Dr. Stefan Richter, Timo Dörzbach
Wahrscheinlichkeitstheorie 2
Sommersemester 2023

Präsenzblatt 7

Aufgabe P1 (Konvergenz der Verteilungsfkt. der Residuen bei einer Regression).
Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Betrachtet wird das einfache Regressionsmodell

$$Y_i = a_0 + Z_i, \quad i = 1, ..., n$$

wobei $a_0 \in \mathbb{R}$ ein unbekannter Parameter ist, und Z_i i.i.d. Zufallsvariablen mit $\mathbb{E}Z_1 = 0$, $\mathbb{E}[Z_1^2] < \infty$. Sei $P = \mathbb{P}^{Z_1}$, und es bezeichne $\hat{P}_n$, $\mathbb{G}_n$ die empirischen Maße bzw. Prozesse bzgl. $Z_1, ..., Z_n$.

(a) Zeigen Sie, dass der Kleinste-Quadrate-Schätzer

$$\hat{a}_n = \arg\min_{a \in \mathbb{R}} \hat{R}_n(a), \qquad \hat{R}_n(a) = \frac{1}{n} \sum_{i=1}^n (Y_i - a)^2$$

durch $\hat{a} = \overline{Y_n} = \frac{1}{n} \sum_{i=1}^n Y_i$ gegeben ist.

(b) Betrachtet werden nun die empirischen Residuen

$$\hat{Z}_i = Y_i - \hat{a}_n, \quad i = 1, ..., n$$

Definiere

$$\tilde{F}_Z(z) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{\hat{Z}_i \leq z\}},$$

und für $z \in \mathbb{R}, a \in \mathbb{R}$:

$$f_{z,a}(y) = \mathbb{1}_{\{y \leq z - (a_0 - a)\}}$$

Zeigen Sie, dass gilt:

$$\tilde{F}_Z(z) = \hat{P}_n f_{z,\hat{a}_n}.$$

(c) Begründen Sie, dass

$$\mathcal{F} = \{f_{z,a} : z \in \mathbb{R}, a \in \mathbb{R}\}$$

erfüllt: $N_2(\varepsilon, \mathcal{F}, P) \leq \frac{2}{\varepsilon^2}$ und $J_2(1, \mathcal{F}, P) < \infty$.

Ab nun sei $z_0 \in \mathbb{R}$ fest gewählt. Es soll gezeigt werden, dass

$$\sqrt{n}\big(\tilde{F}_Z(z_0) - F_Z(z_0)\big) \xrightarrow{D} N(0, \sigma^2), \qquad (*)$$

wobei σ^2 zu bestimmen ist. Sei dazu angenommen, dass F_Z in z_0 differenzierbar ist.

(d) Sei $\hat{f}_n := f_{z_0,\hat{a}_n}$ und $f_0 := f_{z_0,a_0}$. Zeigen Sie, dass

$$\int |\hat{f}_n(x) - f_0(x)|^2 dP(x) \xrightarrow{\mathbb{P}} 0.$$

(e) Folgern Sie mit Lemma 1.72: $\mathbb{G}_n(\hat{f}_n - f_0) \xrightarrow{\mathbb{P}} 0$.

(f) Zeigen Sie, dass

$$\sqrt{n}\big(\tilde{F}_Z(z_0)-F_Z(z_0)\big) = \mathbb{G}_n(\hat{f}_n-f_0)+\sqrt{n}(F_Z\big(z_0-(a_0-\hat{a}_n)\big)-F_Z(z_0)\big)+\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{Z_i\le z_0\}}-F_Z(z_0)\Big)$$

(g) Zeigen Sie (*) und ermitteln Sie σ^2.
Hinweis: Delta-Methode in $\mathbb{R}$.

Aufgabe P2 (Abschätzung eines Supremums).

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $(\mathcal{X}, d)$ ein metrischer Raum und $X_i : \Omega \to \mathcal{X}$, $i \in \mathbb{N}$ i.i.d. Zufallsvariablen. Sei $P = \mathbb{P}^{X_1}$. Sei $\mathcal{F} = \{f : \mathcal{X} \to \mathbb{R} \text{ messbar}\}$ und $L : \mathcal{X} \to \mathbb{R}$ eine Einhüllende von $\mathcal{F}$. Zeigen Sie zunächst, dass

$$\mathbb{E}^* \sup_{f,g\in\mathcal{F}, P((f-g)^2)\le\delta^2} |\mathbb{G}_n(f-g)| \lesssim J_2(\delta, \mathcal{F}, P) + \sqrt{n}P\big[L\mathbb{1}_{\{2L>\sqrt{n}a(\delta)\}}\big],$$

wobei $a(\delta) = \frac{\delta}{\sqrt{Log N_2(\delta,\mathcal{F},P)}}$.

Es bezeichne $\hat{F}_n$ die empirische Verteilungsfunktion von X_1 und $F(x) = \mathbb{P}(X \le x)$. Zeigen Sie, dass für $\delta \le 1$:

$$\mathbb{E} \sup_{|F(x)-F(y)|\lesssim\delta^2} \big|\big(\hat{F}_n(x) - F(x)\big) - \big(\hat{F}_n(y) - F(y)\big)\big| \lesssim \frac{1}{\sqrt{n}} \cdot \Big(\delta\sqrt{Log(1/\delta)} + \frac{\sqrt{Log(1/\delta)}}{\delta}\Big)$$

Hinweis: Nutzen Sie ohne Beweis, dass für $\delta \le 1$ gilt: $\int_0^\delta \sqrt{Log(1/\varepsilon)}d\varepsilon \lesssim \delta\sqrt{Log(1/\delta)}$.

Vorlesungswebseite:
`https://ssp.math.uni-heidelberg.de/wt2-ss23/`