UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

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

9. Übungsblatt

Aufgabe 33	Aufgabe 34	Aufgabe 35	Aufgabe 36	Summe:

Übungsgruppe: Tutor(in):

Namen:

Aufgabe 33 (Gleichmäßig besserer und zulässiger Schätzer, 4 = 1 + 1 + 0.5 + 1.5 Punkte).

Sei $\mathcal{S}$ eine vorgegebene Menge von Schätzern. Ein Schätzer $S \in \mathcal{S}$ heißt *gleichmäßig besser* als ein Schätzer $S' \in \mathcal{S}$, wenn für alle $\theta \in \Theta$, $R(\theta, S) \leq R(\theta, S')$ gilt und ein θ_0 existiert, so dass $R(\theta_0, S) < R(\theta_0, S')$. Ein Schätzer heißt *zulässig* in $\mathcal{S}$, wenn es keinen gleichmäßig besseren Schätzer in $\mathcal{S}$ gibt. Zeigen Sie folgende Aussagen:

(a) Ist S ein Minimax-Schätzer und eindeutig (in $\mathcal{S}$) in dem Sinne, dass jeder andere Minimax-Schätzer die gleiche Risikofunktion R besitzt, so ist S zulässig in $\mathcal{S}$.

(b) Ist S zulässig (in $\mathcal{S}$) mit konstanter Risikofunktion R (in θ), so ist S ein Minimax-Schätzer.

(c) Ist S_Λ ein Bayes-Schätzer bezüglich einer Verteilung Λ und eindeutig (in $\mathcal{S}$) in dem Sinne, dass jeder andere Bayes-Schätzer S'_Λ bezüglich Λ die gleiche Risikofunktion R besitzt, so ist S zulässig (in $\mathcal{S}$).

(d) Sei Θ nun ein metrischer Raum ausgestattet mit der Borel-σ-Algebra. Ist S_Λ ein Bayes-Schätzer (in $\mathcal{S}$) bezüglich einer Verteilung Λ, so ist S_Λ zulässig in $\mathcal{S}$, falls

  (a) $R(\theta, S_\Lambda) < \infty$,

  (b) jede offene Menge $U \in \Theta$, $\Lambda(U) > 0$,

  (c) für jeden weiteren Schätzer S' (in $\mathcal{S}$) mit $R_\Lambda(S') \leq R_\Lambda(S)$, die Abbildung $\theta \mapsto R(\theta, S')$ stetig ist.

Aufgabe 34 (Teil 1: Beispiel für nichtparametrische Regression - der gleitende Mittelwert, 3 = 2 + 1 Punkte).

Sei $n \in \mathbb{N}$. Wir betrachten das einfache Regressionsmodell

$$X_i = f(i/n) + \varepsilon_i, \qquad i = 1, ..., n.$$

Dabei sind ε_i identisch und unabhängig verteilte Zufallsvariablen mit $\mathbb{E}[\varepsilon_1] = 0$ und $\mathrm{Var}(\varepsilon) = 1$ und $f : [0,1] \to \mathbb{R}$ eine stetige Funktion. *Wir gehen also davon aus, dass für eine wachsende*

Anzahl von Beobachtungen n, die Funktion f genauer abgetastet wird. Wir wollen also aus den Beobachtungen die Funktion f schätzen. Sei dazu $h = h_n \to 0$ für $n \to \infty$ eine Folge nichtnegativer reeller Zahlen (die „Bandbreite"). Für $x \in (0,1)$ definieren wir den *gleitenden Mittelwert* durch

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \mathbb{1}_{[x-\frac{h}{2}, x+\frac{h}{2}]}\big(\frac{i}{n}\big) X_i.$$

Wir wollen hierbei zeigen, dass $\hat{f}_h(x) \xrightarrow{\mathbb{P}} f(x)$ für $nh_n \to \infty$ und $x \in (0,1)$ gilt, das heißt $\hat{f}_h$ ist ein punktweise konsistenter Schätzer für f. Zeigen Sie dazu:

(a) $\mathbb{E}[\hat{f}_h(x)] - f(x) \to 0$ für $n \to \infty$
Hinweis: Sie können auf elementarem Weg geschickt eine komplizierte Null mit dem Term $\pm f(x)\frac{1}{nh}\sum_{i=1}^{n} \mathbb{1}_{[x-\frac{h}{2}, x+\frac{h}{2}]}\big(\frac{i}{n}\big)$ einfügen und abschätzen. Alternativ verwenden Sie den Mittelwertsatz der Integralrechnung.

(b) $\mathrm{Var}(\hat{f}_h(x)) \to 0$ für $n \to \infty$

Dann folgt $\mathbb{E}[(\hat{f}_h(x) - f(x))^2] \to 0$ für $n \to \infty$ und insbesondere die Konsistenz von $\hat{f}_h(x)$.

Aufgabe 35 (Teil 2: Asymptotik des gleitenden Mittelwertes, 5 = 2 + 2 + 1 Punkte).

Wir betrachen das Regressionsmodell und den dazugehörigen Schätzer $\hat{f}_h(x)$ aus Aufgabe 34. Sei f nun zusätzlich Lipschitz-stetig und es existiere $\mathbb{E}|\varepsilon_1|^3 < \infty$. Wir wollen zeigen, dass für $h = h_n \to 0$ mit $h_n \leq 1$, $nh \to \infty$ und $\sqrt{nh^3} \to 0$ die punktweise asymptotische Normalität

$$\sqrt{nh}(\hat{f}_h(x) - f(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2) \tag{$*$}$$

für jedes $x \in (0,1)$ gilt. Zeigen Sie dazu unter Nutzung von Rechnungen aus Aufgabe 34 folgende Aussagen:

(a) $|\mathbb{E}[\hat{f}_h(x)] - f(x)| \leq C(h + \frac{1}{nh})$ für eine von n unabhängige Konstante $C > 0$.

(b) Es gilt
$$\sqrt{nh}(\hat{f}_h(x) - \mathbb{E}[\hat{f}_h(x)]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$
Hinweis: Benutzen Sie den zentralen Grenzwertsatz von Ljapunov (Dreiecksschema): Für jedes $n \in \mathbb{N}$ seien $Z_{n,i}$, $1 \leq i \leq n$, unabhängige Zufallsvariablen mit $\mathbb{E}[Z_{n,i}] = 0$, $\sigma_{n,i}^2 = \mathbb{E}[Z_{n,i}^2] < \infty$ und $s_n^2 = \sum_{i=1}^{n} \sigma_{n,i}^2$. Falls ein $\delta > 0$ existiert, so dass
$$\frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|Z_{n,i}|^{2+\delta}] \to 0,$$
dann gilt
$$\frac{1}{s_n} \sum_{i=1}^{n} Z_{n,i} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$
Betrachten Sie für die Aufgabe $\delta = 1$.

(c) Folgern Sie nun mittels Teilaufgabe (a) und (b) die Konvergenz in $(*)$.

Aufgabe 36 (Der Fluch der Dimension, 4 = 1.5 + 1.5 + 1 Punkte).

In dieser Aufgabe werden wir sehen, dass die Konvergenzrate $n^{-\frac{2\beta}{2\beta+1}}$ (durch eine optimale Wahl der Bandbreite) des Kerndichteschätzers in einer Dimension nicht mehr erreicht werden kann, wenn stattdessen Boebachtungen in höherer Dimension d vorliegen. In diesem Fall

sinkt die Konvergenzrate auf $n^{-\frac{2\beta}{2\beta+d}}$. *Wir zeigen dies exemplarisch für den Fall* $d = 2$.
Sei $(X_1, Y_1), ..., (X_n, Y_n) \in \mathbb{R}$ eine zweidimensionale identisch und unabhängig verteilte Stichprobe mit beschränkter Dichte $f : \mathbb{R}^2 \to \mathbb{R}$ bezüglich des Lebesguemaßes und $h = h_n \to 0$. Ferner erfülle f die Hölderbedingung

$$\forall (x, y),\ (x', y') \in \mathbb{R}^2 : \quad |f(x, y) - f(x', y')| \leq L \cdot \Big(|x - x'|^\beta + |y - y'|^\beta\Big)$$

mit den Konstanten $0 < \beta \leq 1$ und $L > 0$.
Sei weiterhin $K : \mathbb{R} \to \mathbb{R}$ eine beschränkte Funktion mit $\int K(x)\ dx = 1$ und kompaktem Träger auf $[-N, N]$ mit einem festen $N > 0$.
Es bezeichne $\hat{f}_h$ den zweidimensionalen Kerndichteschätzer

$$\hat{f}_h(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \cdot K\left(\frac{Y_i - y}{h}\right), \qquad (x, y) \in \mathbb{R}^2.$$

Sei $(x_0, y_0) \in \mathbb{R}^2$ fest gewählt. Zeigen Sie:

(a) Mit einer Konstanten $c > 0$ unabhängig von (x_0, y_0) gilt für die Varianz des Kerndichteschätzers gilt

$$\mathrm{Var}_f(\hat{f}_h(x_0, y_0)) \leq c \cdot \frac{1}{nh^2}.$$

(b) Für den Bias gilt mit einer Konstanten $c' > 0$ unabhängig von (x_0, y_0),

$$|\mathbb{E}_f[\hat{f}_h(x_0, y_0)] - f(x_0, y_0)| \leq c' \cdot h^\beta.$$

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

$$\mathrm{MSE}_f(\hat{f}_{h^*}(x_0, y_0)) = O\big(n^{-\frac{2\beta}{2\beta+2}}\big).$$

Abgabe:
In Zweiergruppen, bis spätestens Donnerstag, den **28.01.2021, 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`