Übung 7

1. Die Sensitivitätskurve einer Folge von Schätzern $\widehat{\mu_n}$ ist definiert durch

$$SC_n(x) = (n+1) \cdot (\widehat{\mu_{n+1}}(x_1, \ldots, x_n, x) - \widehat{\mu_n}(x_1, \ldots, x_n)).$$

Sie liefert gemäss der im Abschnitt 6.1 im Skript erläuterten Heuristik eine Approximation für die Einflussfunktion.

a) Mache den Grenzübergang $n \to \infty$ für die Sensitivitätskurve des α-gestutztes Mittels (im Fall $[n\alpha] = [(n+1)\alpha]$; vergleiche Serie 1, Aufgabe 1a)

$$\widehat{\mu_n} = \frac{1}{n - 2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} x_{(i)}.$$

Nimm zur Vereinfachung an, dass die Verteilung F symmetrisch um den Lageparameter θ ist.

b) Beweise unter der Annahme, dass $F(dx) = f(x)dx$ mit strikt positiver Dichte f, dass $Q_\alpha(F) = F^{-1}(\alpha)$ für festes $\alpha \in (0,1)$ die Einflussfunktion

$$IF(x, Q_\alpha, F) = \begin{cases} (\alpha - 1)/f(F^{-1}(\alpha)) & \text{für } x < F^{-1}(\alpha) \\ \alpha/f(F^{-1}(\alpha)) & \text{für } x > F^{-1}(\alpha) \end{cases}$$

hat.

c) Berechne die Einflussfunktion von $\widehat{\mu_n}$ mit Hilfe der Funktionaldefinition der Einflussfunktion: Ergänze die im Skript (S.65 unten) angegebene Berechnung der Einflussfunktion des gestutzten arithmetischen Mittels und benutze b). Vergleiche mit a).

d) Skizziere die Sensitivitätskurve des α-Winsorisierten Mittels (im Fall $[n\alpha] = [(n+1)\alpha]$):

$$\widehat{\mu_n} = \frac{1}{n} \left([n\alpha] \cdot x_{([n\alpha]+1)} + \sum_{i=[n\alpha]+1}^{n-[n\alpha]} x_{(i)} + [n\alpha] \cdot x_{(n-[n\alpha])} \right).$$

(Hierbei wird ein Anteil von je α der Beobachtungen auf beiden Seiten nicht gestutzt, sondern stattdessen auf den $[n\alpha] + 1$.höchsten bzw. -niedrigsten Wert gesetzt.)
Mache den Grenzübergang $n \to \infty$ und vergleiche mit a).

2. Seien $X_1, \ldots, X_n$ i.i.d. mit den drei möglichen Werten 1,2 und 3. Die entsprechenden Wahrscheinlichkeiten seien $p_1 = P[X_i = 1] = \theta^2$, $p_2 = P[X_i = 2] = 2\theta(1-\theta)$ und $p_3 = P[X_i = 3] = (1-\theta)^2$.

Dieses Modell ist in der Genetik bekannt als das *Hardy-Weinberg-Modell*: X_i bezeichnet den Phänotypen AA, Aa oder aa, und θ ist die unbekannte Häufigkeit des Allels A. Obige Wahrscheinlichkeiten ergeben sich als Gleichgewichtszustand bei zufälligen Paarungen.

a) Wir betrachten die Zufallsvariablen $N_k = \sum_{i=1}^n 1_{\{X_i = k\}}$. Welche Verteilung hat $(N_1, N_2, N_3)^T$? Zeige, dass $(N_1/n, N_2/n)^T$ asymptotisch $\mathcal{N}((p_1, p_2)^T, V/n)$-verteilt ist, und bestimme die asymptotische Kovarianzmatrix V.

b) Wir betrachten die folgenden 2 Schätzer für θ: $T_n^{(1)} = \sqrt{N_1/n}$ und $T_n^{(2)}$ =Maximum-Likelihood Schätzer. Zeige, dass beide Schätzer konsistent und asymptotisch normal sind, und berechne die Varianzen der Grenzverteilungen. Welchen Schätzer würdest Du vorziehen?

3. **a)** Bestimme eine varianzstabilisierende Transformation für

$$X_i \text{ i.i.d.} \sim \text{Poisson}(\lambda),\ T_n = \frac{1}{n}\sum_{i=1}^n X_i.$$

b) ... und für

$$X_i \text{ i.i.d.} \sim \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\right),\ T_n = \frac{\sum_{i=1}^n X_{i,1}X_{i,2}}{\sqrt{\sum_{i=1}^n X_{i,1}^2 \sum_{i=1}^n X_{i,2}^2}}.$$

(Die asymptotische Varianz von T_n ist $(1-\rho^2)^2$, siehe Skript.)

c) Bestimme ein genähertes 95%-Konfidenzintervall für ρ bei $n = 25$ und $T_n = 0.5$, basierend auf den approximativen Pivots (i) $\frac{\sqrt{n}(T_n - \rho)}{1 - T_n^2}$ und (ii) $\sqrt{n}(g(T_n) - g(\rho))$, wobei g die in b) bestimmte Transformation ist. Diskutiere die Unterschiede.

Abgabe: Montag, 4. Februar