INSTITUT FÜR STOCHASTIK
UNIVERSITÄT KARLSRUHE
Priv.-Doz. Dr. D. Kadelka

SS 2008

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 9.9.2008

Musterlösungen

Aufgabe A1

Gegeben sei eine Urliste mit den Paaren $(x_1, y_1), \ldots, (x_{12}, y_{12})$

j	1	2	3	4	5	6	7	8	9	10	11	12
x_j	-2.1	-1	0.1	1.1	1.6	2.9	4.2	5	5.7	7	8.3	9
y_j	7.1	3.3	6.4	5.9	4.1	1.2	1.3	2.7	0	-3.3	-5.7	-7.1

a) Berechnen Sie die Stichprobenmittel $\bar{x}$, $\bar{y}$, die Stichproben-Standardabweichungen s_x, s_y und den empirischen Korrelationskoeffizienten r_{xy}.
Lösung: Direkt aus den Daten ergibt sich gemäß Definition 1.8 und Paragraph 1.5 unter Ausnützung der Beziehung

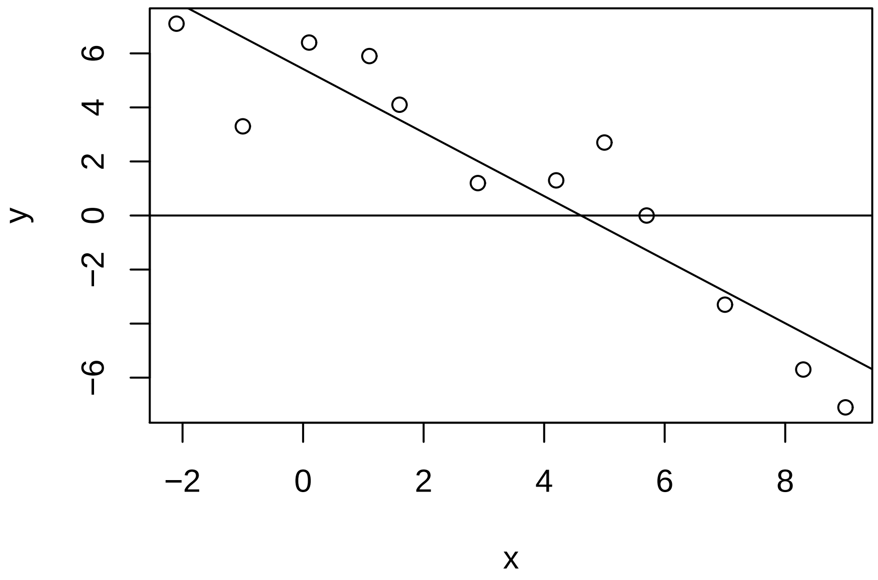
$$\sum_{j=1}^{n}(x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 3.48 \qquad s_x = 3.643$$
$$\bar{y} = 1.32 \qquad s_y = 4.642$$
$$r_{xy} = -0.9233$$

b) Bestimmen Sie die zugehörige Regressionsgerade $y = a^* + b^* \cdot x$ von y auf x.
Lösung: Nach Paragraph 1.5 ist $b^* = r_{xy} \cdot \dfrac{s_y}{s_x}$ und $a^* = \bar{y} - b^* \cdot \bar{x}$, also

$$\begin{aligned} b^* &= -1.176 \\ a^* &= 5.42 \end{aligned}$$

und die Regressionsgerade $y = 5.42 - 1.176 \cdot x$.



Punkte und Regressionsgerade $y = a^* + b^* \cdot x$

Für die Lösung der letzten beiden Aufgabenteile benötigen wir die aufsteigend sortierten y-Werte. Es ist

$$y_{()} = (-7.1,\ -5.7,\ -3.3,\ 0,\ 1.2,\ 1.3,\ 2.7,\ 3.3,\ 4.1,\ 5.9,\ 6.4,\ 7.1)$$

c) Berechnen Sie das 0.1-getrimmte Stichprobenmittel $\bar{y}_{0.1}$ von $(y_1, \ldots, y_{12})$.
Lösung: Mit $k = [12 \cdot 0.1] = 1$ ergibt sich

$$\bar{y}_{0.1} = \frac{1}{12 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(11)}) = 1.59$$

d) Berechnen Sie den Quartilsabstand von $(y_1, \ldots, y_{12})$.
Lösung: Da $0.25 \cdot 12 = 3$ und $0.75 \cdot 12 = 9$ beide ganzzahlig sind, ergibt sich mit $k_1 = 3$ und $k_2 = 9$

$$\begin{aligned}
\tilde{y}_{0.25} &= \frac{y_{(k_1)} + y_{(k_1+1)}}{2} = \frac{y_{(3)} + y_{(4)}}{2} = -1.65 \\
\tilde{y}_{0.75} &= \frac{y_{(k_2)} + y_{(k_2+1)}}{2} = \frac{y_{(9)} + y_{(10)}}{2} = 5
\end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 6.65$.

Aufgabe A2

Die Zufallsvariable X besitze die Verteilung $\mathcal{N}(15, 25)$ und die Zufallsvariable Y die Verteilung $\mathcal{N}(-12, 144)$. X und Y seien stochastisch unabhängig.

a) Bestimmen Sie $\mathbb{P}(X \geq 22)$.

b) Bestimmen Sie die Standardabweichung σ_{X+Y} von $X + Y$.

c) Welche Verteilung besitzt $Z := X/5 - Y/4$?

d) Berechnen Sie die Kovarianz $C(Z, X)$ von Z und X. Sind X und Z positiv korreliert, negativ korreliert oder unkorreliert?

e) Für welches $a > 0$ gilt $\mathbb{P}(|X - 15| \leq a) = 0.9545$?

Lösung:

a) Wegen der Stetigkeit von $\mathcal{N}(15, 25)$ gilt nach (9.6) und Tabelle A.1

$$\mathbb{P}(X \geq 22) = 1 - \mathbb{P}(X < 22) = 1 - \mathbb{P}(X \leq 22) = 1 - \Phi\left(\frac{22 - 15}{5}\right)$$
$$= 1 - \Phi(1.4) = 1 - 0.9192 = 0.0808$$

b) Wegen Satz 12.23 f) gilt $V(X + Y) = V(X) + V(Y) = 25 + 144 = 169$, da X und Y stochastisch unabhängig sind, und damit

$$\sigma_{X+Y} = \sqrt{V(X + Y)} = \sqrt{169} = 13.$$

c) Wegen Satz 9.7 a) gilt $X/5 = \frac{1}{5} \cdot X \sim \mathcal{N}(\frac{1}{5} \cdot 15, \frac{1}{5^2} \cdot 25) = \mathcal{N}(3, 1)$ und $-Y/4 = \frac{-1}{4} \cdot Y \sim \mathcal{N}(\frac{-1}{4} \cdot (-12), \frac{1}{4^2} \cdot 144) = \mathcal{N}(3, 9)$, also wegen der Unabhängigkeit von $X/5$ und $-Y/4$ und der Faltungsformel 11.16

$$X/5 - Y/4 \sim \mathcal{N}(3 + 3, 1 + 9) = \mathcal{N}(6, 10).$$

d) Wegen Satz 12.23 und der Unabhängigkeit von $X/5$ und $-Y/4$ gilt

$$C(Z, X) = C(X/5 - Y/4, X) = C(X/5, X) - C(Y/4, X) = \frac{1}{5}C(X, X) - 0 = \frac{1}{5}V(X) = 5.$$

Hieraus erhält man ohne weitere Rechnung, dass $\rho(Z, X) > 0$ ist. Daher sind X und Z positiv korreliert.

e) Wir wenden die $k \cdot \sigma$-Regel aus (9.7) an mit $k = 2$. Danach gilt

$$\mathbb{P}(|X - \mu| \leq 2 \cdot \sigma) = 2 \cdot \Phi(2) - 1 = 0.9545,$$

falls $X \sim \mathcal{N}(\mu, \sigma^2)$, hier also mit $\mu = 15$, $\sigma = 5$ und $t = 2$ wegen (9.7)

$$\mathbb{P}(|X - 15| \leq 10) = 0.9545.$$

Damit gilt $a = 10$.

Aufgabe A3

Aus einer Sendung von 30 Maschinenteilen, unter denen sich 9 defekte und 21 intakte befinden, wählt man zur Kontrolle 4 Stück zufällig nacheinander und ohne Zurücklegen aus.

a) Wie groß ist die Wahrscheinlichkeit p_1, dass das erste herausgegriffene Maschinenteil intakt ist?

b) Welche Verteilung besitzt X, die zufällige Anzahl der herausgegriffenen intakten Maschinenteile und welche Verteilung besitzt Y, die zufällige Anzahl der herausgegriffenen defekten Maschinenteile?

c) Man berechne die Wahrscheinlichkeit p_2, dass alle 4 entnommenen Teile intakt sind und die Wahrscheinlichkeit p_3, dass unter den 4 entnommenen Teilen mindestens ein defektes ist.

d) Man bestimme die Wahrscheinlichkeit p_4, dass das erste entnommene Teil intakt ist, aber dass nicht alle 4 Teile intakt sind.

e) Man berechne die bedingte Wahrscheinlichkeit q, dass das erste herausgegriffene Maschinenteil intakt ist unter der Bedingung, dass mindestens ein defektes Teil ausgewählt wurde.

Lösung:

a) Sei A das Ereignis, dass das erste herausgegriffene Maschinenteil intakt ist. Mit $r := 9$, die Anzahl der defekten Maschinenteile, und $s := 21$, die Anzahl der intakten Maschinenteile, ist

$$\mathbb{P}(A) = \frac{s}{r+s} = 0.7 \ .$$

b) Da die herausgegriffenen Teile nicht zurückgelegt werden, gilt mit $n = 4$

$$\begin{aligned} X &\sim Hyp(n,s,r) &= Hyp(4,21,9) \\ Y &\sim Hyp(n,r,s) &= Hyp(4,9,21). \end{aligned}$$

c) Gesucht ist

$$\mathbb{P}(\text{„Alle 4 entnommenen Teile intakt“}) = \mathbb{P}(X=4) = \frac{\binom{s}{4}\binom{r}{0}}{\binom{r+s}{4}} = \frac{\binom{21}{4}}{\binom{30}{4}}$$

$$= \frac{21\cdot 20\cdot 19\cdot 18}{4!} \Big/ \frac{30\cdot 29\cdot 28\cdot 27}{4!} = \frac{21\cdot 20\cdot 19\cdot 18}{30\cdot 29\cdot 28\cdot 27} = 0.2184 \ .$$

und

$$\mathbb{P}(\text{„Mindestens 1 defektes Teil“}) = 1 - \mathbb{P}(\text{„Alle 4 entnommenen Teile intakt“}) = 0.7816 \ .$$

d) Sei C das Ereignis „Alle 4 entnommenen Teile intakt“. Wegen c) gilt $\mathbb{P}(C) = 0.2184$. Gesucht ist

$$\mathbb{P}(A\setminus C) = \mathbb{P}(A) - \mathbb{P}(C) = \frac{21}{30} - \frac{21\cdot 20\cdot 19\cdot 18}{30\cdot 29\cdot 28\cdot 27} = \frac{419}{870} = 0.4816 \ ,$$

da das Ereignis C in A enthalten ist.

e) Gesucht ist

$$\mathbb{P}(A \mid C^c) \;=\; \frac{\mathbb{P}(A \cap C^c)}{\mathbb{P}(C^c)} \;=\; \frac{\mathbb{P}(A \setminus C)}{1 - \mathbb{P}(C)} \;=\; \frac{0.4816}{0.7816} \;=\; 0.6162\ .$$

Aufgabe A4

Eine Zufallsvariable Y habe die Verteilungsfunktion

$$t \to F_Y(t) := \begin{cases} 0 & , t \le 1, \\ 1 - \dfrac{1}{t^c} & , t > 1, \end{cases}$$

wobei $c > 0$ ein fester Parameter ist.

a) Berechnen Sie das q-Quantil von Y für $0 < q < 1$.

b) Bestimmen Sie die Dichte $y \to f_Y(y)$ von Y für $y > 1$.

c) Es sei $c > 2$. Berechnen Sie den Erwartungswert $\mathbb{E}Y$, das zweite Moment $\mathbb{E}Y^2$ und die Varianz $V(Y)$.
Hinweis: $\displaystyle\int_1^\infty x^{-k} dx = \frac{1}{k-1}$ für $k > 1$.

d) Es sei $Z := \ln(Y)$. Berechnen Sie $\mathbb{P}(Z \le t)$ für $t > 0$.

e) Die Zufallsvariable Z besitzt eine Exponentialverteilung $Exp(\alpha)$ mit einem gewissen Parameter $\alpha > 0$. Bestimmen Sie α.

Lösung:

a) Nach Definition 12.19 gilt, dass das q-Quantil t_q die Lösung ist von

$$F_Y(t_q) = 1 - \frac{1}{t_q^c} = q$$

und damit

$$\begin{aligned} 1 - q &= \frac{1}{t_q^c} \\ t_q^c &= \frac{1}{1-q} \\ t_q &= \sqrt[c]{\frac{1}{1-q}}\ . \end{aligned}$$

b) Die Verteilungsfunktion F_Y von Y ist stetig und bis auf die Stelle $y = 1$ auch stetig differenzierbar. Wegen Satz 8.12 folgt, dass

$$f_Y(y) = F_Y'(y) = \frac{c}{y^{c+1}}, \quad y > 1,$$

die gesuchte Dichte ist.

c) Unter Ausnützung des Hinweises erhält man

$$\begin{aligned}
\mathbb{E}Y &= \int_{-\infty}^{\infty} y\ f_Y(y)\ dy = \int_1^{\infty} \frac{c}{y^c}\ dy = \frac{c}{c-1} \\
\mathbb{E}Y^2 &= \int_{-\infty}^{\infty} y^2\ f_Y(y)\ dy = \int_1^{\infty} \frac{c}{y^{c-1}}\ dy = \frac{c}{c-2} \\
V(Y) &= \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = \frac{c}{c-2} - \left(\frac{c}{c-1}\right)^2 = \frac{c}{(c-2)\cdot(c-1)^2}
\end{aligned}$$

d) Wegen $e^t > 1$ für $t > 0$ gilt

$$\mathbb{P}(Z \le t) = \mathbb{P}(\ln(Y) \le t) = \mathbb{P}(Y \le e^t) = 1 - \frac{1}{(e^t)^c} = 1 - \frac{1}{e^{c\cdot t}} = 1 - e^{-c\cdot t}.$$

e) Die Exponentialverteilung $Exp(\alpha)$ besitzt die Verteilungsfunktion $F(t) = 1 - e^{-\alpha\cdot t}, t > 0$. Durch Vergleich mit der Verteilungsfunktion von Z in d) folgt direkt $\alpha = c$.

Aufgabe A5

Ein Merkmal besitze die Dichte

$$f_\vartheta(x) = \begin{cases} \frac{2}{\sqrt{\pi}}\ \vartheta^{3/2}\ x^{1/2}\ e^{-\vartheta x}, & \text{falls } x > 0 \\ 0, & \text{falls } x \le 0 \end{cases}$$

mit unbekanntem Parameter $\vartheta > 0$.

a) Der Maximum-Likelihood-Schätzer für ϑ^* zur Stichprobe $x = (x_1, \dots, x_n)$, wobei $x_i > 0$ für $i = 1, \dots, n$ gilt, ist von der Form

$$T(x_1, \dots, x_n) = c \cdot \frac{1}{\sum_{i=1}^n x_i}$$

mit einer geeigneten Konstanten c. Bestimmen Sie c.

b) $x \to f_\vartheta(x)$ ist die Dichte einer speziellen $\Gamma(\alpha, \beta)$-Verteilung. Wie groß ist α und wie groß ist β ?

c) $X_1, \dots, X_n$ seien die unabhängigen Stichprobenvariablen mit der Dichte f_ϑ. Welche Verteilung hat $Y := \sum_{i=1}^n X_i$?

d) Bestimmen Sie $\mathbb{E}_\vartheta T(X_1, \dots, X_n)$.
<u>Hinweis:</u> Hat eine Zufallsvariable Z die Verteilung $\Gamma(\alpha, \beta)$ mit $\alpha > 1$, so gilt bekanntlich $\mathbb{E}Z^{-1} = \frac{\beta}{\alpha-1}$.

Lösung:

a) Es gilt

$$\ln(f_\vartheta(x)) = \ln(2/\sqrt{\pi}) + \frac{3}{2}\ln(\vartheta) + \frac{1}{2}\ln(x) - \vartheta \cdot x, \quad x > 0.$$

Damit ergibt sich die Loglikelihood-Funktion zu

$$
\begin{aligned}
M_x(\vartheta) = \sum_{i=1}^n \ln(f_\vartheta(x_i)) &= \sum_{i=1}^n \left(\ln(2/\sqrt{\pi}) + \frac{3}{2}\ln(\vartheta) + \frac{1}{2}\ln(x_i) - \vartheta \cdot x_i\right) \\
&= n \cdot \ln(2/\sqrt{\pi}) + n \cdot \frac{3}{2} \cdot \ln(\vartheta) + \frac{1}{2}\sum_{i=1}^n \ln(x_i) - \sum_{i=1}^n \vartheta \cdot x_i \\
&= n \cdot \ln(2/\sqrt{\pi}) + \frac{1}{2}\sum_{i=1}^n \ln(x_i) + \frac{3n}{2} \cdot \ln(\vartheta) - n \cdot \bar{x} \cdot \vartheta,
\end{aligned}
$$

wobei wir $n \cdot \bar{x} = \sum_{i=1}^n x_i$ ausgenützt haben. Ableiten nach ϑ ergibt

$$M_x'(\vartheta) = \frac{3n}{2 \cdot \vartheta} - n \cdot \bar{x}$$

und

$$M_x''(\vartheta) = -\frac{3n}{2 \cdot \vartheta^2} < 0.$$

Wegen $M_x''(\vartheta) < 0$ für alle $\vartheta > 0$ erhalten wir den Maximum-Likelihood-Schätzwert $\hat{\vartheta}(x)$ als Nullstelle ϑ_0 der Gleichung $M_x'(\vartheta_0) = 0$, also $\dfrac{3n}{2 \cdot \vartheta_0} = n \cdot \bar{x}$. Aufgelöst nach ϑ_0 ergibt sich

$$\hat{\vartheta}(x) = \vartheta_0 = \frac{3}{2\bar{x}} = \frac{3n}{2 \cdot \sum_{i=1}^n x_i} = T(x_1, \ldots, x_n).$$

Damit gilt $c = \frac{3n}{2}$.

b) Nach Definition 9.3 besitzt allgemein $\Gamma(\alpha, \beta)$ die Dichte

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta x}, \quad x > 0.$$

Der Vergleich mit der Dichte f_ϑ liefert $\alpha - 1 = 1/2$, also $\alpha = \frac{3}{2}$ und $\beta = \vartheta$. (Tatsächlich stimmen wegen $\Gamma(3/2) = 1/2 \cdot \Gamma(1/2) = \sqrt{\pi}/2$ (vergl. Skript 9.2) dann f_ϑ und f überein.)

c) Wegen $X_i \sim \Gamma(\frac{3}{2}, \vartheta)$ und der Unabhängigkeit der X_i gilt nach der Faltungsformel

$$Y = \sum_{i=1}^n X_i \sim \Gamma\left(\underbrace{\frac{3}{2} + \ldots + \frac{3}{2}}_{n\text{-mal}}, \vartheta\right) = \Gamma\left(\frac{3n}{2}, \vartheta\right).$$

d) Es ist $T(X_1, \ldots, X_n) = \frac{3n}{2Y} = \frac{3n}{2} \cdot Y^{-1}$ und damit

$$\mathbb{E}_\vartheta T(X_1, \ldots, X_n) = \mathbb{E}_\vartheta \frac{3n}{2} \cdot Y^{-1} = \frac{3n}{2} \cdot \mathbb{E}_\vartheta Y^{-1} \overset{\text{c) und Hinweis}}{=} \frac{3n}{2} \frac{\vartheta}{\frac{3n}{2} - 1} = \vartheta \cdot \frac{1}{1 - \frac{2}{3n}}.$$