INSTITUT FÜR STOCHASTIK
UNIVERSITÄT KARLSRUHE
Dr. B. Klar

WS 2007/08

Klausur

Wahrscheinlichkeitstheorie und Statistik für Studierende der Informatik

vom 05.03.2008

Musterlösungen

A

Aufgabe A1

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	1.1	2.1	3.1	4	4.9	6.1	7	8	9.1	9.8
y_j	11.1	9.5	9.3	5.4	4.7	2.3	1.4	1.5	−4.2	−3.3

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

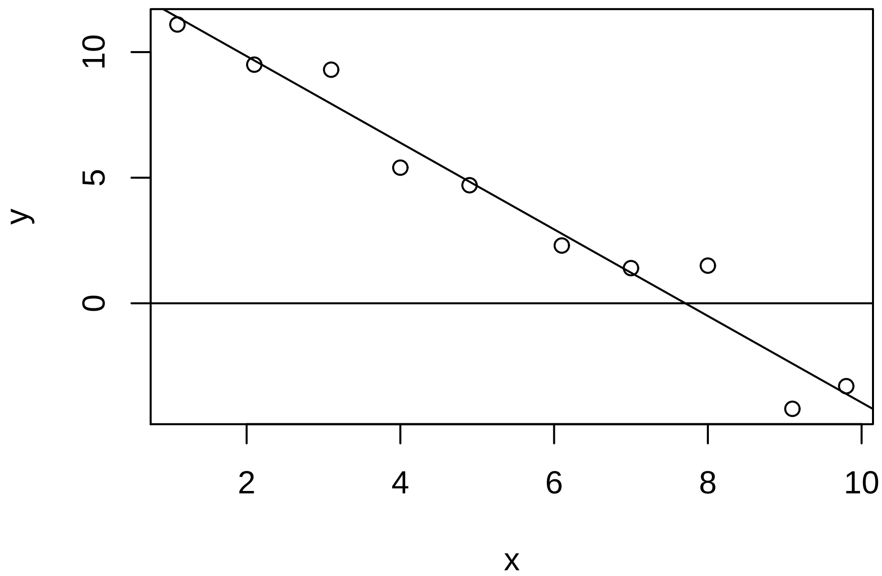
$$\sum_{j=1}^{n}(x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 5.52 \qquad s_x = 2.974$$
$$\bar{y} = 3.77 \qquad s_y = 5.24$$
$$r_{xy} = -0.9784$$

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

$$\begin{aligned} b^* &= -1.724 \\ a^* &= 13.29 \end{aligned}$$

und die Regressionsgerade $y = 13.29 - 1.724 \cdot x$.



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

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

$$y_{()} = (-4.2,\ -3.3,\ 1.4,\ 1.5,\ 2.3,\ 4.7,\ 5.4,\ 9.3,\ 9.5,\ 11.1)$$

c) Berechnen Sie das 0.1-getrimmte Stichprobenmittel $\bar{y}_{0.1}$ von $(y_1, \ldots, y_{10})$.
Lösung: Mit $k = [10 \cdot 0.1] = 1$ ergibt sich

$$\bar{y}_{0.1} = \frac{1}{10 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(9)}) = 3.85$$

d) Bestimmen Sie das Stichproben-0.2-Quantil $\tilde{y}_{0.2}$ von $(y_1, \ldots, y_{10})$.
Lösung: Da $10 \cdot 0.2 = 2$ ganzzahlig ist, ist mit $k = [2] = 2$

$$\tilde{y}_{0.2} = \frac{y_{(k)} + y_{(k+1)}}{2} = \frac{y_{(2)} + y_{(3)}}{2} = -0.95$$

e) 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 $k_2 = [7.5] = 7$

$$\begin{aligned} \tilde{y}_{0.25} &= y_{(k_1+1)} = y_{(3)} = 1.4 \\ \tilde{y}_{0.75} &= y_{(k_2+1)} = y_{(8)} = 9.3 \end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 7.9$.

Aufgabe A2

Die Zufallsvariable X besitze die Verteilung $\mathcal{N}(1,1)$. Weiter sei $Y := 1 - 3X$.

a) Berechnen Sie den Erwartungswert $\mathbb{E}Y$ und die Varianz $V(Y)$.
Lösung: Mit $\mathbb{E}X = 1$ und $V(X) = 1$ folgt

$$\begin{aligned} \mathbb{E}Y &= 1 - 3\mathbb{E}X = -2, \\ V(Y) &= (-3)^2 V(X) = 9. \end{aligned}$$

b) Welche Verteilung besitzt Y?
Lösung: Y ist $\mathcal{N}(-2,9)$-verteilt.

c) Bestimmen Sie die Wahrscheinlichkeit $\mathbb{P}(-2 \leq Y \leq 4)$.
Lösung: Es ist

$$\begin{aligned} \mathbb{P}(-2 \leq Y \leq 4) &= \mathbb{P}(Y \leq 4) - \mathbb{P}(Y \leq -2) \\ &= \Phi\left(\frac{4-(-2)}{3}\right) - \Phi\left(\frac{-2-(-2)}{3}\right) = \Phi(2) - \Phi(0). \end{aligned}$$

Dabei ist Φ die Verteilungsfunktion von $\mathcal{N}(0,1)$. Wegen $\Phi(2) = 0.9772$ und $\Phi(0) = 0.5$ (aus Tabelle A.1) folgt

$$\mathbb{P}(-1 < Y \leq 3) = 0.9772 - 0.5 = 0.4772.$$

d) Bestimmen Sie das 0.1-Quantil $q_{0.1}$ der Zufallsvariablen Y.
Lösung: Die Verteilungsfunktion von Y sei mit F_Y bezeichnet. Dann ist $q_{0.1}$ die Lösung q der Gleichung

$$F_Y(q) = \Phi\left(\frac{q-(-2)}{3}\right) = 0.1.$$

Wegen $\Phi(1.28) = 0.9$ und $\Phi(-x) = 1 - \Phi(x)$ gilt $\Phi(-1.28) = 0.1$ und somit

$$\frac{q+2}{3} = -1.28$$

Hieraus folgt $q_{0.1} = q = 3 \cdot (-1.28) - 2 = -5.84$.

e) Bestimmen Sie die Kovarianz $C(X,Y)$ und den Korrelationskoeffizienten $\rho(X,Y)$.
Lösung: Es gilt

$$\begin{aligned} C(X,Y) &= C(X, 1-3X) = C(X,-3X) = -3\,C(X,X) = -3\,V(X) = -3, \\ \rho(X,Y) &= C(X,Y)/\sqrt{V(X)V(Y)} = (-3)/\sqrt{1 \cdot 9} = -1. \end{aligned}$$

Aufgabe A3

X sei eine Zufallsvariable mit Werten in $\{1, 2\}$ und Y eine Zufallsvariable mit Werten in $\{0, 1, c\}$. Hierbei sei c eine feste Zahl aus $\{2, 3, 4, \ldots\}$. Die folgende Tabelle gibt die gemeinsame Verteilung $\mathbb{P}(X = i, Y = j)$ des Zufallsvektors (X, Y) für die Werte $i = 1, 2$ und $j = 0, 1, c$ an.

	$j = 0$	$j = 1$	$j = c$
$i = 1$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
$i = 2$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$

a) Berechnen Sie die Randverteilung von Y und den Erwartungswert $\mathbb{E}Y$. Für welche c gilt $\mathbb{E}Y = 4$?
Lösung:

$$\mathbb{P}(Y = 0) = \frac{3}{8}, \quad \mathbb{P}(Y = 1) = \frac{2}{8}, \quad \mathbb{P}(Y = c) = \frac{3}{8}$$

$$\mathbb{E}Y = 0 \cdot \frac{3}{8} + 1 \cdot \frac{2}{8} + c \cdot \frac{3}{8} = (3c + 2)/8$$

Für $c = 10$ ist $\mathbb{E}X = 4$.

b) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(X = 2, Y > 0)$.
Lösung:

$$\mathbb{P}(X = 2, Y > 0) = \mathbb{P}(X = 2, Y = 1) + \mathbb{P}(X = 2, Y = c) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}$$

c) Berechnen Sie die bedingte Wahrscheinlichkeit $\mathbb{P}(X = 2 | Y > 0)$.
Lösung:

$$\begin{aligned} \mathbb{P}(X = 2 | Y > 0) &= \frac{\mathbb{P}(X = 2, Y > 0)}{\mathbb{P}(Y > 0)} = \frac{\mathbb{P}(X = 2, Y > 0)}{\mathbb{P}(Y = 1) + \mathbb{P}(Y = c)} \\ &= \frac{\frac{1}{8} + \frac{2}{8}}{\frac{2}{8} + \frac{3}{8}} = \frac{3}{5} \end{aligned}$$

d) Es sei $c = 2$ sowie $Z := X \cdot Y$. Berechnen Sie den Erwartungswert $\mathbb{E}Z$, das zweite Moment $\mathbb{E}Z^2$ und die Varianz $V(Z)$.
Lösung: Es sei $f(i, j) := \mathbb{P}(X = i, Y = j)$. Dann gilt

$$\begin{aligned} \mathbb{E}Z &= \mathbb{E}[X \cdot Y] = \sum_{i,j:\ f(i,j)>0} i \cdot j \cdot f(i, j) = 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 4 \cdot \frac{1}{4} = \frac{13}{8}, \\ \mathbb{E}Z^2 &= \mathbb{E}[X^2 \cdot Y^2] = \sum_{i,j:\ f(i,j)>0} i^2 \cdot j^2 \cdot f(i, j) = 1 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} + 16 \cdot \frac{1}{4} \\ &= \frac{41}{8}, \\ V(Z) &= \mathbb{E}Z^2 - (\mathbb{E}Z)^2 = \frac{41}{8} - \frac{169}{64} = \frac{159}{64}. \end{aligned}$$

Aufgabe A4

X und Y seien die zufälligen Wartezeiten [in Minuten] von zwei Kunden A und B, die an unterschiedlichen Kassen stehen. Wir nehmen an, dass X und Y stochastisch unabhängig und jeweils exponentialverteilt mit Parameter 0.5 sind.

a) Mit welcher Wahrscheinlichkeit wartet Kunde A länger als 4 Minuten?
Lösung: Die Verteilungsfunktion der Wartezeit X von Kunde A ist $F_X(t) = 1 - \exp(-0.5t)$ für $t > 0$. Somit gilt:

$$\mathbb{P}(X > 4) = 1 - F_X(4) \;=\; e^{-2} \;=\; 0.135.$$

b) Wie groß ist die Wahrscheinlichkeit, dass beide Kunden länger als 4 Minuten warten müssen?
Lösung: Die Unabhängigkeit von X, Y liefert

$$\mathbb{P}(X > 4, Y > 4) = \mathbb{P}(X > 4) \cdot \mathbb{P}(Y > 4) \;=\; e^{-4} \;=\; 0.018.$$

c) Wie groß ist die Wahrscheinlichkeit, dass mindestens eine Kunde länger als 4 Minuten warten muss?
Lösung: Es ist

$$\begin{aligned}\mathbb{P}(\max(X,Y) > 4) &= 1 - \mathbb{P}(X \le 4, Y \le 4) \;=\; 1 - \mathbb{P}(X \le 4) \cdot \mathbb{P}(Y \le 4) \\ &= 1 - \left(1 - e^{-2}\right)^2 \;=\; e^{-2}\left(2 - e^{-2}\right) \;=\; 0.252.\end{aligned}$$

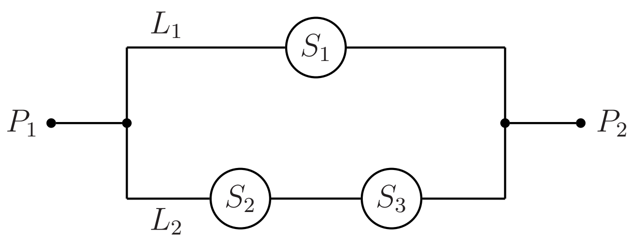
d) Geben Sie die gemeinsame Dichte von X und Y an.
Lösung: Die Dichte von X bzw. Y ist $f_X(t) = f_Y(t) = 0.5\exp(-0.5t)$ für $t > 0$. Wegen der Unabhängigkeit von X und Y ist

$$f(x,y) \;=\; f_X(x) \cdot f_Y(y) \;=\; \begin{cases} 0.25e^{-0.5(x+y)}, & x, y > 0 \\ 0, & \text{sonst} \end{cases}$$

(eine) gemeinsame Dichte von X und Y.

Aufgabe A5

Zwischen zwei Punkten P_1 und P_2 verläuft folgendes Leitungsnetz:



Dabei sind S_1, S_2, S_3 störanfällige Stellen. Leitung L_1 läuft durch die störanfällige Stelle S_1, Leitung L_2 läuft durch die störanfälligen Stellen S_2 und S_3. Die Zufallsvariablen X_i seien definiert als

$$X_i = \begin{cases} 0, & S_i \text{ ist unterbrochen,} \\ 1, & S_i \text{ ist nicht unterbrochen,} \end{cases} \quad i = 1, 2, 3.$$

Ferner seien X_1, X_2, X_3 stochastisch unabhängig mit $\mathbb{P}(X_i = 1) = q$, $i = 1, 2, 3$, für ein q mit $0 < q < 1$.
A sei das Ereignis, dass Leitung L_1 frei ist, d.h. dass S_1 nicht unterbrochen ist. B sei das Ereignis, dass Leitung L_2 frei ist.

a) Wie groß ist $\mathbb{P}(B)$, also die Wahrscheinlichkeit, dass L_2 frei ist?
Lösung: Wegen der Unabhängigkeit von X_2 und X_3 gilt

$$\mathbb{P}(B) \;=\; \mathbb{P}(\{X_2 = 1\} \cap \{X_3 = 1\}) \;=\; \mathbb{P}(X_2 = 1) \cdot \mathbb{P}(X_3 = 1) \;=\; q^2.$$

b) Wie groß ist die Wahrscheinlichkeit, dass L_2 frei ist, unter der Bedingung, dass L_1 frei ist?
Lösung: Wegen der Unabhängigkeit von X_1, X_2 und X_3 sind auch A und B unabhängig, und es gilt

$$\mathbb{P}(B|A) = \mathbb{P}(B) \;=\; q^2.$$

c) Wie groß ist die Wahrscheinlichkeit, dass L_2 frei ist, aber L_1 nicht?
Lösung:

$$\mathbb{P}(A^c \cap B) \;=\; \mathbb{P}(A^c) \cdot \mathbb{P}(B) \;=\; (1 - q)q^2.$$

d) Wie groß ist die Wahrscheinlichkeit, dass genau eine der beiden Leitungen frei ist?
Lösung: Mit $\mathbb{P}(A \cap B^c) = \mathbb{P}(A) \cdot \mathbb{P}(B^c) = q(1 - q^2)$ folgt

$$\mathbb{P}((A \cap B^c) \cup (A^c \cap B)) \;=\; \mathbb{P}(A \cap B^c) + \mathbb{P}(A^c \cap B) \;=\; q(1-q^2) + (1-q)q^2 \;=\; q(1-q)(1+2q).$$

e) C sei das Ereignis {„P_1 ist mit P_2 verbunden“}. Bestimmen Sie $\mathbb{P}(C)$.
Lösung: Mit $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B) = q \cdot q^2 = q^3$ folgt

$$\mathbb{P}(C) \;=\; \mathbb{P}(A \cup B) \;=\; \mathbb{P}(A) + \mathbb{P}(A) - \mathbb{P}(A \cap B) \;=\; q + q^2 - q^3.$$

f) Die Zufallsvariable $Y := \min\{X_2, X_3\}$ besitzt die Binomial-Verteilung $Bin(n, p)$. Bestimmen Sie die Parameter n und p.
Lösung: Die Zufallsvariable Y nimmt die Werte 0 und 1 an. Wegen

$$\mathbb{P}(Y = 1) \;=\; \mathbb{P}(\min\{X_2, X_3\} = 1) \;=\; \mathbb{P}(X_2 = 1, X_3 = 1) \;=\; \mathbb{P}(X_2 = 1) \cdot \mathbb{P}(X_3 = 1) \;=\; q^2$$

gilt $\mathbb{P}(Y = 0) = 1 - q^2$. Somit ist Y $Bin(1, q^2)$-verteilt.

Aufgabe A6

Es soll der unbekannte Parameter $\vartheta \in \Theta = (0, \infty)$ für die Verteilung mit der Dichte

$$f(x;\vartheta) = \frac{1}{\sqrt{2\pi\vartheta}} \exp\left(-\frac{2x^2}{\vartheta}\right), \qquad x \in \mathbb{R}.$$

bestimmt werden.

a) Geben Sie die zur Stichprobe $x = (x_1, \ldots, x_n)$ gehörende Likelihood-Funktion $L_x(\vartheta)$ an.
Lösung: Es ist

$$L_x(\vartheta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\vartheta}} \exp\left(-\frac{2x_i^2}{\vartheta}\right) = \left(\frac{1}{2\pi\vartheta}\right)^{n/2} \exp\left(-\frac{2}{\vartheta}\sum_{i=1}^{n} x_i^2\right).$$

b) Berechnen Sie die Loglikelihood-Funktion $M_x(\vartheta) := \log L_x(\vartheta)$.
Lösung: Folglich ist

$$M_x(\vartheta) = \log L_x(\vartheta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\vartheta - \frac{2}{\vartheta}\sum_{i=1}^{n} x_i^2$$

c) Berechnen Sie den Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ.
Lösung: Für die Ableitung von M_x nach ϑ gilt:

$$\frac{d}{d\vartheta} M_x(\vartheta) = -\frac{n}{2\vartheta} + \frac{2}{\vartheta^2}\sum_{i=1}^{n} x_i^2 = 0 \quad \Leftrightarrow \quad \vartheta = \frac{4}{n}\sum_{i=1}^{n} x_i^2.$$

Weiter gilt mit $\hat{\vartheta}_n := \frac{4}{n}\sum_{i=1}^{n} x_i^2$

$$\frac{d^2}{d\vartheta^2} M_x(\hat{\vartheta}_n) = \frac{n}{2\hat{\vartheta}_n^2} - \frac{n}{\hat{\vartheta}_n^2} < 0.$$

Somit liegt eine lokale Maximalstelle vor, und

$$\hat{\vartheta}_n(x) = \frac{4}{n}\sum_{i=1}^{n} x_i^2$$

ist der gesuchte Maximum-Likelihood-Schätzer für ϑ.