Institut für Angewandte Mathematik
Sommersemester 2024
Andreas Eberle, Stefan Oberdörster

11. Übungsblatt „Einführung in die Statistik"

Abgabe bis Dienstag 2.7.

---

1. (Robustheitsmaße) Seien $X_1, X_2, \ldots$ unabhängige reelle Stichprobenwerte von einer Wahrscheinlichkeitsverteilung μ mit Verteilungsfunktion F und stetiger Dichtefunktion f.

a) Bestimmen Sie die asymptotischen Bruchpunkte der folgenden Statistiken:

(i) Spannweite, (ii) Stichprobenquantile $\hat{q}_\gamma$ $(0 < \gamma < 1)$, (iii) MAD.

b) Skizzieren Sie die Sensitivitätsfunktion des getrimmten Mittelwerts mit Parameter $\tau \in (0, 1/2)$.

c) Bestimmen Sie für $h > 0$ und $x \in \mathbb{R}$ den Median der gestörten Verteilung $\mu_h = (1-h)\mu + h\delta_x$, und berechnen Sie die Einflussfunktion des Medians.

2. (Stochastische Dominanz) Seien μ und ν zwei Wahrscheinlichkeitsverteilungen auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, und seien $G_\mu(u) = \inf\{c \in \mathbb{R} : F_\mu(c) \geq u\}$ und $G_\nu(u)$ die verallgemeinerten Inversen der Verteilungsfunktionen von μ und ν.

a) Zeigen Sie, dass die folgenden Aussagen alle äquivalent sind:

(i) μ wird von ν *stochastisch dominiert* $(\mu \preceq \nu)$, d.h.

$$\mu[(c,\infty)] \;\leq\; \nu[(c,\infty)] \quad \text{für alle } c \in \mathbb{R}.$$

(ii) Für alle $c \in \mathbb{R}$ gilt $F_\mu(c) \geq F_\nu(c)$.

(iii) Für alle $u \in (0,1)$ gilt $G_\mu(u) \leq G_\nu(u)$.

(iv) Es gibt Zufallsvariablen X und Y mit Verteilungen μ und ν auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, so dass $\mathbb{P}[X \leq Y] \;=\; 1$.

(v) Für jede monoton wachsende Funktion $h : \mathbb{R} \to [0,\infty)$ gilt $\int h\, d\mu \leq \int h\, d\nu$.

b) Zeigen Sie für $n, m \in \mathbb{N}$ und $p, q \in [0,1]$:

(i) $\mathrm{Bin}(n,p) \preceq \mathrm{Bin}(n,q)$ genau dann, wenn $p \leq q$.

(ii) Für $n \leq m$ gilt $\mathrm{Bin}(n,p) \preceq \mathrm{Bin}(m,p)$.

3. (Mann-Whitney U-Test a.k.a. Wilkoxon Rangsummentest) Wir beobachten die Realisierungen von $n = k + l$ unabhängigen reellen Zufallsvariablen $X_1, \ldots, X_{k+l}$ mit unbekannten Verteilungen $X_1, \ldots X_k \sim \mu$ und $X_{k+1}, \ldots, X_{k+l} \sim \nu$. Wir nehmen an, dass die Verteilungsfunktionen stetig sind, so dass die Stichproben mit Wahrscheinlichkeit 1 alle verschieden sind. Im Mann-Whitney U-Test für das Testproblem $H_0 : \mu = \nu$ vs. $H_1 : \mu \prec \nu$ verwirft man die Nullhypothese, falls der Wert der *U-Statistik*

$$U = \sum_{i=1}^{k} \sum_{j=k+1}^{k+l} 1_{\{X_i > X_j\}}$$

strikt unterhalb eines vorgegebenen Schwellenwerts c liegt. Hierbei bedeutet $\mu \prec \nu$, dass $\mu \preceq \mu$ und $\nu \neq \mu$, siehe Aufgabe 2.

a) Zeigen Sie, dass sich die U-Statistik auch darstellen lässt als

$$U = W - \frac{k(k+1)}{2} \quad \text{mit} \quad W = \sum_{i=1}^{k} R_i \quad \text{(Wilcoxons Rangsumme).}$$

Hierbei bezeichnet R_i den (fast sicher eindeutigen) Rang von X_i unter $X_1, \ldots X_{k+l}$.

b) Zeigen Sie, dass $\mathbb{E}[U] = kl/2$ und $\mathrm{Var}[U] = kl(k+l+1)/12$.

c) Man kann beweisen, dass die U-Statistik asymptotisch normalverteilt ist, das heißt $(U - \mathbb{E}[U]) / \sqrt{\mathrm{Var}[U]}$ konvergiert für $k, l \to \infty$ in Verteilung gegen die Standardnormalverteilung, siehe z.B. Satz 11.28 im Buch von Georgii. Leiten Sie daraus eine Wahl für den Schwellenwert c her, so dass der Test asymptotisch das Niveau α hat.

d) In einer medizinischen Studie wurde ein physiologischer Parameter bei zwölf Diabetikern und Nichtdiabetikern gemessen. Die Frage ist, ob es einen systematischen Unterschied zwischen beiden Gruppen in Bezug auf diesen Parameter gibt. Ermitteln Sie einen approximativen p-Wert.

Diab.	11,5	12,1	16,1	17,8	24,0	28,8	33,9	40,7	51,3	56,2	61,7	69,2
Nichtdiab.	4,1	6,3	7,8	8,5	8,9	10,4	11,5	12,0	13,8	17,6	24,3	37,2

4. (Wilcoxon Vorzeichen-Rang-Test) Seien $X_1, ..., X_n$ unabhängige reelle Zufallsvariablen mit identischer Verteilung μ und Verteilungsfunktion F. Die Verteilung sei absolutstetig und bezüglich 0 symmetrisch, d.h. $F(-c) = 1 - F(c)$ für alle $c \in \mathbb{R}$. Für jedes $1 \leq i \leq n$ sei $Z_i = 1_{X_i > 0}$ und R_i der Rang von $|X_i|$ unter $|X_1|, \ldots, |X_n|$. Sei $W_n = \sum_{i=1}^{n} Z_i R_i$ die zugehörige Vorzeichen-Rangsumme. Zeigen Sie:

a) Für jedes i sind Z_i und $|X_i|$ unabhängig.

b) Der Zufallsvektor $R = (R_1, \ldots, R_n)$ ist unabhängig von der zufälligen Menge

$$Z = \{1 \leq i \leq n : Z_i = 1\}.$$

c) Z ist gleichverteilt auf der Potenzmenge von $\{1, \ldots, n\}$, und R ist gleichverteilt auf der Menge $\mathcal{S}_n$ aller Permutationen von $\{1, \ldots, n\}$.

d) Für jedes $0 \leq l \leq n(n+1)/2$ gilt

$$\mathbb{P}[W_n = l] = 2^{-n} \left|\{A \subseteq \{1, \ldots, n\} : \sum_{i \in A} i = l\}\right|.$$