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

WS 2008/09

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 3.3.2009

Musterlösungen

Aufgabe B1

Gegeben sei eine Urliste mit den Paaren $(x_1, y_1), \ldots, (x_{11}, y_{11})$

j	1	2	3	4	5	6	7	8	9	10	11
x_j	-2.8	-1.6	-0.7	0.4	1.4	1.8	3.5	3.8	5.1	6.1	6.7
y_j	10	5.4	4.5	5.9	2.4	5.4	1.9	5.6	-1.7	0.3	-4.9

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

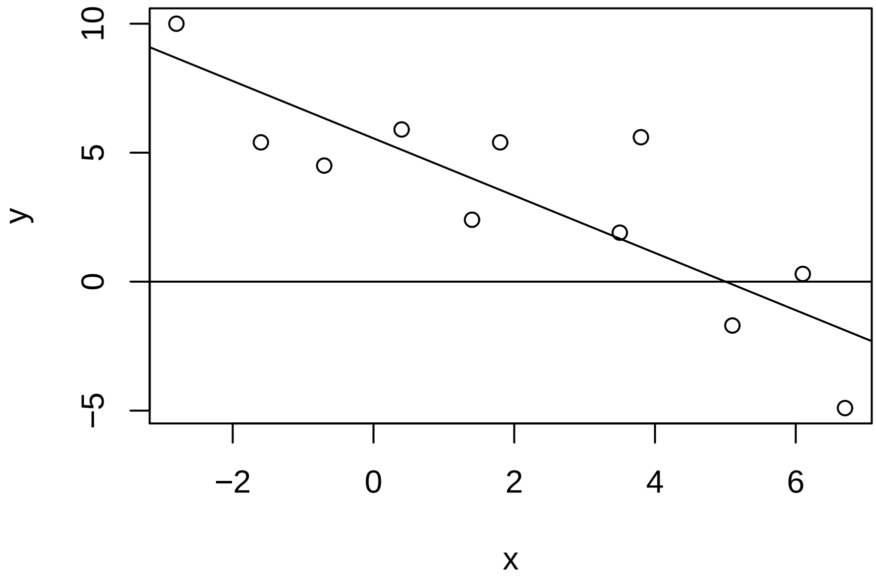
$$\sum_{j=1}^{n}(x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 2.15 \qquad s_x = 3.162$$
$$\bar{y} = 3.16 \qquad s_y = 4.135$$
$$r_{xy} = -0.8494$$

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

$$\begin{aligned} b^* &= -1.111 \\ a^* &= 5.56 \end{aligned}$$

und die Regressionsgerade $y = 5.56 - 1.111 \cdot x$.



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

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

$$y_{()} = (-4.9,\ -1.7,\ 0.3,\ 1.9,\ 2.4,\ 4.5,\ 5.4,\ 5.4,\ 5.6,\ 5.9,\ 10)$$

c) Berechnen Sie das 0.2-getrimmte Stichprobenmittel $\bar{y}_{0.2}$ von $(y_1, \ldots, y_{11})$.
Lösung: Mit $k = [11 \cdot 0.2] = 2$ ergibt sich

$$\bar{y}_{0.2} = \frac{1}{11 - 2 \cdot 2} \cdot (y_{(3)} + \ldots + y_{(9)}) = 3.643$$

d) Bestimmen Sie das Stichproben-0.2-Quantil $\tilde{y}_{0.2}$ von $(y_1, \ldots, y_{11})$.
Lösung: Da $11 \cdot 0.2 = 2.2$ nicht ganzzahlig ist, ist mit $k = [2.2] = 2$

$$\tilde{y}_{0.2} = y_{(k+1)} = y_{(3)} = 0.3$$

e) Berechnen Sie den Quartilsabstand von $(y_1, \ldots, y_{11})$.
Lösung: Da $0.25 \cdot 11 = 2.75$ und $0.75 \cdot 11 = 8.25$ beide nicht ganzzahlig sind, ergibt sich mit $k_1 = [2.75] = 2$ und $k_2 = [8.25] = 8$

$$\begin{aligned}\tilde{y}_{0.25} &= y_{(k_1+1)} = y_{(3)} = 0.3\\ \tilde{y}_{0.75} &= y_{(k_2+1)} = y_{(9)} = 5.6\end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 5.3$.

Aufgabe B2

Ein LKW wird insgesamt n-mal durch einen Kran mit Kies beladen. Bei jedem Beladungsvorgang i erfasst der Kran unabhängig von den anderen Beladungsvorgängen eine zufällige Masse X_i an Kies, die normalverteilt ist mit Mittelwert $\mu = 300$ (in kg) und Varianz $\sigma^2 = 4000$ (in kg^2), wobei $i = 1, \ldots, n$.

a) Welche Verteilung besitzt S_n, die zufällige Gesamtmasse an Kies bei n unabhängigen Beladungsvorgängen?

b) Der LKW selbst hat eine Masse von 5000 kg. Welche Verteilung besitzt G_n, die Masse des LKW's inklusive der Beladung nach n unabhängigen Beladungsvorgängen?

c) Bestimmen Sie den Erwartungswert $\mathbb{E}G_n$ und die Varianz $V(G_n)$.

d) Wie groß ist die Wahrscheinlichkeit, dass die Gesamtmasse S_{25} mindestens 7200 kg beträgt, wenn der LKW mit $n = 25$ Kranfüllungen beladen wird.

e) Wie groß muss n mindestens sein, damit $\mathbb{P}(S_n \geq 6000) \geq 0.99$ gilt?
Hinweis:

n	$\frac{15n-300}{\sqrt{10n}}$
20	0.00
21	1.04
22	2.02
23	2.97

Lösung: $S_n = X_1 + \ldots + X_n$ ist die zufällige Gesamtmasse.

a) Nach Voraussetzung gilt $X_i \sim \mathcal{N}(300, 4000)$ für $i = 1, \ldots, n$. Da $X_1, \ldots, X_n$ stochastisch unabhängig sind, gilt nach der Faltungsformel (Skriptum, 11.16) $S_n \sim \mathcal{N}(300 \cdot n, 4000 \cdot n)$.

b) Wegen $G_n = 5000 + S_n$ und a) gilt nach Satz 9.7 $G_n \sim \mathcal{N}(5000 + 300 \cdot n, 4000 \cdot n)$.

c) Nach den Tabellen im Skriptum 12.7 und in 12.13 gilt wegen b) $\mathbb{E}G_n = 5000 + 300 \cdot n$ und $V(G_n) = 4000 \cdot n$.

d) Für den Fall $n = 25$ gilt $S_n \sim \mathcal{N}(n \cdot 300, n \cdot 4000) = \mathcal{N}(7500, 100000)$. Daher ist die gesuchte Wahrscheinlichkeit

$$\begin{aligned}\mathbb{P}(S_{25} \geq 7200) = 1 - \mathbb{P}(S_{25} < 7200) &\overset{(*)}{=} 1 - \mathbb{P}(S_{25} \leq 7200) = 1 - \Phi\left(\frac{7200 - 7500}{\sqrt{100000}}\right) \\ &= 1 - \Phi(-0.9487) \approx 1 - \Phi(-0.95) = \Phi(0.95) = 0.8289\end{aligned}$$

An Stelle $(*)$ wurde Satz 8.11 verwendet und ausgenützt, dass S_{25} eine stetige Zufallsvariable ist.

e) Gesucht ist das kleinste n, so dass $\mathbb{P}(S_n \geq 6000) \geq 0.99$ gilt, d.h. wie in d), dass

$$\begin{aligned}\mathbb{P}(S_n \geq 6000) = 1 - \mathbb{P}(S_n \leq 6000) = 1 - \Phi\left(\frac{6000 - 300n}{\sqrt{4000n}}\right) &= 1 - \Phi\left(\frac{6000 - 300n}{20 \cdot \sqrt{10n}}\right) \\ = 1 - \Phi\left(\frac{300 - 15n}{\sqrt{10n}}\right) &\overset{(**)}{=} \Phi\left(\frac{15n - 300}{\sqrt{10n}}\right) \geq 0.99\end{aligned}$$

gilt, wobei an Stelle $(**)$ Skriptum, (9.5) ausgenützt wurde. Wegen dem Hinweis und wegen $\Phi(1.04) = 0.8508 < 0.99$, $\Phi(2.02) = 0.9783 < 0.99$ und $\Phi(2.97) = 0.9985 > 0.99$ gilt also $\mathbb{P}(S_n \geq 6000) \geq 0.99$ für $n = 23, 24, \ldots$. n muss also mindestens 23 sein.

Aufgabe B3

Eine Urne enthalte 10 gleichgroße Kugeln in den Farben rot und blau, die 20 g oder 50 g wiegen, entsprechend den Anzahlen:

	rot	blau
20 g	3	2
50 g	4	1

Man bestimme die Wahrscheinlichkeiten

a) p_1, dass eine „rein zufällig“ herausgezogene Kugel rot ist,

b) p_2, dass eine „rein zufällig“ herausgezogene Kugel blau ist und 50 g wiegt,

c) p_3, dass eine „rein zufällig“ herausgezogene Kugel 50 g wiegt.

d) Berechnen Sie die bedingte Wahrscheinlichkeit p_4, dass eine „rein zufällig“ herausgezogene Kugel rot ist, wenn bekannt ist, dass sie 50 g wiegt.

Berechnen Sie die Wahrscheinlichkeiten p_5 bzw. p_6 dafür, dass bei „rein zufälliger“ aufeinanderfolgender Entnahme von Kugeln zuerst eine blaue zu 50 g, dann eine beliebig schwere rote, dann eine beliebig schwere blaue und zuletzt eine blaue Kugel zu 50 g gezogen wird, wenn die entnommenen Kugeln

e) jeweils vor der nächsten Ziehung wieder zurückgelegt werden,

f) nicht wieder zurückgelegt werden?

Hinweis: Die auftretenden Brüche brauchen nicht ausgerechnet zu werden.

Lösung: Sei A das Ereignis, dass eine rote Kugel gezogen wird und B das Ereignis, dass die gezogene Kugel 50 g wiegt.

a) Da in der Urne insgesamt 10 Kugeln sind und davon 7 rot sind, gilt $p_1 = \mathbb{P}(A) = \frac{7}{10} = 0.7$.

b) Da in der Urne insgesamt 10 Kugeln sind und davon 1 blau ist und 50 g wiegt, ist die gesuchte Wahrscheinlichkeit $p_2 = \mathbb{P}(A^c \cap B) = \frac{1}{10} = 0.1$.

c) Da in der Urne insgesamt 10 Kugeln sind und davon 5 20 g wiegen, ist die gesuchte Wahrscheinlichkeit $p_3 = \mathbb{P}(B) = \frac{5}{10} = 0.5$.

d) Gesucht ist die bedingte Wahrscheinlichkeit $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$. $A \cap B$ ist das Ereignis, dass die gezogene Kugel rot ist und 50 g wiegt. Da in der Urne 4 derartige Kugeln sind, gilt $\mathbb{P}(A \cap B) = \frac{4}{10} = 0.4$ und damit

$$p_4 = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{4/10}{5/10} = \frac{4}{5}.$$

e) Da die entnommenen Kugeln jedesmal zurückgelegt werden, liegt ein ideales Zufallsexperiment vor, dass viermal durchführt wird und die gesuchte Wahrscheinlichkeit ist wegen der Produktregel

$$p_5 = \mathbb{P}(A^c \cap B) \cdot \mathbb{P}(A) \cdot \mathbb{P}(A^c) \cdot \mathbb{P}(A^c \cap B) = \frac{1}{10} \cdot \frac{7}{10} \cdot \frac{3}{10} \cdot \frac{1}{10} = \frac{21}{10000}.$$

f) Die gesuchte Wahrscheinlichkeit ist $p_6 = 0$, da es unmöglich ist, zweimal eine blaue Kugel zu 50 g zu ziehen, wenn die gezogene Kugel nicht zurückgelegt wird und insgesamt nur eine derartige Kugel vorhanden ist.

Aufgabe B4

Ein Autohändler verkauft je Tag X Autos des Typs **1** und Y Autos des Typs **2**. Die Tabelle zeigt die gemeinsame Zähldichte $f_{X,Y}(i,j) = \mathbb{P}(X = i, Y = j)$ von X und Y für $i, j = 0, 1, 2$.

i \ j	0	1	2
0	0.1	0.1	0.0
1	0.1	0.3	0.2
2	0.0	0.1	0.1

Es ist also z.B. $\mathbb{P}(X = 1, Y = 2) = f_{X,Y}(1, 2) = 0.2$.

a) Bestimmen Sie die Zähldichten f_X von X und f_Y von Y.

b) Berechnen Sie die bedingte Wahrscheinlichkeit $\mathbb{P}(X \geq 1 \mid Y = 1)$.

c) Sind X und Y stochastisch unabhängig ?

d) Berechnen Sie die Erwartungswerte $\mathbb{E}X$, $\mathbb{E}Y$ und die Varianzen $V(X)$ und $V(Y)$.

e) Berechnen Sie den Korrelationskoeffizienten $\rho(X, Y)$ von X und Y.

Lösung:

a) Die Bestimmung der Zähldichte von X bzw. Y geschieht durch Summation der Wahrscheinlichkeiten der gemeinsamen Verteilung über die möglichen Werte von Y bzw. X (vergl. Skriptum Beispiel 6.7 und Tabelle 6.1)

i \ j	0	1	2	Σ ($f_X(i)$)
0	0.1	0.1	0.0	0.2
1	0.1	0.3	0.2	0.6
2	0.0	0.1	0.1	0.2
Σ ($f_Y(j)$)	0.2	0.5	0.3	1.0

b) Nach Definition gilt $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, falls $\mathbb{P}(B) > 0$. Da $A \to \mathbb{P}(A|B)$ ein Wahrscheinlichkeitsmaß ist (Satz 10.11), gilt

$$\begin{aligned}
\mathbb{P}(X \geq 1 \mid Y = 1) &= \mathbb{P}(X = 1 \mid Y = 1) + \mathbb{P}(X = 2 \mid Y = 1) \\
&= \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(Y = 1)} + \frac{\mathbb{P}(X = 2, Y = 1)}{\mathbb{P}(Y = 1)} = \frac{0.3}{0.5} + \frac{0.1}{0.5} = \frac{4}{5} = 0.8
\end{aligned}$$

c) X und Y sind nicht stochastisch unabhängig, denn $\mathbb{P}(X = 0, Y = 2) = 0.0 \neq 0.2 \cdot 0.3 = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 2)$.

d) Es gilt

$$\begin{aligned}\mathbb{E}X &= 0\cdot f_X(0)+1\cdot f_X(1)+2\cdot f_X(2) = 0\cdot 0.2+1\cdot 0.6+2\cdot 0.2 = 1.0\\ \mathbb{E}X^2 &= 0^2\cdot f_X(0)+1^2\cdot f_X(1)+2^2\cdot f_X(2) = 0\cdot 0.2+1\cdot 0.6+4\cdot 0.2 = 1.4\end{aligned}$$

und damit dann $V(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 1.4 - 1.0^2 = 0.4$.

Analog

$$\begin{aligned}\mathbb{E}Y &= 0\cdot f_Y(0)+1\cdot f_Y(1)+2\cdot f_Y(2) = 0\cdot 0.2+1\cdot 0.5+2\cdot 0.3 = 1.1\\ \mathbb{E}Y^2 &= 0^2\cdot f_Y(0)+1^2\cdot f_Y(1)+2^2\cdot f_Y(2) = 0\cdot 0.2+1\cdot 0.5+4\cdot 0.3 = 1.7\end{aligned}$$

und damit dann $V(Y) = \mathbb{E}Y^2 - (\mathbb{E}Y)^2 = 1.7 - 1.1^2 = 0.49$.

e) Es gilt $C(X,Y) = \mathbb{E}(X\cdot Y) - \mathbb{E}X\cdot\mathbb{E}Y$. Aus der Tabelle ergibt sich

$$\begin{aligned}\mathbb{E}(X\cdot Y) &= 1\cdot 1\cdot f_{X,Y}(1,1)+1\cdot 2\cdot f_{X,Y}(1,2)+2\cdot 1\cdot f_{X,Y}(2,1)+2\cdot 2\cdot f_{X,Y}(2,2)\\ &= 1\cdot 0.3+2\cdot 0.2+2\cdot 0.1+4\cdot 0.1 = 1.3\end{aligned}$$

und damit $C(X,Y) = 1.3 - 1.1\cdot 1.0 = 0.2$, also

$$\rho(X,Y) = \frac{C(X,Y)}{\sqrt{V(X)\cdot V(Y)}} = \frac{0.2}{\sqrt{0.49\cdot 0.4}} = 0.4518$$

Aufgabe B5

Ein Merkmal habe die Dichte der Erlang-Verteilung $E(4, 1/\vartheta)$

$$t \to f_\vartheta(t) := \frac{t^3}{6\vartheta^4}\cdot e^{-t/\vartheta}, \qquad t > 0,$$

wobei $\vartheta \in \Theta := (0,\infty)$ ein unbekannter Parameter ist. ϑ soll aufgrund einer unabhängigen Stichprobe $x = (x_1,\ldots,x_n)$ geschätzt werden, wobei $x_1 > 0,\ldots,x_n > 0$ vorausgesetzt wird.

a) Bestimmen Sie die Loglikelihood-Funktion zur Stichprobe $x = (x_1,\ldots,x_n)$.

b) Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat\vartheta(x_1,\ldots,x_n)$ für ϑ.

c) Berechnen Sie den Erwartungswert von $\hat\vartheta(X) = \hat\vartheta(X_1,\ldots,X_n)$ bei Vorliegen von ϑ für den Fall, dass die unabhängigen Stichprobenvariablen $X_1,\ldots,X_n$ die Erlang-Verteilung $E(4, 1/\vartheta)$ besitzen.

Hinweis: Bestimmen Sie zuerst $\mathbb{E}_\vartheta(X_j)$ für $j = 1,\ldots,n$.

d) Welche der drei folgenden Eigenschaften sind für den Maximum-Likelihood-Schätzer $\hat\vartheta(x_1,\ldots,x_n)$ erfüllt?

d_1) $\hat\vartheta(x_1,\ldots,x_n)$ ist erwartungstreu für ϑ.

d_2) Die Schätzfolge $(\hat\vartheta(x_1,\ldots,x_n))_{n\in\mathbb{N}}$ ist asymptotisch erwartungstreu für ϑ.

d_3) Die Schätzfolge $(\hat\vartheta(x_1,\ldots,x_n))_{n\in\mathbb{N}}$ ist konsistent für ϑ.

Lösung:

a) Es ist

$$\log(f_\vartheta(t)) \;=\; 3\log(t) - \log(6) - 4\log(\vartheta) - \frac{t}{\vartheta}\,, \qquad t > 0,$$

also

$$\begin{aligned} M_x(\vartheta) &= \sum_{j=1}^n \log(f_\vartheta(x_j)) \;=\; \sum_{j=1}^n \Big(3\log(x_j) - \log(6) - 4\log(\vartheta) - \frac{x_j}{\vartheta}\Big) \\ &= 3\sum_{j=1}^n \log(x_j) - n\log(6) - 4n\log(\vartheta) - \frac{1}{\vartheta}\sum_{j=1}^n x_j. \end{aligned}$$

b) Es gilt

$$M_x'(\vartheta) = -\frac{4n}{\vartheta} + \frac{1}{\vartheta^2}\sum_{j=1}^n x_j \;=\; -\frac{4n}{\vartheta^2}\left(\vartheta - \frac{1}{4n}\sum_{j=1}^n x_j\right) \begin{array}{c} > \\ = \\ < \end{array} 0$$

genau dann, wenn (beachte $-\frac{4n}{\vartheta^2} < 0$)

$$\vartheta \begin{array}{c} < \\ = \\ > \end{array} \frac{1}{4n}\sum_{j=1}^n x_j =: \hat{\vartheta}(x).$$

Daher ist $\hat{\vartheta}(x)$ die einzige Maximumstelle von $\vartheta \to M_x(\vartheta)$, insbesondere ist

$$(x_1, \dots, x_n) \to \hat{\vartheta}(x) = \frac{1}{4n}\sum_{j=1}^n x_j$$

der gesuchte Maximum-Likelihood-Schätzer für ϑ.

c) Da die Stichprobenvariablen $X_1, \dots, X_n$ die Erlang-Verteilung $\Gamma(4, 1/\vartheta)$ (eine spezielle Gamma-Verteilung) besitzen, gilt nach der Tabelle vor Satz 12.8 $\mathbb{E}_\vartheta X_j = 4\vartheta$ und damit

$$\mathbb{E}_\vartheta \hat{\vartheta}(X) = \mathbb{E}_\vartheta\left(\frac{1}{4n}\sum_{j=1}^n X_j\right) = \frac{1}{4n}\sum_{j=1}^n \mathbb{E}_\vartheta X_j \;=\; \frac{1}{4n}\cdot n\cdot 4\vartheta \;=\vartheta.$$

d) Direkt aus c) und der Definition der Erwartungstreue in Definition 17.11 folgt, dass $\hat{\vartheta}(x_1, \dots, x_n)$ erwartungstreu für ϑ ist. Damit gilt d_1). Wegen

$$\lim_{n\to\infty} \mathbb{E}_\vartheta(\hat{\vartheta}(X_1, \dots, X_n)) = \lim_{n\to\infty} \vartheta = \vartheta$$

ist die Schätzfolge $(\hat{\vartheta}(x_1, \dots, x_n))_{n\in\mathbb{N}}$ damit auch asymptotisch erwartungstreu für ϑ und es gilt d_2). Um zu begründen, dass die Schätzfolge $(\hat{\vartheta}(x_1, \dots, x_n))_{n\in\mathbb{N}}$ auch konsistent für ϑ ist, kann man darauf verweisen, dass Maximum-Likelihood-Schätzer unter schwachen Regularitätsvoraussetzungen (die hier erfüllt sind und was wir nicht zeigen wollen) konsistent sind. Eine präzisere Begründung (für die Interessierten) erhält man durch Verweis auf Satz 17.19a):

Der Schätzer $\bar{x}$ ist konsistent für $\gamma(\vartheta) := \mathbb{E}_\vartheta(X_1) = 4\vartheta$ und daher gilt

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}_\vartheta(|\hat{\vartheta}(X_1, \dots, X_n) - \vartheta| \geq \epsilon) &= \lim_{n\to\infty} \mathbb{P}_\vartheta(|4\hat{\vartheta}(X_1, \dots, X_n) - 4\vartheta| \geq 4\epsilon) \\ &= \lim_{n\to\infty} \mathbb{P}_\vartheta(|\bar{X} - \gamma(\vartheta)| \geq 4\epsilon) = 0 \end{aligned}$$

für beliebiges $\epsilon > 0$.