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

WS 2004/05

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 8.3.2005

Musterlösungen

Aufgabe B1:

Gegeben sei eine Urliste mit den Paaren $(x_1, y_1), \ldots, (x_{10}, y_{10})$

j	1	2	3	4	5	6	7	8	9	10
x_j	0.8	1.8	3.1	4	4.9	6	6.8	8.2	9.1	10.2
y_j	15.2	12.7	8.5	12.2	9.1	4.8	9.5	-1.6	0	-1.9

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

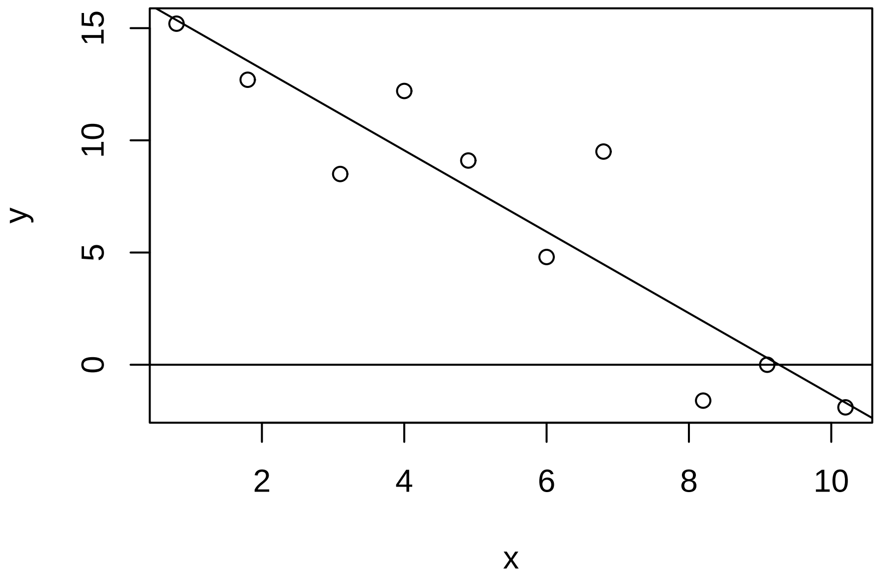
$$\sum_{j=1}^{n} (x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 5.49 \qquad s_x = 3.135$$
$$\bar{y} = 6.85 \qquad s_y = 6.205$$
$$r_{xy} = -0.9164$$

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

$$b^* = -1.814$$
$$a^* = 16.81$$

und die Regressionsgerade $y = 16.81 - 1.814 \cdot x$.



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

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

$$y_{()} = (-1.9,\ -1.6,\ 0,\ 4.8,\ 8.5,\ 9.1,\ 9.5,\ 12.2,\ 12.7,\ 15.2)$$

c) Berechnen Sie das 0.14-getrimmte Stichprobenmittel $\bar{y}_{0.14}$ von $(y_1, \ldots, y_{10})$.
Lösung: Mit $k = [10 \cdot 0.14] = 1$ ergibt sich

$$\bar{y}_{0.14} = \frac{1}{10 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(9)}) = 6.9$$

d) Berechnen Sie den Quartilsabstand von $(y_1, \ldots, y_{10})$.
Lösung: Da $0.25 \cdot 10 = 2.5$ und $0.75 \cdot 10 = 7.5$ beide nicht ganzzahlig sind, ergibt sich mit $k_1 = [2.5] = 2$ und $k_2 = [7.5] = 7$

$$\begin{aligned} \tilde{y}_{0.25} &= y_{(k_1+1)} = y_{(3)} = 0 \\ \tilde{y}_{0.75} &= y_{(k_2+1)} = y_{(8)} = 12.2 \end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 12.2$.

Aufgabe B2

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 $i, j = 0, 1, 2$.

$i \backslash j$	0	1	2
0	0.2	0.1	0.0
1	0.1	0.2	0.1
2	0.0	0.2	0.1

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

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

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

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)

$i \backslash j$	0	1	2	Σ	
0	0.2	0.1	0.0	0.3	
1	0.1	0.2	0.1	0.4	$f_X(i)$
2	0.0	0.2	0.1	0.3	
Σ	0.3	0.5	0.2	1.0	
	$f_Y(j)$				

Damit ergibt sich für die Zähldichten f_X und f_Y

i	0	1	2
$f_X(i)$	0.3	0.4	0.3
$f_Y(i)$	0.3	0.5	0.2

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

$$\begin{aligned}
\mathbb{P}(X \geq 1 \mid Y = 1) &= \mathbb{P}(X = 1 \mid Y = 1) + \mathbb{P}(X = 2 \mid Y = 1) \\
&= \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(Y = 1)} + \frac{\mathbb{P}(X = 2, Y = 1)}{\mathbb{P}(Y = 1)} = \frac{0.2}{0.5} + \frac{0.2}{0.5} = \frac{4}{5} = 0.8000
\end{aligned}$$

c) Es gilt

$$\begin{aligned}
\mathbb{E}X &= 0 \cdot f_X(0) + 1 \cdot f_X(1) + 2 \cdot f_X(2) = 0 \cdot 0.3 + 1 \cdot 0.4 + 2 \cdot 0.3 = 1.0 \\
\mathbb{E}X^2 &= 0^2 \cdot f_X(0) + 1^2 \cdot f_X(1) + 2^2 \cdot f_X(2) = 0 \cdot 0.3 + 1 \cdot 0.4 + 4 \cdot 0.3 = 1.6
\end{aligned}$$

und damit dann $V(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 1.6 - 1.0^2 = 0.60$.

Analog

$$\mathbb{E}Y = 0 \cdot f_Y(0) + 1 \cdot f_Y(1) + 2 \cdot f_Y(2) = 0 \cdot 0.3 + 1 \cdot 0.5 + 2 \cdot 0.2 = 0.9$$

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

$$\begin{aligned}
\mathbb{E}(X \cdot Y) &= 1 \cdot 1 \cdot f_{X,Y}(1,1) + 1 \cdot 2 \cdot f_{X,Y}(1,2) + 2 \cdot 1 \cdot f_{X,Y}(2,1) + 2 \cdot 2 \cdot f_{X,Y}(2,2) \\
&= 1 \cdot 0.2 + 2 \cdot 0.1 + 2 \cdot 0.2 + 4 \cdot 0.1 = 1.2
\end{aligned}$$

und damit $C(X,Y) = 1.2 - 1.0 \cdot 0.9 = 0.3$, insbesondere $\rho(X,Y) > 0$. X und Y sind also positiv korreliert.

Aufgabe B3

Ein Unternehmen beschäftigt 12 Verkäufer, deren tägliche Umsätze voneinander unabhängig und normalverteilt seien. 10 dieser Verkäufer erzielen dabei einen mittleren täglichen Umsatz von 1600 € bei einer Standardabweichung von 480 €, während die restlichen 2 einen mittleren Umsatz von 2000 € bei einer Standardabweichung von 600 € erreichen. Sei Z der tägliche Gesamtumsatz aller Verkäufer.

a) Man berechne den Erwartungswert und die Varianz von Z.

b) Welche Verteilung besitzt $\bar{Z} = \frac{Z}{12}$, der tägliche mittlere Umsatz aller Verkäufer?

c) Mit welcher Wahrscheinlichkeit p ist der tägliche Gesamtumsatz Z größer als 19600 €?

d) Bestimmen Sie denjenigen Gesamtumsatz t, so dass $\mathbb{P}(Z > t) = 0.025$ gilt.

Lösung:

a) Seien $Z_1, \ldots, Z_{10} \sim \mathcal{N}(1600, 480^2)$ die täglichen Einzelumsätze der 10 Verkäufer und $Z_{11}, \ldots, Z_{12} \sim \mathcal{N}(2000, 600^2)$ die täglichen Einzelumsätze der restlichen 2 Verkäufer. Dann gilt

$$\mathbb{E}Z = \mathbb{E}(Z_1 + \ldots + Z_{12}) = \mathbb{E}Z_1 + \ldots + \mathbb{E}Z_{12} = 10 \cdot 1600 + 2 \cdot 2000 = 20000$$

und wegen der Unabhängigkeit der Einzelumsätze nach Satz 12.23 f)

$$V(Z) = V(Z_1 + \ldots + Z_{12}) = V(Z_1) + \ldots + V(Z_{12}) = 10 \cdot 480^2 + 2 \cdot 600^2 = 3024000.$$

Damit gilt $\mathbb{E}Z = 20000$ und $V(Z) = 3024000$.

b) Wegen der Unabhängigkeit der Einzelumsätze gilt nach der Tabelle auf S. 114

$$\begin{aligned} Z &= Z_1 + \ldots + Z_{12} \\ &\sim \mathcal{N}(1600, 480^2) * \ldots * \mathcal{N}(1600, 480^2) * \mathcal{N}(2000, 600^2) * \mathcal{N}(2000, 600^2) \\ &\sim \mathcal{N}(10 \cdot 1600 + 2 \cdot 2000, 10 \cdot 480^2 + 2 \cdot 600^2) = \mathcal{N}(20000, 3024000) \end{aligned}$$

Nach Satz 7. gilt dann

$$\bar{Z} = \frac{1}{12} \cdot Z \sim \mathcal{N}\left(\frac{20000}{12}, \frac{3024000}{12^2}\right) = \mathcal{N}(1666.67, 21000)$$

c) Zu berechnen ist $p = \mathbb{P}(Z > 19600)$.

Wegen Satz 9.6 und wegen $\Phi(-x) = 1 - \Phi(x), x \in \mathbb{R}$ gilt

$$\begin{aligned} \mathbb{P}(Z > 19600) &= 1 - \mathbb{P}(Z \leq 19600) = 1 - \Phi\left(\frac{19600 - 20000}{\sqrt{3024000}}\right) \\ &= 1 - \Phi(-0.2300) = \Phi(0.2300) \approx \Phi(0.23) = 0.5910 \end{aligned}$$

nach Anhang A.1 im Skriptum. Die gesuchte Wahrscheinlichkeit ist also

$$p = 0.5910$$

d) Gesucht ist $t \in \mathbb{R}$ mit $\mathbb{P}(Z > t) = 1 - \mathbb{P}(Z \leq t) = 0.025$, also mit

$$0.975 = \mathbb{P}(Z \leq t) = \Phi\left(\frac{t - 20000}{\sqrt{3024000}}\right)$$

Wegen 12.20 d) oder nach Tabelle A.1 gilt $\Phi(1.96) = 0.975$, d.h. 1.96 ist das 0.975-Quantil von $\mathcal{N}(0, 1)$. Damit ergibt sich

$$1.96 = \frac{t - 20000}{\sqrt{3024000}}$$

also

$$t = 20000 + 1.96 \cdot \sqrt{3024000} = 23410.$$

Aufgabe B4

Ein Radiobastler hört an drei Tagen seinen Lieblingshit von 4-minütiger Dauer auf einem alten Radioempfänger. Die zufällig Anzahl von Knacklauten, die der Empfänger während einer Zeit von t Minuten produziert, genüge einer Poissonverteilung $Po(\lambda \cdot t)$ mit dem festen Parameter $\lambda = 0.3$. Die Anzahlen an den drei Tagen seien stochastisch unabhängig voneinander.

a) Welche Verteilung besitzt die zufällige Anzahl von Knacklauten während einer Übertragung des Lieblingshits?

b) Mit welcher Wahrscheinlichkeit p_1 wird während einer 4-minütigen Übertragung des Lieblingshits eines Hörers kein Knacken das Vergnügen beeinträchtigen?

c) Bestimmen Sie die Wahrscheinlichkeit p_2, dass der Radiobastler an allen drei Tagen zusammen mindestens 3 Knacklaute während seiner Lieblingshits hört.

d) Berechnen Sie die Wahrscheinlichkeit p_3, dass der Radiobastler an keinem der drei Tage seinen Lieblingshit ohne Knacklaut hören kann.

Lösung:

a) Sei X_1 bzw. X_2, X_3 die zufällige Anzahl von Knacklauten während des Lieblingshits am ersten, bzw. zweiten und dritten Tag. Da der Lieblingshit 4 Minuten dauert, gilt

$$X_1 \sim Po(1.2).$$

(Analog gilt $X_2 \sim Po(1.2)$ und $X_3 \sim Po(1.2)$.)

b) Gesucht ist $p_1 = \mathbb{P}(X_1 = 0)$. Wegen a) gilt

$$p_1 = \mathbb{P}(X_1 = 0) = e^{-1.2} = 0.3012$$

c) Nach Voraussetzung sind X_1, X_2 und X_3 stochastisch unabhängig und daher gilt nach der Faltungsformel für die Gesamtzahl der Knacklaute

$$X = X_1 + X_2 + X_3 \sim Po(1.2 + 1.2 + 1.2) = Po(3.6)$$

Gesucht ist $p_2 = \mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X \leq 2)$. Da X die Zähldichte

$$f_X(k) = e^{-3.6} \cdot \frac{3.6^k}{k!}, k = 0, 1, \ldots$$

besitzt, gilt

$$\begin{aligned} \mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = f_X(0) + f_X(1) + f_X(2) \\ &= e^{-3.6} + e^{-3.6} \cdot \frac{3.6}{1} + e^{-3.6} \cdot \frac{3.6^2}{2} = 0.3027 \end{aligned}$$

und damit $p_2 = \mathbb{P}(X \geq 3) = 1 - 0.3027 = 0.6973$.

d) Jetzt ist $p_3 = \mathbb{P}(X_1 \geq 1, X_2 \geq 1, X_3 \geq 1)$ gesucht. Wegen der Unabhängigkeit von X_1, X_2 und X_3 gilt

$$\begin{aligned} p_3 &= \mathbb{P}(X_1 \geq 1, X_2 \geq 1, X_3 \geq 1) = \mathbb{P}(X_1 \geq 1) \cdot \mathbb{P}(X_2 \geq 1) \cdot \mathbb{P}(X_3 \geq 1) \\ &= \mathbb{P}(X_1 \geq 1)^3 = (1 - \mathbb{P}(X_1 = 0))^3 \overset{b)}{=} (1 - 0.3012)^3 = 0.3412 \end{aligned}$$

da X_1, X_2 und X_3 alle dieselbe Verteilung $Po(1.2)$ besitzen.

Aufgabe B5

Ein Merkmal besitze eine logarithmische Normalverteilung $\mathcal{LN}(\vartheta, 9)$ mit der Dichte

$$f_\vartheta(x) = \begin{cases} 0 & , x \leq 0 \\ \frac{1}{3 \cdot x \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{18}(\ln(x)-\vartheta)^2} & , x > 0, \end{cases}$$

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 $\mathcal{LN}(\vartheta, 9)$. 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

$$\ln(f_\vartheta(x)) = -\ln(3 \cdot x \cdot \sqrt{2\pi}) - \frac{1}{18}(\ln(x) - \vartheta)^2$$

ergibt sich die Loglikelihood-Funktion zu

$$\begin{aligned} M_x(\vartheta) &= \sum_{i=1}^n \ln(f_\vartheta(x_i)) = \sum_{i=1}^n \left(-\ln(3 \cdot x_i \cdot \sqrt{2\pi}) - \frac{1}{18}(\ln(x_i) - \vartheta)^2 \right) \\ &= -\sum_{i=1}^n \ln(3 \cdot x_i \cdot \sqrt{2\pi}) - \frac{1}{18} \sum_{i=1}^n (\ln(x_i) - \vartheta)^2 \end{aligned}$$

b) Ihre Ableitung ist

$$M_x'(\vartheta) = -\frac{1}{18} \sum_{i=1}^n -2 \cdot (\ln(x_i) - \vartheta) = \frac{1}{9} \sum_{i=1}^n (\ln(x_i - \vartheta) = \frac{1}{9} \left(\sum_{i=1}^n \ln(x_i) - n \cdot \vartheta \right)$$

c) Da die zweite Ableitung

$$M_x''(\vartheta) = -\frac{n}{9} < 0.$$

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 $\sigma^2 = 9$. Also besitzt Z die Verteilung $\mathcal{N}(\vartheta, 9)$.

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 $\mathcal{LN}(\vartheta, 9)$ besitzt, gilt nach der Tabelle auf S. 124 im Skriptum $\mathbb{E}_\vartheta(X_1) = e^{\vartheta+1/2\cdot 9} = e^{\vartheta+4.5}$ und damit

$$\begin{aligned} e^{\vartheta+4.5} &= \bar{x} \\ \vartheta + 4.5 &= \ln(\bar{x}) \\ \vartheta &= \ln(\bar{x}) - 4.5 \end{aligned}$$

Der gesuchte Momentenschätzer ist also $\hat{\vartheta}_m(x) := \ln(\bar{x}) - 4.5$.