INSTITUT FÜR MATHEMATISCHE STOCHASTIK WS 2003/04
UNIVERSITÄT KARLSRUHE
Dr. B. Klar

Klausur

Wahrscheinlichkeitstheorie und Statistik für Informatiker

vom 26.2.2004

Musterlösungen

B

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	1.1	1.6	2.9	3.9	5.2	6.1	7.1	7.6	8.9	9.8
y_j	14.2	14.3	1.2	7.4	5.6	3.9	-0.9	1	-8.9	-7.5

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

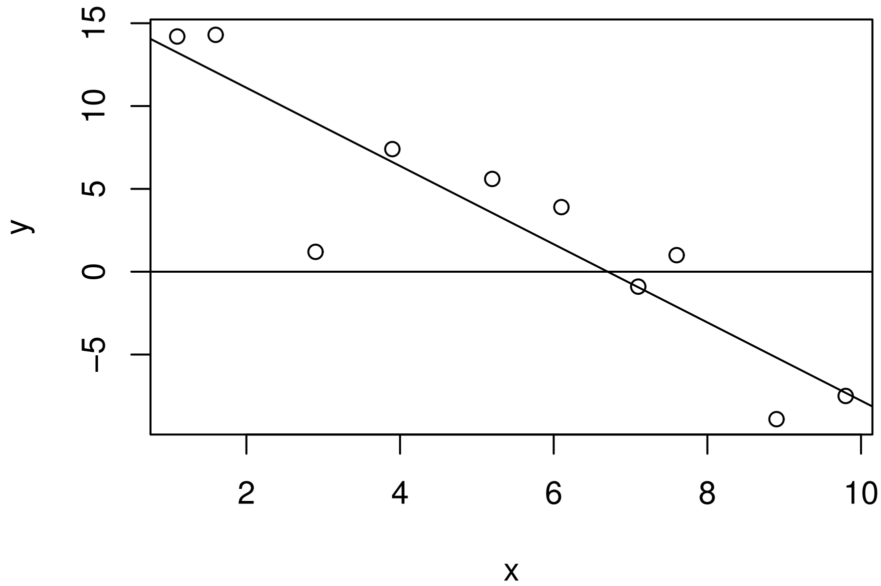
$$\sum_{j=1}^{n}(x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\begin{aligned} \bar{x} &= 5.42 & s_x &= 3.005 \\ \bar{y} &= 3.03 & s_y &= 7.848 \\ r_{xy} &= -0.9043 \end{aligned}$$

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

$$\begin{aligned} b^* &= -2.361 \\ a^* &= 15.83 \end{aligned}$$

und die Regressionsgerade $y = 15.83 - 2.361 \cdot x$.



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

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

$$y_{()} = (-8.9,\ -7.5,\ -0.9,\ 1,\ 1.2,\ 3.9,\ 5.6,\ 7.4,\ 14.2,\ 14.3)$$

c) Berechnen Sie das 0.15-getrimmte Stichprobenmittel $\bar{y}_{0.15}$ von $(y_1, \ldots, y_{10})$.
Lösung: Mit $k = [10 \cdot 0.15] = 1$ ergibt sich

$$\bar{y}_{0.15} = \frac{1}{10 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(9)}) = 3.112$$

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 $k_2 = [7.5] = 7$

$$\begin{aligned} \tilde{y}_{0.25} &= y_{(k_1+1)} = y_{(3)} = -0.9 \\ \tilde{y}_{0.75} &= y_{(k_2+1)} = y_{(8)} = 7.4 \end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 8.3$.

Aufgabe B2

X, Y seien zwei Zufallsvariablen mit Werten in $\{0, 1, c\}$, $c \in \{2, 3, 4, \ldots\}$, bzw. $\{0, 1, 2\}$. Die folgende Tabelle gibt die gemeinsame Verteilung $\mathbb{P}(X = i, Y = j)$ des Zufallsvektors (X, Y) für die Werte $(i, j) \in \{0, 1, c\} \times \{0, 1, 2\}$ an.

$Y \backslash X$	$i = 0$	$i = 1$	$i = c$
$j = 0$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{12}$
$j = 1$	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
$j = 2$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

a) Berechnen Sie die Randverteilung von X und den Erwartungswert $\mathbb{E}X$. Für welche c gilt $\mathbb{E}X = \frac{5}{4}$?
Lösung:

$$\mathbb{P}(X = 0) = \frac{1}{4}, \quad \mathbb{P}(X = 1) = \frac{1}{2}, \quad \mathbb{P}(X = c) = \frac{1}{4}$$
$$\mathbb{E}X = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + c \cdot \frac{1}{4} = \frac{1}{4}c + \frac{1}{2}$$

Für $c = 3$ ist $\mathbb{E}X = 5/4$.

b) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(X > 0, Y = 1)$.
Lösung:

$$\mathbb{P}(X > 0, Y = 1) = \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = c, Y = 1) = \frac{1}{6} + \frac{1}{12} = \frac{1}{4}$$

c) Berechnen Sie die bedingte Wahrscheinlichkeit $\mathbb{P}(Y = 1 | X > 0)$.
Lösung:

$$\begin{aligned} \mathbb{P}(Y = 1 | X > 0) &= \frac{\mathbb{P}(X > 0, Y = 1)}{\mathbb{P}(X > 0)} = \frac{\mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = c, Y = 1)}{\mathbb{P}(X = 1) + \mathbb{P}(X = c)} \\ &= \frac{\frac{1}{6} + \frac{1}{12}}{\frac{1}{2} + \frac{1}{4}} = \frac{1}{3} \end{aligned}$$

d) Es sei $c = 3$ 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

$$\begin{aligned} \mathbb{E}Z &= \mathbb{E}[X \cdot Y] = \sum_{i,j:\, f(i,j)>0} i \cdot j \cdot f(i, j) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{12} + 3 \cdot \frac{1}{12} + 6 \cdot \frac{1}{12} = \frac{13}{12}, \\ \mathbb{E}Z^2 &= \mathbb{E}[X^2 \cdot Y^2] = \sum_{i,j:\, f(i,j)>0} i^2 \cdot j^2 \cdot f(i, j) = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{12} + 9 \cdot \frac{1}{12} + 36 \cdot \frac{1}{12} \\ &= \frac{17}{4}, \\ V(Z) &= \mathbb{E}Z^2 - (\mathbb{E}Z)^2 = \frac{17}{4} - \frac{169}{144} = \frac{443}{144}. \end{aligned}$$

Aufgabe B3

Es sei X eine standardnormal-verteilte Zufallsvariable, kurz $X \sim \mathcal{N}(0,1)$. Weiter sei $Y := 3X - 1$.

a) Berechnen Sie den Erwartungswert $\mathbb{E}Y$ und die Varianz $V(Y)$.
Lösung: Mit $\mathbb{E}X = 0$ und $V(X) = 1$ folgt

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

b) Bestimmen Sie die Wahrscheinlichkeit $\mathbb{P}(-1 < Y \leq 2)$.
Lösung: Mit der Definition von Y und 8.11 Satz folgt

$$\mathbb{P}(-1 < Y \leq 2) \;=\; \mathbb{P}(-1 < 3X - 1 \leq 2) \;=\; \mathbb{P}(0 < X \leq 1) \;=\; \Phi(1) - \Phi(0).$$

Dabei ist Φ die Verteilungsfunktion von X. Wegen $\Phi(2) = 0.8413$ und $\Phi(0) = 0.5$ (aus Tabelle A.1) folgt

$$\mathbb{P}(-1 < Y \leq 2) = 0.3413.$$

c) Bestimmen Sie das 0.975-Quantil $q_{0.975}$ der Zufallsvariablen Y.
Lösung: Die Verteilungsfunktion von Y sei mit F_Y bezeichnet. Dann ist $q_{0.975}$ die Lösung q der Gleichung

$$F_Y(q) \;=\; \Phi\left(\frac{q-(-1)}{3}\right) = 0.975.$$

Wegen $\Phi(1.96) = 0.975$ (aus Tabelle A.1) gilt also

$$\frac{q+1}{3} = 1.96$$

und damit

$$q_{0.975} = q = 3 \cdot 1.96 - 1 = 4.88. \tag{1}$$

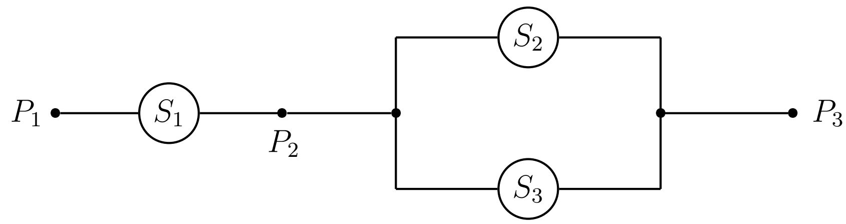
Der Zusammenhang (1) kann auch direkt aus Bemerkung 12.20 c) abgelesen werden.

d) Bestimmen Sie die Kovarianz $C(X,Y)$.
Lösung: Wegen $\mathbb{E}X = 0$ und $\mathbb{E}X^2 \; (= V(X) + (\mathbb{E}X)^2) \; = 1$ gilt

$$\begin{aligned} C(X,Y) &= \mathbb{E}[X \cdot Y] - \mathbb{E}X \cdot \mathbb{E}Y \;=\; \mathbb{E}[X \cdot (3X-1)] - 0 \cdot (-1) \;=\; \mathbb{E}[3X^2 - X] \\ &= 3\mathbb{E}X^2 - \mathbb{E}X \;=\; 3. \end{aligned}$$

Aufgabe B4

Zwischen 3 Punkten P_1, P_2 und P_3 verläuft folgendes Leitungsnetz:



Dabei sind S_1, S_2, S_3 störanfällige Stellen. 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 $\mathbb{P}(X_i = 1) = p$, $i = 1, 2, 3$, für ein $0 < p < 1$.

a) Stellen Sie das Ereignis $A := \{„P_2 \text{ ist mit } P_3 \text{ verbunden“}\}$ mit Hilfe der Zufallsvariablen X_2 und X_3 dar.
Lösung:

$$A = \{X_2 = 1\} \cup \{X_3 = 1\} = \{\max\{X_2, X_3\} = 1\} = \{X_2 + X_3 \geq 1\}$$

b) Stellen Sie das Ereignis $B := \{„P_1 \text{ ist mit } P_3 \text{ verbunden“}\}$ mit Hilfe der Zufallsvariablen X_1 und des Ereignisses A dar.
Lösung: $B = \{X_1 = 1\} \cap A$

c) Bestimmen Sie die Wahrscheinlichkeit $\mathbb{P}(B)$.
Lösung: Zunächst erhält man mit dem Additionssatz und der Unabhängigkeit von X_2 und X_3

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(\{X_2 = 1\} \cup \{X_3 = 1\}) \\ &= \mathbb{P}(X_2 = 1) + \mathbb{P}(X_3 = 1) - \mathbb{P}(\{X_2 = 1\} \cap \{X_3 = 1\}) \\ &= p + p - \mathbb{P}(X_2 = 1) \cdot \mathbb{P}(X_3 = 1) = 2p - p^2 = p(2 - p). \end{aligned}$$

Wegen der Unabhängigkeit von $\{X_1 = 1\}$ und A (Blockungslemma) gilt

$$\mathbb{P}(B) = \mathbb{P}(X_1 = 1) \cdot \mathbb{P}(A) = p \cdot p(2 - p) = p^2(2 - p).$$

d) Die Zufallsvariable $Y := X_1 + X_2$ besitzt die Binomial-Verteilung $Bin(n, q)$. Bestimmen Sie die Parameter n und q.
Lösung: Die Zufallsvariable Y ist die Faltung zweier $Bin(1, p)$-verteilten Zufallsvariablen. Mit Tabelle S.114 erhält man damit $n = 2$ und $q = p$.

e) Bestimmen Sie die erzeugende Funktion $g_Y(s)$, $|s| \le 1$, von Y.
Lösung: Nach 13.1 Definition gilt

$$\begin{aligned} g_{X_1}(s) &= \sum_{k=0}^{1} \mathbb{P}(X_1 = k)s^k = \mathbb{P}(X_1 = 0)s^0 + \mathbb{P}(X_1 = 1)s^1 \\ &= (1-p) + p \cdot s \end{aligned}$$

für $|s| \le 1$. Weiter ist $g_{X_1} = g_{X_2}$. Nach 13.3c) folgt

$$g_Y(s) = g_{X_1}(s) \cdot g_{X_2}(s) = ((1-p) + p \cdot s)^2$$

für $|s| \le 1$.
Da $Y \sim Bin(2, p)$ ist, siehe auch 13.2 Beispiel a).

Aufgabe B5

Ein Merkmal habe die Dichte

$$t \mapsto f_\vartheta(t) := \begin{cases} \dfrac{1}{\sqrt{\pi}\, t} \exp\left(-(\log(t) - \vartheta)^2\right), & \text{falls } t > 0, \\ 0 \qquad , & \text{falls } t \le 0, \end{cases}$$

wobei $\vartheta \in \mathbb{R}$ ein unbekannter Parameter ist. Der Parameter ϑ soll aufgrund einer unabhängigen Stichprobe $x = (x_1, \ldots, x_n)$ geschätzt werden, wobei $x_1 > 0, \ldots, x_n > 0$ sind. Der Maximum-Likelihood-Schätzer $\hat{\vartheta}(x_1, \ldots, x_n)$ für ϑ ist von der Form

$$\hat{\vartheta}(x_1, \ldots, x_n) = c_1 \cdot \sum_{j=1}^{n} \log(c_2 \cdot x_j - c_3)$$

mit gewissen Konstanten c_1, c_2, c_3. Bestimmen Sie die Konstanten c_1, c_2 und c_3.
Lösung: Es ist

$$\log(f_\vartheta(t)) = -\log(\sqrt{\pi}) - \log(t) - (\log(t) - \vartheta)^2, \quad t > 0,$$

also

$$\begin{aligned} M_x(\vartheta) &= \sum_{j=1}^{n} \log(f_\vartheta(x_j)) = \sum_{j=1}^{n} \left(-\log(\sqrt{\pi}) - \log(x_j) - (\log(x_j) - \vartheta)^2\right) \\ &= -n\log(\sqrt{\pi}) - \sum_{j=1}^{n} \log(x_j) - \sum_{j=1}^{n} (\log(x_j) - \vartheta)^2, \\ M_x'(\vartheta) &= 2\sum_{j=1}^{n} (\log(x_j) - \vartheta) = 2\sum_{j=1}^{n} \log(x_j) - 2n\vartheta, \\ M_x''(\vartheta) &= -2n < 0. \end{aligned}$$

Wegen

$$M_x'(\vartheta) = 0 \iff \vartheta = \frac{1}{n}\sum_{j=1}^{n} \log(x_j)$$

ist

$$(x_1, \ldots, x_n) \mapsto \hat{\vartheta}(x) := \frac{1}{n} \sum_{j=1}^{n} \log(x_j)$$

der gesuchte Maximum-Likelihood-Schätzer für ϑ.

Aufgabe B6

Eine Münze mit den Symbolen „Zahl" und „Wappen" soll auf ihre Echtheit überprüft werden, d.h es soll überprüft werden ob,

$$\mathbb{P}(\{\text{„Zahl" geworfen}\}) \;=\; \mathbb{P}(\{\text{„Wappen" geworfen}\}) \;=\; \frac{1}{2}$$

gilt. Dem folgenden Zufallsexperiment werden die Zufallsvariablen

$$X_i \;=\; \begin{cases} 1, & \text{falls „Zahl" geworfen,} \\ 0, & \text{falls „Wappen" geworfen,} \end{cases} \qquad i = 1, \ldots, n,$$

zugrunde gelegt. Dabei seien $X_1, X_2, \ldots, X_n$ stochastisch unabhängig und jeweils $Bin(1, \vartheta)$-verteilt. Der Parameter $0 < \vartheta < 1$ sei unbekannt. Nun wird die Münze $n = 1000$ Mal geworfen und 487 Mal das Symbol „Zahl" beobachtet.

a) Bestimmen Sie das Stichproben-Mittel $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.

Lösung: Mit $\sum_{i=1}^{n} X_i = 487$ erhält man $\bar{X}_n = \frac{487}{1000} = 0.487$.

b) Ein asymptotisches Konfidenzintervall $\mathcal{C} = [l_n^*, L_n^*]$ für ϑ zum (Konfidenz-)Niveau $1 - \alpha$, $0 < \alpha < 1$, ist durch

$$\begin{aligned} l_n^* \;&=\; l_n^*(X_1, \ldots, X_n) \;=\; \bar{X}_n \boxed{} \frac{h}{\sqrt{n}} \sqrt{\bar{X}_n(1 - \bar{X}_n)}\,, \\ L_n^* \;&=\; L_n^*(X_1, \ldots, X_n) \;=\; \bar{X}_n \boxed{} \frac{h}{\sqrt{n}} \sqrt{\bar{X}_n(1 - \bar{X}_n)} \end{aligned}$$

gegeben, wobei $h = \boxed{}$ ist. Vervollständigen Sie die fehlenden Angaben in den Formel für l_n^*, L_n^* und h. Berechnen Sie $\mathcal{C}$ für $\alpha = 0.05$.

Lösung: Mit 18.10 Beispiel und 18.11 Bemerkung 1. folgt

$$\begin{aligned} l_n^* \;&=\; l_n^*(X_1, \ldots, X_n) \;=\; \bar{X}_n - \frac{h}{\sqrt{n}} \sqrt{\bar{X}_n(1 - \bar{X}_n)}\,, \\ L_n^* \;&=\; L_n^*(X_1, \ldots, X_n) \;=\; \bar{X}_n + \frac{h}{\sqrt{n}} \sqrt{\bar{X}_n(1 - \bar{X}_n)}\,, \end{aligned}$$

wobei $h = u_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ ist. Für $\alpha = 0.05$ ist $h = u_{0.975} = 1.96$ (aus Tabelle A.1). Mit $n = 1000$ und $\bar{X}_n = 0.487$ erhält man $l_n^* = 0.456$ und $L_n^* = 0.518$.