Eingangsprüfung Stochastik, 14.10.2016

Wir gehen stets von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ aus.

Aufgabe 1 (18 Punkte)
Sei $B \in \mathcal{A}$ mit $P(B) > 0$. Sei $Q : \mathcal{A} \longrightarrow \mathbb{R}$,

$$Q(A) := \frac{P(A \cap B)}{P(B)}.$$

(a) Für $A \in \mathcal{A}$ sei $1_A : \Omega \longrightarrow \{0, 1\}$ mit

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{sonst} \end{cases}$$

die Indikatorfunktion von A. Zeigen Sie für alle $A, B \in \mathcal{A}$:

(i) $1_A \cdot 1_B = 1_{A \cap B}$

(ii) $\displaystyle\int_\Omega 1_B \cdot 1_A \, dP = P(A \cap B)$

(b) Sei $X \geq 0$ eine Zufallsvariable. Beweisen Sie

$$P(B) \int_\Omega X dQ = \int_\Omega 1_B X \, dP.$$

Aufgabe 2 (18 Punkte) Sei $a > 0$. Seien $X_1, X_2 : \Omega \longrightarrow [0, \infty)$ Zufallsvariablen mit stetiger Verteilungsfunktion F_1, F_2, die $F_1(0) = F_2(0) = 0$ und $F_1(a) = 1$ erfüllen. Betrachte nun eine Zufallsvariable $X : \Omega \longrightarrow [0, \infty)$ mit Verteilungsfunktion F, wobei

$$F(x) = \begin{cases} vF_1(x) & \text{falls } x \leq a \\ v + (1 - v)F_2(x - a) & \text{sonst} \end{cases}$$

mit $v \in (0, 1)$ ist.

(a) Beweisen Sie $E(X) = vE(X_1) + (1 - v)(a + E(X_2))$.
Hinweis: Sie können $E(X) = \displaystyle\int_0^\infty 1 - F(x) \, dx$ *ohne Beweis verwenden.*

(b) Zeigen Sie, dass $P(X - a \leq x | X > a) = F_2(x)$ für alle $x > 0$ gilt.

(c) Bestimmen Sie $P(X \leq x | X \leq a)$ für $x \in (0, a)$.

Aufgabe 3 (14 Punkte) Gegeben seien zwei Zufallsvariablen Y und N. Dabei sei $Y \sim \Gamma(\alpha, \beta)$ und die bedingte Verteilung von N gegeben $Y = y$ sei eine Poisson-Verteilung $\mathrm{Poi}(y)$.

(a) Bestimmen Sie $E(N)$ und $\mathrm{Var}(N)$.

(b) Bestimmen Sie die Wahrscheinlichkeit $P(N = 0)$.

Hinweis: $\int\limits_0^\infty t^{x-1} e^{-at}\, dt = \frac{\Gamma(x)}{a^x}$.

Aufgabe 4 (18 Punkte)
Die Dichte $f : \mathbb{R} \longrightarrow [0, \infty)$ einer Zufallsvariablen X sei gegeben durch

$$f(x) = \frac{1}{\vartheta} e^{-\frac{x}{\vartheta}} 1_{(0,\infty)}(x)$$

mit $\vartheta > 0$.

(a) Begründen Sie $E(X) = \vartheta$.

(b) Bestimmen Sie einen Maximum-Likelihood Schätzer $\hat{\vartheta}$ für ϑ.

(c) Seien im Folgenden $X_1, \ldots, X_n$ unabhängige, identisch wie X verteilte Zufallsvariablen.

(i) Betrachten Sie in der Folge den Schätzer $\hat{\vartheta}_1 = \frac{1}{n} \sum\limits_{i=1}^n X_i$. Ist dieser Schätzer erwartungstreu?

(ii) Bestimmen Sie die Verteilung von $\hat{\vartheta}_1$ aus (i).

Aufgabe 5 (22 Punkte)
In der folgenden Tabelle sind die Handspanne (gemessen von Spitze des Daumens bis zur Spitze des kleinen Fingers) und die Körpergröße in cm von zufällig ausgewählten Personen aufgezeichnet.

Nr.	Handspanne	Körpergröße	Nr.	Handspanne	Körpergröße
1	22,0	187,96	7	18,0	175,00
2	16,7	157,48	8	19,2	176,50
3	17,5	150,00	9	16,0	159,00
4	22,5	177,00	10	18,5	167,00
5	21,0	177,00	11	23,2	180,00
6	19,1	177,00			

Eine graphische Darstellung finden Sie in Abbildung 1.
Bezeichnet man mit x_i die Handspanne und y_i Körpergröße, so wird für die Zufallsvariable Y_i das Modell

$$Y_i = a + bx_i + \varepsilon_i,\ i = 1, \ldots, 11$$

untersucht. Die ε_i seien unabhängig und $N(0, \sigma^2)$-verteilt.

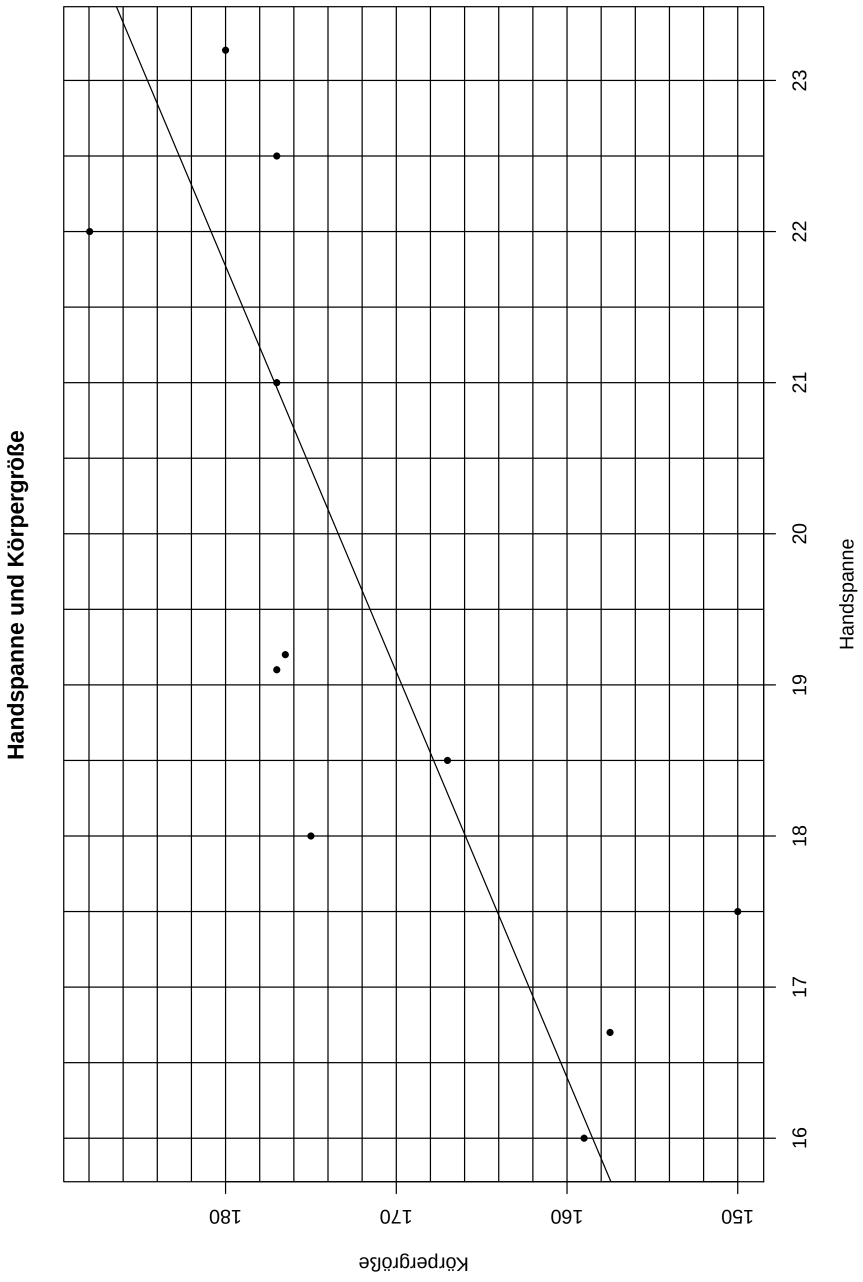

Abbildung 1: Beobachtete Handspannen, Körpergrößen und Ausgleichsgerade. Beobachtungen sind mit Punkten • eingetragen, senkrechte bzw. waagerechte Linien sind in Intervallen von 0,5 bzw. 2 eingezeichnet.

(a) Bestimmen Sie einen approximativen Schätzwert $\hat{b}$ von b aus Abbildung 1.

(b) Untersuchen Sie die Hypothese, dass $b = 0$ gilt und geben Sie eine Interpretation Ihres Ergebnisses. Die Aussage soll zu einem Niveau von $\alpha = 10$ % erfolgen. Sie können verwenden:

$$\begin{aligned}
\overline{x} &= 19,4 \\
\sum_{i=1}^{11}(x_i - \overline{x})^2 &= 58,7 \\
\sum_{i=1}^{11}(y_i - \hat{y}_i)^2 &= 491,2 \qquad (\text{mit } \hat{y}_i = \hat{a} + \hat{b}x_i)
\end{aligned}$$

(c) Bestimmen Sie mit Hilfe der Abbildung einen Schätzwert für den Erwartungswert der Körpergröße bei einer Handspanne von 20 cm und ein 95 % Schätzintervall für diesen Erwartungswert.

(d) Eine Statistik-Software gibt das 95 % Prognose-Intervall [155,9;190,9] aus. Vergleichen Sie dies mit dem Ergebnis aus (c) und kommentieren Sie Ihre Feststellung.

Lösungsvorschläge

Aufgabe 1 [(a) 4+2 Punkte, (b) 12]
Lösung
Zu (a)
(i) Es gilt

$$
\begin{aligned}
1_A(\omega)\cdot 1_B(\omega)=1 &\iff 1_A(\omega)=1 \text{ und } 1_B(\omega)=1 \iff \omega\in A \text{ und } \omega\in B\\
&\iff \omega\in A\cap B \iff 1_{A\cap B}(\omega)=1.
\end{aligned}
$$

(ii) Mit (i) gilt gemäß der Definition des Integrals

$$
\int_\Omega 1_B\cdot 1_A dP \quad = \quad \int_\Omega 1_{B\cap A} dP = P(A\cap B).
$$

Zu (b)
Sei $X=\sum_{i=1}^n \alpha_i 1_{A_i}$, $\alpha_i\geq 0$, $A_i\in\mathcal{A}$, $i=1,\dots,n$ eine Treppenfunktion.
Dann gilt laut Konstruktion des Integrals

$$
\begin{aligned}
\int_\Omega X\,dQ &= \sum_{i=1}^n \alpha_i Q(A_i) = \sum_{i=1}^n \alpha_i \,\frac{P(A_i\cap B)}{P(B)} \text{ also}\\
P(B)\int_\Omega X\,dQ &= \sum_{i=1}^n \alpha_i P(A_i\cap B).
\end{aligned}
$$

Andererseits gilt wegen (a)(ii)

$$
\int_\Omega 1_B X dP \quad = \quad \int_\Omega 1_B\sum_{i=1}^n \alpha_i 1_{A_i} dP = \sum_{i=1}^n \alpha_i \int_\Omega 1_B 1_{A_i} dP \overset{(a)}{=} \sum_{i=1}^n \alpha_i P(A_i\cap B)
$$

und somit die Behauptung für Treppenfunktionen.
Sei nun $X\geq 0$. Es gibt eine Folge $(X_n)_{n\in\mathbb{N}}$ von Treppenfunktionen mit $X_n\geq 0$ und $X_n\nearrow X$.
Da X_n eine Treppenfunktion ist, gilt wie oben gezeigt

$$
P(B)\int_\Omega X_n\,dQ = \int_\Omega 1_B X_n\,dP.
$$

Aus $X_n\nearrow X$ folgt $X_n 1_B\nearrow X 1_B$ und mit dem Satz von der monotonen Konvergenz folgt für die beiden Seiten

$$
\begin{aligned}
\lim_{n\to\infty}\int_\Omega X_n\,dQ &= \int_\Omega X\,dQ,\\
\lim_{n\to\infty}\int_\Omega 1_B X_n\,dP &= \int_\Omega 1_B X\,dP
\end{aligned}
$$

und somit insgesamt:

$$
P(B)\int_\Omega X\,dQ = \int_\Omega 1_B X\,dP.
$$

Aufgabe 2 [(a) 8, (b) 5, (c) 5]
Zu (a)
Es gilt mit dem Hinweis

$$\begin{aligned}
E(X) &= \int_0^a 1 - F(x)\,dx + \int_a^\infty 1 - F(x)\,dx \\
&= \int_0^a 1 - vF_1(x)\,dx + \int_a^\infty 1 - (v + (1-v)F_2(x-a))\,dx \\
&= \int_0^a (1-v) + v - vF_1(x)\,dx + \int_a^\infty 1 - v - (1-v)F_2(x-a)\,dx \\
&= a(1-v) + v\int_0^a 1 - F_1(x)\,dx + (1-v)\int_a^\infty 1 - F_2(x-a)\,dx \\
&= a(1-v) + vE(X_1) + (1-v)\int_0^\infty 1 - F_2(y)\,dy \\
&= a(1-v) + vE(X_1) + (1-v)E(X_2).
\end{aligned}$$

In der vorletzten Zeile wird für den zweiten Summanden verwendet, dass

$$\begin{aligned}
&F_1(a) = 1 \Longrightarrow \forall x \geq a: \; F_1(x) = 1 \\
&E(X_1) = \int_0^\infty 1 - F_1(x)\,dx = \int_0^a 1 - F_1(x)\,dx
\end{aligned}$$

gilt und beim dritten Summanden wird die Substitution $y = x - a$ verwendet.
Zu (b)
Sei $x > 0$.

$$\begin{aligned}
P(X - a \leq x | X > a) &= \frac{P(X - a \leq x, X > a)}{P(X > a)} = \frac{P(X \leq x + a, X > a)}{1 - P(X \leq a)} \\
&= \frac{P(a < X \leq x + a)}{1 - v} = \frac{P(X \leq a + x) - P(X \leq a)}{1 - v} \\
&= \frac{v + (1-v)F_2(x + a - a) - v}{1 - v} = \frac{(1-v)F_2(x)}{1 - v} = F_2(x).
\end{aligned}$$

Zu (c)
Sei $x \in (0, a)$.

$$\begin{aligned}
P(X \leq x | X \leq a) &= \frac{P(X \leq x, X \leq a)}{P(X \leq a)} = \frac{P(X \leq x)}{v} = \frac{vF_1(x)}{v} \\
&= F_1(x)
\end{aligned}$$

Aufgabe 3 [(a) 6, (b) 8]
Zu (a)
Nach Voraussetzung gilt

$$E(N|Y) = Y, \; \mathrm{Var}(N|Y) = Y.$$

Mit dem iterierten Erwartungswert gilt

$$E(N) = E(E(N|Y)) = E(Y) = \frac{\alpha}{\beta}$$
$$\mathrm{Var}(N) = \mathrm{Var}(E(N|Y)) + E(\mathrm{Var}(N|Y)) = \mathrm{Var}(Y) + E(Y) = \frac{\alpha}{\beta} + \frac{\alpha}{\beta^2}.$$

Zu (b)
Sei $f_Y(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$, $y > 0$ die Dichte von Y nach Formelsammlung. Es gilt

$$\begin{aligned} P(N=0) &= \int_0^\infty P(N=0|Y=y) f_Y(y)\,dy = \int_0^\infty e^{-y} \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}\,dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-(\beta+1)y}\,dy = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{(1+\beta)^\alpha} = \frac{\beta^\alpha}{(1+\beta)^\alpha}. \end{aligned}$$

Aufgabe 4[(a) 3 (b) 9 (c) 3 + 3]
Zu (a)
Es gilt $X \sim \mathrm{Exp}\left(\frac{1}{\vartheta}\right)$ und somit $E(X) = \left(\frac{1}{\vartheta}\right)^{-1} = \vartheta$.
Zu (b)
Seien $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Exp}\left(\frac{1}{\vartheta}\right)$. Es ergibt sich für $(x_1, \ldots, x_n) \in (0,\infty)^n$

$$L(\vartheta) = \prod_{i=1}^n \frac{1}{\vartheta} \exp(-\frac{x_i}{\vartheta}) \qquad \text{(Likelihood)}$$
$$\ell(\vartheta) := \ln L(\vartheta) = \sum_{i=1}^n (-\ln\vartheta - \frac{1}{\vartheta} x_i)$$
$$\ell'(\vartheta) = \sum_{i=1}^n \left(-\frac{1}{\vartheta} + \frac{1}{\vartheta^2} x_i\right) = -\frac{n}{\vartheta} + \frac{1}{\vartheta^2} \sum_{i=1}^n x_i$$
$$\ell''(\vartheta) = \frac{n}{\vartheta^2} - \frac{2}{\vartheta^3} \sum_{i=1}^n x_i$$
$$\ell'(\hat\vartheta) = 0 \Longrightarrow \frac{n}{\hat\vartheta} = \frac{1}{\hat\vartheta^2} \sum_{i=1}^n x_i \Longrightarrow \hat\vartheta = \frac{1}{n} \sum_{i=1}^n x_i =: \overline{x}$$
$$\ell''(\hat\vartheta) = \frac{n}{\overline{x}^2} - \frac{2}{\overline{x}^3} \sum_{i=1}^n x_i = \frac{n}{\overline{x}^2} - \frac{2n}{\overline{x}^2} = -\frac{n}{\overline{x}^2} < 0.$$

Somit ist $\hat\vartheta = \frac{1}{n} \sum_{i=1}^n X_i$ ML-Schätzer.
Zu (c) (i)
Wegen

$$E(\hat\vartheta_1) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X)$$

ist $\hat{\vartheta}_1$ wegen (a) erwartungstreu.

(ii) Wegen $X_i \sim \operatorname{Exp}\left(\frac{1}{\vartheta}\right) = \Gamma\left(1, \frac{1}{\vartheta}\right)$ gilt $\sum_{i=1}^{n} X_i \sim \Gamma\left(n, \frac{1}{\vartheta}\right)$ und $\frac{1}{n}\sum_{i=1}^{n} X_i \sim \Gamma\left(n, \frac{n}{\vartheta}\right)$.

Aufgabe 5 [(a) 6, (b) 6, (c) 6, (d) 4]

Zu (a)

Aus der Graphik werden die Punkte (17,5/164) und (21/177) auf der Ausgleichsgeraden abgelesen. Die Steigung der Ausgleichsgeraden ist damit

$$\hat{b} = \frac{177 - 164}{21 - 17,5} = 3,71.$$

Zu (b)

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{9} \cdot 491,2 = 54,57 \\ \operatorname{se}(\hat{b}) &= \sqrt{\frac{54,57}{58,7}} = 0,964 \\ T &= \frac{\hat{b}}{\operatorname{se}(\hat{b})} = \frac{3,71}{0,964} = 3,85. \end{aligned}$$

H_0 wird abgelehnt, denn $|T| > t_{9;0,95} = 1,833$. Somit wird die Hypothese, dass die Körpergröße nicht von der Handspanne abhängt, verworfen.

Zu (c)

Man liest auf der Ausgleichsgeraden bei der Handspanne von 20 cm den Wert von ca. 173,5 ab.

Mit $t_{9;0,975} = 2,262$,

$$\operatorname{se}(\hat{a} + \hat{b}x) = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(x - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} = 7,34\sqrt{\frac{1}{11} + \frac{(20 - 19,4)^2}{58,7}} = 2,3$$

ergibt sich (gerundet) $[168,3; 178,7]$.

Zu (d)

Das Intervall in (d) ist deutlich breiter als in (c). Dort ist es ein Schätzintervall für den Erwartungswert. Das Prognoseintervall bezieht sich auf die Realisierungen der Zufallsvariablen mit $x = 20$, deren Schwankung deutlich höher ist.