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

SS 2009

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 14.9.2009

Musterlösungen

Aufgabe 1:

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	−0.8	0.1	0.9	2.1	3	4	4.8	6.1	7.2	8.1	9.2	10
y_j	8.6	6.5	7.3	7.8	5.9	2.3	−0.4	−0.8	−8.6	−0.3	−6.3	−8.1

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

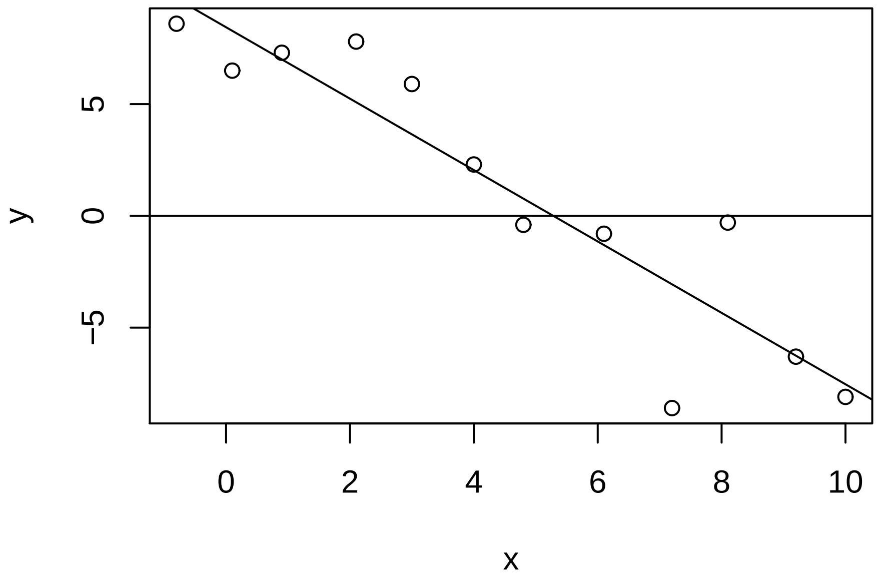
$$\sum_{j=1}^{n}(x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 4.56 \qquad s_x = 3.616$$
$$\bar{y} = 1.16 \qquad s_y = 6.285$$
$$r_{xy} = -0.9193$$

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

$$\begin{aligned} b^* &= -1.598 \\ a^* &= 8.44 \end{aligned}$$

und die Regressionsgerade $y = 8.44 - 1.598 \cdot x$.



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

Für die Lösung der nächsten drei Aufgabenteile benötigen wir die aufsteigend sortierten y-Werte. Es ist

$$y_{()} = (-8.6,\ -8.1,\ -6.3,\ -0.8,\ -0.4,\ -0.3,\ 2.3,\ 5.9,\ 6.5,\ 7.3,\ 7.8,\ 8.6)$$

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

$$\bar{y}_{0.1} = \frac{1}{12 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(11)}) = 1.39$$

d) Bestimmen Sie das Stichproben-0.2-Quantil $\tilde{y}_{0.2}$ von $(y_1, \ldots, y_{12})$.
Lösung: Da $12 \cdot 0.2 = 2.4$ nicht ganzzahlig ist, ist mit $k = [2.4] = 2$

$$\tilde{y}_{0.2} = y_{(k+1)} = y_{(3)} = -6.3$$

e) 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 $k_2 = 9$

$$\begin{aligned} \tilde{y}_{0.25} &= \frac{y_{(k_1)} + y_{(k_1+1)}}{2} = \frac{y_{(3)} + y_{(4)}}{2} = -3.55 \\ \tilde{y}_{0.75} &= \frac{y_{(k_2)} + y_{(k_2+1)}}{2} = \frac{y_{(9)} + y_{(10)}}{2} = 6.9 \end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 10.45$.

Aufgabe 2

Die gemeinsame Zähldichte $f_{X,Y}(i,j)$ zweier Zufallszahlen X und Y ist in der folgenden Tabelle gegeben.

$j \backslash i$	-1	0	1	2
0	$\frac{2}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$
1	$\frac{2}{20}$	0	$\frac{2}{20}$	$\frac{3}{20}$
2	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{2}{20}$

So ist z.B. $f_{X,Y}(-1,0) = \frac{2}{20}$ und $f_{X,Y}(2,1) = \frac{3}{20}$.

a) Bestimmen Sie die Zähldichte f_X von X.
Lösung: Gemäß (6.4) erhält man die Zähldichte f_X von X als Spaltensumme der obigen Matrix. Es ist also

i	-1	0	1	2
$f_X(i)$	$\frac{8}{20}$	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{6}{20}$

b) Berechnen Sie den Erwartungswert $\mathbb{E}(X)$ und die Varianz $V(X)$ der Zufallsvariablen X.
Lösung:

$$\mathbb{E}(X) = \sum_{i=-1}^{2} i \cdot f_X(i) = (-1) \cdot \frac{8}{20} + 0 \cdot \frac{2}{20} + 1 \cdot \frac{4}{20} + 2 \cdot \frac{6}{20} = \frac{8}{20} = 0.4$$

Mit $V(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ und

$$\mathbb{E}(X^2) = \sum_{i=-1}^{2} i^2 \cdot f_X(i) = (-1)^2 \cdot \frac{8}{20} + 0^2 \cdot \frac{2}{20} + 1^2 \cdot \frac{4}{20} + 2^2 \cdot \frac{6}{20} = \frac{36}{20} = 1.8$$

ergibt sich

$$V(X) = 1.8 - 0.4^2 = 1.64\ .$$

c) Berechnen Sie die Kovarianz $C(X,Y)$ und den Korrelationskoeffizienten $\rho(X,Y)$ von X und Y.
Hinweis: Verwenden Sie zusätzlich zu b) noch $\mathbb{E}Y = \frac{23}{20}$, $\mathbb{E}Y^2 = \frac{39}{20}$ und $\mathbb{E}(X \cdot Y) = \frac{2}{5}$.
Lösung: Wegen Satz 12.23 a) gilt $C(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$ und wegen Satz 12.8

$$\begin{aligned}
\mathbb{E}(X \cdot Y) &= \sum_{i=-1}^{2} \sum_{j=0}^{2} i \cdot j \cdot f_{X,Y}(i,j) \\
&= (-1) \cdot 1 \cdot \frac{2}{20} + (-1) \cdot 2 \cdot \frac{4}{20} + 1 \cdot 1 \cdot \frac{2}{20} + 1 \cdot 2 \cdot \frac{1}{20} + 2 \cdot 1 \cdot \frac{3}{20} + 2 \cdot 2 \cdot \frac{2}{20} \\
&= \frac{8}{20} = \frac{2}{5}
\end{aligned}$$

Aus der Tabelle ergibt sich noch $f_Y(0) = \frac{5}{20}$, $f_Y(1) = \frac{7}{20}$ und $f_Y(2) = \frac{8}{20}$ und damit

$$\mathbb{E}Y = 1 \cdot \frac{7}{20} + 2 \cdot \frac{8}{20} = \frac{23}{20} ,$$

$$\mathbb{E}Y^2 = 1^2 \cdot \frac{7}{20} + 2^2 \cdot \frac{8}{20} = \frac{39}{20} ,$$

und damit auch $V(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = \frac{39}{20} - \left(\frac{23}{20}\right)^2 = \frac{251}{400} = 0.6275$. Insgesamt

$$C(X,Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y = \frac{2}{5} - \frac{2}{5} \cdot \frac{23}{20} = -\frac{6}{100} = -0.06$$

und

$$\rho(X,Y) = \frac{C(X,Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{-0.06}{\sqrt{1.64 \cdot 0.6275}} = -0.0591$$

d) Bestimmen Sie $\mathbb{P}(Y = 0 \mid X = 1)$, $\mathbb{P}(Y = 1 \mid X = 1)$ und $\mathbb{P}(Y = 2 \mid X = 1)$.
Lösung: Nach Definition und wegen a) ist

$$\mathbb{P}(Y = j \mid X = 1) = \frac{\mathbb{P}(X = 1, Y = j)}{\mathbb{P}(X = 1)} = \frac{f_{X,Y}(1,j)}{f_X(1)} = 5 \cdot f_{X,Y}(1,j), \qquad j = 0, 1, 2.$$

j	0	1	2
$f_{Y\mid X}(1,j)$	$\frac{5}{20}$	$\frac{10}{20}$	$\frac{5}{20}$

e) Berechnen Sie $\mathbb{P}(Y \geq 1 \mid X = 1)$.
Lösung: Mit d) erhält man

$$\begin{aligned} \mathbb{P}(Y \geq 1 \mid X = 1) &= \mathbb{P}(Y = 1 \mid X = 1) + \mathbb{P}(Y = 2 \mid X = 1) \\ &= \frac{10}{20} + \frac{5}{20} = \frac{15}{20} = \frac{3}{4}. \end{aligned}$$

Aufgabe 3

Es sei X eine normalverteilte Zufallsvariable mit dem Erwartungswert -3 und der Varianz 9. Weiter sei Y eine normalverteilte Zufallsvariable mit dem Erwartungswert 2 und der Varianz 16. Die Zufallsvariablen X und Y seien stochastisch unabhängig.

a) Setzen Sie die richtigen Parameter ein.

$$X \text{ besitzt die Verteilung } \mathcal{N}\Big(\qquad , \qquad \Big).$$

Lösung: Da für eine $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsvariable X stets $\mathbb{E}X = \mu$ und $V(X) = \sigma^2$ gilt, ist hier

$$X \sim \mathcal{N}(-3, 9).$$

b) Berechnen Sie den Erwartungswert und die Varianz der Zufallsvariablen $Z := \frac{1}{3}X - \frac{1}{3}$ und $W := \frac{X-Y}{5}$.
Lösung: Wie in a) erhält man $Y \sim \mathcal{N}(2, 16)$. Wegen Satz 12.6 und Satz 12.11 und wegen der Unabhängigkeit von X und Y gilt:

$$\begin{aligned}
\mathbb{E}Z &= \frac{1}{3}\mathbb{E}X - \frac{1}{3} = \frac{1}{3} \cdot (-3) - \frac{1}{3} = -\frac{4}{3}, \\
V(Z) &= \frac{1}{3^2}V(X) = \frac{1}{9} \cdot 9 = 1, \\
\mathbb{E}W &= \frac{\mathbb{E}X - \mathbb{E}Y}{5} = \frac{-3-2}{5} = -1, \\
V(W) &= \frac{1}{5^2}V(X+Y) = \frac{1}{25}(V(X) + V(Y)) = \frac{25}{25} = 1.
\end{aligned}$$

c) Berechnen Sie die Kovarianz $C(W, Z)$ der Zufallsvariablen W und Z.
Lösung: Wegen Satz 12.23 und der Unabhängigkeit von X und Y, also $C(X, Y) = 0$, gilt

$$\begin{aligned}
C(W, Z) &= C\left(\frac{1}{5} \cdot (X - Y), \frac{1}{3}X - \frac{1}{3}\right) \\
&= \frac{1}{5} \cdot \frac{1}{3} \cdot C(X - Y, X) = \frac{1}{15} \cdot [C(X, X) - C(Y, X)] \\
&= \frac{V(X)}{15} = \frac{9}{15} = 0.6.
\end{aligned}$$

d) Bestimmen Sie $\mathbb{P}(Y > -4)$.
Lösung: Wegen $Y \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu := \mathbb{E}Y = 2$ und $\sigma^2 := V(Y) = 16$ und (9.6) gilt

$$\begin{aligned}
\mathbb{P}(Y > -4) &= 1 - \mathbb{P}(Y \leq -4) = 1 - \Phi_{\mu,\sigma^2}(-4) = 1 - \Phi\left(\frac{-4-\mu}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{-4-2}{4}\right) = 1 - \Phi(-1.5) = \Phi(1.5).
\end{aligned}$$

Wegen $\Phi(1.5) = 0.9332$ (aus Tabelle A.1) folgt

$$\mathbb{P}(Y > -4) = 0.9332.$$

e) Bestimmen Sie das 0.95-Quantil $q_{0.95}$ der Zufallsvariablen Y.
Lösung: $q_{0.95}$ ist die Lösung q der Gleichung

$$F_Y(q) = \Phi_{2,16}(q) = \Phi\left(\frac{q-2}{4}\right) = 0.95.$$

Wegen $\Phi(1.6449) = 0.95$ (Skript 12.20 d)) gilt also

$$\frac{q-2}{4} = 1.6449$$

und damit

$$q_{0.95} = q = 1.6449 \cdot 4 + 2 = 8.58.$$

Aufgabe 4

Eine Serienschaltung bestehe aus vier elektrischen Elementen. Die Lebensdauern X_1, X_2, X_3, X_4 der vier Elemente seien stochastisch unabhängig und jeweils $Exp(1)$ verteilt.

a) Die Gesamtlebensdauer der Serienschaltung Y ist definiert durch

$$Y := \min\{X_1, X_2, X_3, X_4\}.$$

Bestimmen Sie die Verteilungsfunktion F_Y von Y. Welche Verteilung besitzt Y?
Lösung: Wegen der stochastischen Unabhängigkeit von X_1, X_2, X_3 und X_4 und wegen $X_1, X_2, X_3, X_4 \sim Exp(1)$ gilt für $t \geq 0$ (vergl. Skript Satz 11.18):

$$\begin{aligned} F_Y(t) &= 1-(1-F_{X_1}(t))\cdot(1-F_{X_2}(t))\cdot(1-F_{X_3}(t))\cdot(1-F_{X_4}(t)) \\ &= 1-(1-(1-e^{-t}))^4 = 1-(e^{-t})^4 = 1-e^{-4t} \end{aligned}$$

und für $t < 0$ gilt dann $F_Y(t) = 0$. Insbesondere gilt $Y \sim Exp(4)$.

b) Berechnen Sie $\mathbb{P}(Y > 2 \mid X_1 > 1)$.

Lösung: Aus Definition 10.5 mit

$$\{X_1 > 2\} \cap \{X_1 > 1\} = \{X_1 > 2\}$$

folgt

$$\begin{aligned} \mathbb{P}(Y > 2 \mid X_1 > 1) &= \frac{\mathbb{P}(Y > 2, X_1 > 1)}{\mathbb{P}(X_1 > 1)} \\ &= \frac{\mathbb{P}(\min\{X_1, X_2, X_3, X_4\} > 2, X_1 > 1)}{\mathbb{P}(X_1 > 1)} \\ &= \frac{\mathbb{P}(X_1 > 2, X_2 > 2, X_3 > 2, X_4 > 2, X_1 > 1)}{\mathbb{P}(X_1 > 1)} \\ &= \frac{\mathbb{P}(X_1 > 2, X_2 > 2, X_3 > 2, X_4 > 2)}{\mathbb{P}(X_1 > 1)} \\ &= \frac{\mathbb{P}(Y > 2)}{\mathbb{P}(X_1 > 1)} = \frac{e^{-8}}{e^{-1}} = e^{-7} = 9.12 \cdot 10^{-4}. \end{aligned}$$

Sei in den folgenden drei Aufgabenteilen die Zufallsvariable $Z \sim Exp(5)$.

c) Bestimmen Sie den Erwartungswert $\mathbb{E}Z$ und die Varianz $V(Z)$ von Z.
Lösung: Da $Z \sim Exp(5)$ ist, gilt

$$\mathbb{E}Z = \frac{1}{5}$$

und

$$V(Z) = \frac{1}{25}$$

nach den Tabellen auf S. 124 und S. 128.

d) Berechnen Sie $\mathbb{P}(Z > 0.3)$.
Lösung: Mit a) erhält man

$$\mathbb{P}(Z > 0.3) = 1 - F_Z(0.3) = e^{-5 \cdot 0.3} = e^{-1.5} = 0.2231.$$

e) Geben Sie den größten Wert t an, für welche die Wahrscheinlichkeit $\mathbb{P}(Z > t) \geq 0.7$ ist.
Lösung: Aus dem Ansatz

$$\begin{aligned} \mathbb{P}(Z > t) \geq 0.7 &\iff e^{-5t} \geq 0.7 \\ &\iff -5t \geq \log(0.7) \\ &\iff t \leq -\log(0.7)/5 \;=\; 0.0713 \end{aligned}$$

erhält man $t = 0.0713$.

Aufgabe 5

Ein Merkmal besitze eine logarithmische Normalverteilung $\mathcal{LN}(\vartheta, 4)$ mit der Dichte

$$f_\vartheta(x) = \begin{cases} 0 & , x \leq 0 \\ \frac{1}{2 \cdot x \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{8}(\ln(x)-\vartheta)^2} & , x > 0, \end{cases}$$

und unbekanntem Parameter $\vartheta \in \mathbb{R}$.

a) Bestimmen Sie die Loglikelihood-Funktion $M_x(\vartheta)$ zur Stichprobe $x = (x_1, \ldots .x_n)$, wobei $x_i > 0$ für $i = 1, \ldots, n$.

b) Berechnen Sie die Ableitung $M_x'(\vartheta)$.

c) Bestimmen Sie den Maximum-Likelihhod-Schätzer $\hat{\vartheta}(x)$ für ϑ zur Stichprobe x.

d) Sei Y eine Zufallsvariable mit der Verteilung $\mathcal{LN}(\vartheta, 4)$. Welche Verteilung besitzt die Zufallsvariable $Z := \ln(Y)$, ihr natürlicher Logarithmus?

e) Bestimmen Sie den Momentenschätzer $\hat{\vartheta}_m(x)$ für ϑ zur Stichprobe $x = (x_1, \ldots, x_n)$.

Lösung:

a) Mit

$$\ln(f_\vartheta(x)) = -\ln(2 \cdot x \cdot \sqrt{2\pi}) - \frac{1}{8}(\ln(x) - \vartheta)^2$$

ergibt sich die Loglikelihood-Funktion zu

$$\begin{aligned} M_x(\vartheta) &= \sum_{i=1}^n \ln(f_\vartheta(x_i)) = \sum_{i=1}^n \left(-\ln(2 \cdot x_i \cdot \sqrt{2\pi}) - \frac{1}{8}(\ln(x_i) - \vartheta)^2 \right) \\ &= -\sum_{i=1}^n \ln(\sqrt{2\pi} \cdot 2 \cdot x_i) - \frac{1}{8}\sum_{i=1}^n (\ln(x_i) - \vartheta)^2 \end{aligned}$$

b) Ihre Ableitung ist

$$M_x'(\vartheta) \;=\; -\frac{1}{8}\sum_{i=1}^n -2 \cdot (\ln(x_i) - \vartheta) = \frac{1}{4}\sum_{i=1}^n (\ln(x_i - \vartheta) = \frac{1}{4}\left(\sum_{i=1}^n \ln(x_i) - n \cdot \vartheta\right)$$

c) Da die zweite Ableitung

$$M_x''(\vartheta) \quad = \quad -\frac{n}{4} < 0.$$

von M_x überall kleiner als 0 ist, ergibt sich die Maximumstelle $\hat{\vartheta}(x)$ von $\vartheta \to M_x(\vartheta)$ als Lösung von $M_x'(\vartheta) = 0$, also $\sum_{i=1}^{n} \ln(x_i) = n \cdot \vartheta$ und

$$\hat{\vartheta}(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln(x_i)$$

d) Bekannt ist (Skriptum 9.10), dass für eine $\mathcal{N}(\mu, \sigma^2)$-verteilte Zufallsvariable Z die Zufallsvariable $Y = e^Z$ die logarithmische Normalverteilung $\mathcal{LN}(\mu, \sigma^2)$ besitzt. Besitzt dann umgekehrt Y die Verteilung $\mathcal{LN}(\mu, \sigma^2)$, so besitzt dann $Z = \ln(Y)$ die Verteilung $\mathcal{N}(\mu, \sigma^2)$. Hier ist $\mu = \vartheta$ und $\sigma^2 = 4$. Also besitzt Z die Verteilung $\mathcal{N}(\vartheta, 4)$.

e) Für den Momentenschätzer ist gemäß Skriptum 17.1.3 die Gleichung

$$m_1(\vartheta) = \mathbb{E}_\vartheta(X_1) = \hat{m}_1(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \bar{x}$$

nach ϑ aufzulösen. Da X_1 die Verteilung $\mathcal{LN}(\vartheta, 4)$ besitzt, gilt nach der Tabelle auf S. 124 im Skriptum $\mathbb{E}_\vartheta(X_1) = e^{\vartheta + 1/2 \cdot 4} = e^{\vartheta + 2}$ und damit

$$\begin{aligned} e^{\vartheta+2} &= \bar{x} \\ \vartheta + 2 &= \ln(\bar{x}) \\ \vartheta &= \ln(\bar{x}) - 2 \end{aligned}$$

Der gesuchte Momentenschätzer ist also $\hat{\vartheta}_m(x) := \ln(\bar{x}) - 2$.