}{nh^{s+1}} \sum_{i=1}^{n} K\Big(\frac{X_i - x}{h}\Big)$$

Sei $x \in \mathbb{R}$ fest gewählt. Zeigen Sie:

(a) Für den Bias $b_f(x) = \mathbb{E}_f[\hat{f}_{h,s}(x)] - f^{(s)}(x)$ gilt mir einer Konstanten $c \in \mathbb{R}$,

$$\sup_{f \in \mathcal{F}(\beta,L)} |b_f(x)| \leq c \cdot h^{\beta - s}.$$

(b) Für die Varianz $\mathrm{Var}_f(\hat{f}_{h,s}(x))$ gilt mit einer Konstanten $c' \in \mathbb{R}$,

$$\sup_{f \in \mathcal{F}(\beta,L)} \mathrm{Var}_f(\hat{f}_{h,s}(x)) \leq \frac{c'}{nh^{2s+1}}.$$

(c) Folgern Sie aus Teilaufgaben (a) und (b), dass $h^* = h_n^*$ so gewählt werden kann, dass für den Mean-Squared-Error gilt:

$$\sup_{f \in \mathcal{F}(\beta,L)} \mathbb{E}_f[\hat{f}_{h^*,\,s}(x) - f^{(s)}(x))^2] = O(n^{-\frac{2(\beta-s)}{2\beta+1}}).$$

Aufgabe 39 (Gleichmäßige Konvergenz von Dichteschätzern, 4 = 1 + 1.5 + 1.5 Bonuspunkte).

In dieser Aufgabe werden wir zeigen, dass der Kerndichteschätzer nicht nur punktweise, sondern gleichmäßig auf Kompakta gegen die wahre Dichte f konvergiert. Seien dazu $X_1, ..., X_n$ unabhängig und identisch verteilt mit stetiger Dichte f und $h = h_n \to 0$ eine Folge von Bandbreiten mit $\frac{nh}{\log(n)} \to \infty$. Sei $K : \mathbb{R} \to \mathbb{R}$ eine Lipschitz-stetige Funktion mit $\int K(x)\,\mathrm{d}x = 1$ und kompaktem Träger $\subseteq [-1, 1]$. Wir betrachten den Kerndichteschätzer

$$\hat{f}_h(x) := \frac{1}{nh} \sum_{i=1}^{n} K\big(\frac{X_i - x}{h}\big).$$

Sei $M \subset \mathbb{R}$ eine kompakte Menge und M_n eine Diskretisierung von M, sodass sich für jedes $x \in M$ eine Zahl $x' \in M_n$ finden lässt mit $|x - x'| \leq c_n := n^{-2}$. Es gelte $|M_n| \leq \lceil \frac{\mathrm{diam}(M)}{c_n} \rceil$.

(a) Zeigen Sie, dass es für jedes $\varepsilon > 0$ ein $C(\varepsilon) > 0$ unabhängig von n, x gibt, sodass

$$\mathbb{P}(|\hat{f}_h(x) - \mathbb{E}\hat{f}_h(x)| > \varepsilon) \leq 2\exp(-C(\varepsilon) \cdot nh).$$

Hinweis: Nutzen Sie die Bernstein-Ungleichung für unabhängige und identisch verteilte Zufallsvariablen $Z_1, ..., Z_n$. Ist $\mathbb{E}Z_1 = 0$, $\mathrm{Var}(Z_1) \leq \sigma^2$ und $L > 0$ eine Konstante mit $|Z_1| \leq L$ f.s., so gilt $\mathbb{P}(|\sum_{i=1}^n Z_i| \geq a) \leq 2\exp\Big(-\frac{a^2}{2(aL+n\sigma^2)}\Big)$ für alle $a > 0$.

(b) Nutzen Sie die Lipschitz-Stetigkeit von K, um zu zeigen, dass für jedes $\varepsilon > 0$ mit n groß genug gilt

$$\mathbb{P}\Big(\sup_{|x-x'| \leq c_n} |(\hat{f}_h(x) - \mathbb{E}\hat{f}_h(x)) - (\hat{f}_h(x') - \mathbb{E}\hat{f}_h(x'))| > \varepsilon\Big) = 0.$$

(c) Zeigen Sie mit Teilaufgaben (a) und (b), dass $\sup_{x \in M} |\hat{f}_h(x) - f(x)| \xrightarrow{p} 0$.
Hinweis: Verwenden Sie zuerst die Abschätzung $\mathbb{P}(a + b > c) \leq \mathbb{P}(a > c/2) + \mathbb{P}(b > c/2)$ für geeignete Terme a, b und $c > 0$, und entfernen dann den Bias-Term. Nutzen Sie weiter für eine geeignete Funktion $g : \mathbb{R} \to \mathbb{R}$ die Zerlegung $\sup_{x \in M} |g(x)| \leq \sup_{|x-x'| \leq c_n} |g(x) - g(x')| + \sup_{x' \in M_n} |g(x)|$ und die Abschätzung $\mathbb{P}(\sup_{x' \in M_n} |g(x')| \geq \varepsilon/2) \leq |M_n| \cdot \sup_{x' \in M_n} \mathbb{P}(|g(x')| \geq \varepsilon/2)$.

Aufgabe 40 (Wiederholung: Konstruktion gleichmäßig bester Tests bei stetigen Verteilungen, 4 Punkte).

Eine Forschungsgruppe untersucht die Halbwertszeit von ^{60}Co. Die bisher vermutete Halbwertszeit beträgt $\lambda_0 = 5.27$ Jahre. Die Forscher beobachten nun in $n = 5$ unabhängigen Experimenten Halbwertszeiten

X_1	X_2	X_3	X_4	X_5
5.75	4.5	6.0	6.5	5.25

.

In ihrem Modell nehmen sie an, dass $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\frac{1}{\lambda})$.
Die Forschungsgruppe will basierend auf ihren Messungen nachweisen, dass die wahre Halbwertszeit größer als die bisher vermutete ist. Formulieren Sie einen gleichmäßig besten Test zum Niveau $\alpha = 0.05$ und entscheiden Sie, ob ein Nachweis einer größeren Halbwertszeit mit den Ergebnissen der Forscher möglich ist.
Hinweis: Sie können ohne Einschränkungen hier eine einseitige Nullhypothese festlegen. Es gilt $\sum_{i=1}^n X_i = \Gamma(n, \frac{1}{\lambda})$. *Das* 0.95*-Quantil der* $\Gamma(5, \frac{1}{\lambda_0})$*-Verteilung ist* $\Gamma(5, \frac{1}{\lambda_0}, 0.95) = 48.25$.

Abgabe:
In Zweiergruppen, bis spätestens Donnerstag, den **04.02.2021, 11:00 Uhr**.
Bitte geben Sie Ihre Lösungsvorschläge im PDF-Format ab. Es reicht, wenn eine der beiden Personen das Dokument hochlädt.

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