# Musterlösung zur Übung 7

**1. a)** Die Sensitivitätskurve wurde (abgesehen vom Faktor $n+1$) in Aufgabe 1a) bereits berechnet und skizziert. Für eine um $\theta$ symmetrische Verteilung $F$ gilt: $\lim_{n\to\infty} x_{([n\alpha])} = F^{-1}(\alpha)$, $\lim_{n\to\infty} \widehat{\mu_n} = \theta$ und $\lim_{n\to\infty} x_{(n-[n\alpha]+1)} = F^{-1}(1-\alpha)$ (die Limites sind als stochastische Konvergenzen zu verstehen).
Weiter ist der Limes der Steigung $\lim_{n\to\infty} \frac{n+1}{n+1-2[n\alpha]} = \frac{1}{1-2\alpha}$.
Somit ergibt sich:

$$\lim_{n\to\infty} SC_n(x) = \begin{cases} (F^{-1}(\alpha)-\theta)/(1-2\alpha) & \text{für } x \le F^{-1}(\alpha) \\ (x-\theta)/(1-2\alpha) & \text{für } F^{-1}(\alpha) < x \le F^{-1}(1-\alpha) \\ (F^{-1}(1-\alpha)-\theta)/(1-2\alpha) & \text{für } x > F^{-1}(1-\alpha) \end{cases}$$

**b)** Sei $G_\epsilon = (1-\epsilon)F + \epsilon\Delta_x$. Betrachte $x \ne F^{-1}(\alpha)$ und $\epsilon < |\alpha - F(x)|$, d.h. $G_\epsilon(G_\epsilon^{-1}(\alpha)) = \alpha$. Damit $G_\epsilon(Q_\alpha(G_\epsilon)) = \alpha$ und

$$\frac{\partial}{\partial\epsilon} G_\epsilon(Q_\alpha(G_\epsilon))\Big|_{\epsilon=0} = 0.$$

Also

$$\begin{aligned} 0 &= \lim_{\epsilon\to 0} \frac{G_\epsilon(Q_\alpha(G_\epsilon)) - G_0(Q_\alpha(G_0))}{\epsilon} \\ &= \lim_{\epsilon\to 0} \frac{(1-\epsilon)F(Q_\alpha(G_\epsilon)) + \epsilon\Delta_x(G_\epsilon) - F(Q_\alpha(G_0))}{\epsilon} \\ &= \lim_{\epsilon\to 0} \left( \frac{F(Q_\alpha(G_\epsilon)) + \epsilon - F(Q_\alpha(G_0))}{\epsilon} + \Delta_x(G_\epsilon) - F(Q_\alpha(G_\epsilon)) \right) \\ &= f(Q_\alpha(F))\, \frac{\partial}{\partial\epsilon} Q_\alpha(G_\epsilon)\Big|_{\epsilon=0} + \Delta_x(Q_\alpha(F)) - \alpha. \end{aligned}$$

(Letzte Gleichung wg. Kettenregel der Differenziation.) Es gilt $IF(x, Q_\alpha, F) = \frac{\partial}{\partial\epsilon} Q_\alpha(G_\epsilon)\big|_{\epsilon=0}$, also

$$IF(x, Q_\alpha, F) = \frac{\alpha - 1_{\{F^{-1}(\alpha)>x\}}}{f(F^{-1}(\alpha))}.$$

**c)** Wir setzen $G_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_y = F - \varepsilon(\Delta_y - F)$. Wenn $G_\varepsilon^{-1} \to F^{-1}(u)$ für $u = \alpha$ und $u = 1-\alpha$, dann gilt:

$$\begin{aligned} (1-2\alpha)(Q(G_\varepsilon) - Q(F)) &= \int_{G_\varepsilon^{-1}(\alpha)}^{G_\varepsilon^{-1}(1-\alpha)} xF(dx) - \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} xF(dx) \\ &\quad + \varepsilon \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x(\Delta_y(dx) - F(dx)) + o(\varepsilon) \\ &= -\int_{F^{-1}(\alpha)}^{G_\varepsilon^{-1}(\alpha)} xF(dx) + \int_{F^{-1}(1-\alpha)}^{G_\varepsilon^{-1}(1-\alpha)} xF(dx) \\ &\quad + \varepsilon \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x(\Delta_y(dx) - F(dx)) + o(\varepsilon) \end{aligned}$$

Wir wollen zuerst das erste Integral in der letzten Summe umformen: Aus der Definition der Einflussfunktion ergibt sich zunächst

$$\begin{aligned} G_\varepsilon^{-1}(\alpha) - F^{-1}(\alpha) &= Q_\alpha(G_\varepsilon) - Q_\alpha(F) = \varepsilon IF(y, Q_\alpha, F) + o(\varepsilon) \\ &= \frac{\varepsilon}{f(F^{-1}(\alpha))}(\alpha - 1_{\{y<F^{-1}(\alpha)\}}) + o(\varepsilon). \end{aligned}$$

und daraus

$$\begin{aligned} -\int_{F^{-1}(\alpha)}^{G_\varepsilon^{-1}(\alpha)} xf(x)dx &= -F^{-1}(\alpha)f(F^{-1}(\alpha)) \cdot \frac{\varepsilon}{f(F^{-1}(\alpha))}(\alpha - 1_{\{y<F^{-1}(\alpha)\}}) + o(\varepsilon) \\ &= -\varepsilon F^{-1}(\alpha)(\alpha - 1_{\{y<F^{-1}(\alpha)\}}) + o(\varepsilon). \end{aligned}$$

Völlig analog ergibt sich für das zweite Integral

$$\int_{F^{-1}(1-\alpha)}^{G_\varepsilon^{-1}(1-\alpha)} xF(dx) = \varepsilon F^{-1}(1-\alpha)(1_{\{y>F^{-1}(1-\alpha)\}} - \alpha) + o(\varepsilon).$$

Schliesslich lässt sich das dritte Integral wie folgt umformen:

$$\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x(\Delta_y(dx) - F(dx)) = y \cdot 1_{\{y\in[F^{-1}(\alpha),F^{-1}(1-\alpha)]\}} - (1-2\alpha)Q(F).$$

Nun erhalten wir die Einflussfunktion einfach:

$$\begin{aligned} (1-2\alpha) \cdot IF(y, Q, F) &= (1-2\alpha) \cdot \lim_{\varepsilon\to 0} \frac{Q(G_\varepsilon) - Q(F)}{\varepsilon} \\ &= -F^{-1}(\alpha)(\alpha - 1_{\{y<F^{-1}(\alpha)\}}) + F^{-1}(1-\alpha)(1_{\{y>F^{-1}(1-\alpha)\}} - \alpha) \\ &\quad + y \cdot 1_{y\in[F^{-1}(\alpha),F^{-1}(1-\alpha)]} - (1-2\alpha)Q(F) \end{aligned}$$

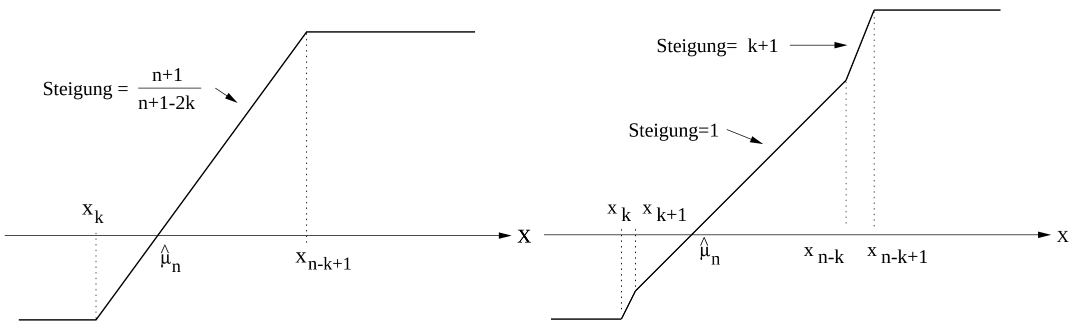
Es ergibt sich dieselbe Funktion wie in Teil a). Das gilt unter bestimmten Regularitätsbedingungen allgemein (siehe jedoch d).

**d)** Sei $k := [n\alpha] = [(n+1)\alpha]$.
Eine ähnliche Rechnung wie im Fall des $\alpha$-gestutztes Mittels ergibt $SC_n(\widehat{\mu_n}) = 0$ und

$$SC_n(x) = \begin{cases} c_1 & \text{für } x \le x_{(k)} \\ (k+1)x + c_2 & \text{für } x_{(k)} < x \le x_{(k+1)} \\ x + c_3 & \text{für } x_{(k+1)} < x \le x_{(n-k)} \\ (k+1)x + c_4 & \text{für } x_{(n-k)} < x \le x_{(n-k+1)} \\ c_5 & \text{für } x > x_{(n-k+1)} \end{cases}$$

wobei $c_1, \dots, c_5$ nur von $x_1, \dots, x_n$ abhängen. Für $n \to \infty$ macht die Sensitivitätskurve des $\alpha$-Winsorisierten Mittels einen Sprung bei $\lim_{n\to\infty} x_{([n\alpha])} = \lim_{n\to\infty} x_{([n\alpha]+1)} = F^{-1}(\alpha)$ und bei $F^{-1}(1-\alpha)$. Dies ist auch das qualitative Verhalten der Einflussfunktion. Will man jedoch die Sprunghöhen der Einflussfunktion durch den Übergang $n \to \infty$ berechnen, kommt man nicht auf die richtigen Resultate (die Regularitätsbedingungen sind nicht erfüllt; gleiches Problem wie beim Median, Beispiel S. 74f. im Skript). Die Einflussfunktion lässt sich mit Funktionalen ähnlich wie in Teil c) korrekt berechnen. Als korrekte Sprunghöhen ergeben sich $\frac{\alpha}{f(F^{-1}(\alpha))}$

(bei $F^{-1}(\alpha)$) und $\frac{\alpha}{f(F^{-1}(1-\alpha))}$ (bei $F^{-1}(1-\alpha)$). Der Unterschied zu a) besteht im Sprung an der Stelle $F^{-1}(\alpha)$. Im $\alpha$-gestutzten Mittel haben also extremere Beobachtungen denselben Einfluss wie $F^{-1}(\alpha)$, während sie beim Winsorisierten Mittel - wo sie nicht gestutzt, sondern genau auf den (asymptotischen) Wert $F^{-1}(\alpha)$ gesetzt werden, einen grösseren Einfluss haben. Das ist intuitiv erstaunlich.




**2.** **a)** $(N_1, N_2, N_3)^T$ ist multinomialverteilt mit Parametern $p_1$, $p_2$ , $p_3$. Betrachte

$$Z_i := \begin{pmatrix} 1_{[X_i=1]} \\ 1_{[X_i=2]} \end{pmatrix} \in \mathbb{N}^2 .$$

Alle $Z_i$ sind unabhängig mit $\mathbf{E}[Z_i] = \binom{p_1}{p_2}$. Für die Kovarianzmatrix $\mathrm{Cov}(Z_i)$ von $Z_i$ brauchen wir

$$\begin{aligned} \mathrm{Var}(1_{[X_i=1]}) &= \mathbf{E}[1_{[X_i=1]}{}^2] - \mathbf{E}[1_{[X_i=1]}]^2 = p_1 - p_1^2 = p_1(1-p_1) \\ \mathrm{Cov}(1_{[X_i=1]}, 1_{[X_i=2]}) &= \mathbf{E}[1_{[X_i=1]}1_{[X_i=2]}] - \mathbf{E}[1_{[X_i=1]}]\,\mathbf{E}[1_{[X_i=2]}] = 0 - p_1p_2 \end{aligned}$$

und erhalten

$$\mathrm{Cov}(Z_i) = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 \\ -p_1p_2 & p_2(1-p_2) \end{bmatrix} .$$

Alle $Z_i$ sind unabhängig, also $\mathrm{Cov}((N_1/n, N_2/n)^T) = \mathrm{Cov}(\frac{1}{n}\sum_{i=1}^n Z_i) = \frac{1}{n}\mathrm{Cov}(Z_i)$. Mit dem zentralen Grenzwertsatz folgt nun

$$\sqrt{n}\left(\begin{pmatrix} N_1/n \\ N_2/n \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}\right) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n Z_i - \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}\right) \xrightarrow{d} Z\,,\ Z \sim \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathrm{Cov}(Z_i)\right),$$

$(n \to \infty)$, oder in anderen Worten

$$\frac{1}{n}\begin{pmatrix} N_1 \\ N_2 \end{pmatrix} \overset{as}{\sim} \mathcal{N}_2\left(\begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \frac{1}{n}V\right) ,$$

wobei $V = \mathrm{Cov}(Z_i)$ die asymptotische Kovarianzmatrix ist.

**b)** Aus dem Gesetz der grossen Zahlen, der Stetigkeit der Wurzelfunktion und der Tatsache, dass Konvergenz in Wahrscheinlichkeit unter fast sicher stetigen Transformationen erhalten bleibt (siehe z. B. Serfling, Theorem in Abschnitt 1.7), folgt

$$T_n^{(1)} = \sqrt{\frac{N_1}{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^n 1_{[X_i=1]}} \xrightarrow{P} \sqrt{\mathbf{E}[1_{[X_i=1]}]} = \sqrt{P[X_i=1]} = \theta \quad (n \to \infty) .$$

Das beweist die Konsistenz von $T_n^{(1)}$. Ausserdem ist nach a) und dem Satz von Cramér-Wold (6.3) auch die erste Komponente von $(N_1/n, N_2/n)^T$ asymptotisch normal, also

$$\frac{N_1}{n} \overset{as}{\sim} \mathcal{N}\big(p_1, \frac{p_1(1-p_1)}{n}\big) = \mathcal{N}\big(\theta^2, \frac{\theta^2(1-\theta^2)}{n}\big)\,.$$

Aus Satz 6.5 folgt mit $h(\cdot) = \sqrt{\cdot}$

$$\begin{aligned} T_n^{(1)} = \sqrt{\frac{N_1}{n}} \quad &\overset{as}{\sim} \quad \mathcal{N}\Big(\sqrt{\theta^2}, \frac{dh}{d\theta}\Big|_{\theta^2} \frac{\theta^2(1-\theta^2)}{n} \frac{dh}{d\theta}\Big|_{\theta^2}\Big) \\ &= \quad \mathcal{N}\Big(\theta, \frac{(\theta^2)^{-\frac{1}{2}}}{2} \frac{\theta^2(1-\theta^2)}{n} \frac{(\theta^2)^{-\frac{1}{2}}}{2}\Big) \\ &= \quad \mathcal{N}\Big(\theta, \frac{1-\theta^2}{4n}\Big)\,. \end{aligned}$$

Als nächstes bestimmen wir den Maximum Likelihood Schätzer:

$$\begin{aligned} P[X_1 = x_1, \ldots, X_n = x_n] &= \prod_{i=1}^n P[X_i = x_i] \\ &= \prod_{i=1}^n (\theta^2)^{1_{[X_i=1]}} (2\theta(1-\theta))^{1_{[X_i=2]}} ((1-\theta)^2)^{1_{[X_i=3]}} \\ &= (\theta^2)^{N_1} (2\theta(1-\theta))^{N_2} ((1-\theta)^2)^{N_3} =: L(\theta) \\ \frac{d}{d\theta} \log L(\theta) &= \frac{2N_1}{\theta} + \frac{N_2}{\theta} - \frac{N_2}{1-\theta} - \frac{2N_3}{1-\theta} \overset{!}{=} 0 \end{aligned}$$

Unter Berücksichtigung von $N_1 + N_2 + N_3 = n$ findet man

$$T_n^{(2)} = \hat{\theta}_{ML} = \frac{N_1}{n} + \frac{N_2}{2n} = \frac{1}{n} \sum_{i=1}^n \big(1_{[X_i=1]} + \frac{1}{2} 1_{[X_i=2]}\big)\,.$$

Konsistenz von $T_n^{(2)}$ folgt aus dem Gesetz der grossen Zahlen

$$T_n^{(2)} \xrightarrow{P} \mathbf{E}\left[1_{[X_i=1]} + \frac{1}{2} 1_{[X_i=2]}\right] = \theta^2 + \frac{1}{2} 2\theta(1-\theta) = \theta \quad (n \to \infty)\,.$$

$T_n^{(2)}$ ist erwartungstreu, und mit Teil a) hat man

$$\begin{aligned} \mathrm{Var}(T_n^{(2)}) &= \frac{1}{n}\Big[\,\mathrm{Var}(1_{[X_i=1]}) + 2\,\mathrm{Cov}(1_{[X_i=1]}, \frac{1}{2} 1_{[X_i=2]}) + \mathrm{Var}\,\big(\frac{1}{2} 1_{[X_i=2]}\big)\Big] \\ &= \frac{1}{n}\Big[\theta^2(1-\theta^2) - 2\theta^2\theta(1-\theta) + \frac{1}{2}\theta(1-\theta)(1-2\theta(1-\theta))\Big] \\ &= \frac{\theta(1-\theta)}{2n}\,. \end{aligned}$$

Schliesslich folgt mit dem zentralen Grenzwertsatz

$$\sqrt{n}(T_n^{(2)} - \theta) \xrightarrow{d} Z\,,\ Z \sim \mathcal{N}\big(0, \frac{\theta(1-\theta)}{2}\big) \quad (n \to \infty)\,,$$

also

$$T_n^{(2)} \overset{as}{\sim} \mathcal{N}\big(\theta, \frac{\theta(1-\theta)}{2n}\big).$$

Der Maximum Likelihood Schätzer ist vorzuziehen, denn für alle $\theta < 1$ gilt

$$\frac{\theta(1-\theta)}{2} < \frac{1-\theta^2}{4} = \frac{(1+\theta)(1-\theta)}{4}.$$

Letzteres ergibt sich aus $\frac{\theta}{2} < \frac{1+\theta}{4}$, und dieses wiederum aus $\frac{\theta}{4} < \frac{1}{4}$ ($\theta < 1$).

**3.** **a)** Die asymptotische Varianz von $T_n$ ist $V(\lambda) = \mathrm{Var}(\sqrt{n}T_n) = \mathrm{Var}(X_i) = \lambda$. Also muss für eine varianzstabilisierende Transformation $h$ gelten $h'(\lambda) = V(\lambda)^{-\frac{1}{2}} = 1/\sqrt{\lambda}$. Die Transformation

$$h(\lambda) = 2\sqrt{\lambda}$$

erfüllt diese Bedingung. Natürlich ist auch $h(\lambda) = \sqrt{\lambda}$ varianzstabilisierend.

**b)** Gemäss Hinweis gilt $V(\rho) = (1-\rho^2)^2$. Aus $h'(\rho) = 1/\sqrt{(1-\rho^2)^2} = 1/(1-\rho^2)$ folgt eine varianzstabilisierende Transformation

$$h(\rho) = \mathrm{artanh}(\rho).$$

**c)** $(i)$ Aus dem ZGS, der Konsistenz von $T_n$ und dem Satz von Slutsky folgt

$$\frac{T_n - \rho}{\sqrt{\frac{1}{n}\widehat{V}(\rho)}} = \frac{\sqrt{n}(T_n - \rho)}{1 - T_n^2} \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Das liefert das approximative 95%-Konfidenzintervall

$$\Big[T_n \pm \Phi^{-1}\big(1 - \frac{\alpha}{2}\big)\frac{1-T_n^2}{\sqrt{n}}\Big] \approx [0.206, 0.794].$$

$(ii)$ Dank Delta-Technik und der Tatsache, dass $\mathrm{artanh}(\cdot)$ varianzstabilisierend ist, folgt

$$\sqrt{n}(\mathrm{artanh}(T_n) - \mathrm{artanh}(\rho)) \xrightarrow{d} Z \sim \mathcal{N}(0,1).$$

Das approximative 95%-Konfidenzintervall ist

$$\tanh\Big[\mathrm{artanh}(T_n) \pm \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{n}}\Big] \approx [0.156, 0.736].$$

Mittels varianzstabilisierender Transformation erhalten wir also ein Konfidenzenzintervall, das nicht mehr symmetrisch um $T_n = 0.5$ ist. Dies macht Sinn, da $\rho$ nur Werte im Intervall $[-1, 1]$ annehmen kann.