Institut für Stochastik
Karlsruher Institut für Technologie (KIT)
Dr. B. Klar

WS 2009/10

Klausur

Wahrscheinlichkeitstheorie und Statistik für Studierende der Informatik

vom 08.02.2010

Musterlösungen

B

Aufgabe B1

Gegeben sei eine Urliste mit den Paaren $(x_1, y_1), \ldots, (x_8, y_8)$

j	1	2	3	4	5	6	7	8
x_j	0.9	2	3	4	4.5	5.8	6.9	8.2
y_j	3.1	3.2	4.9	5.5	3.8	5.3	4.4	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

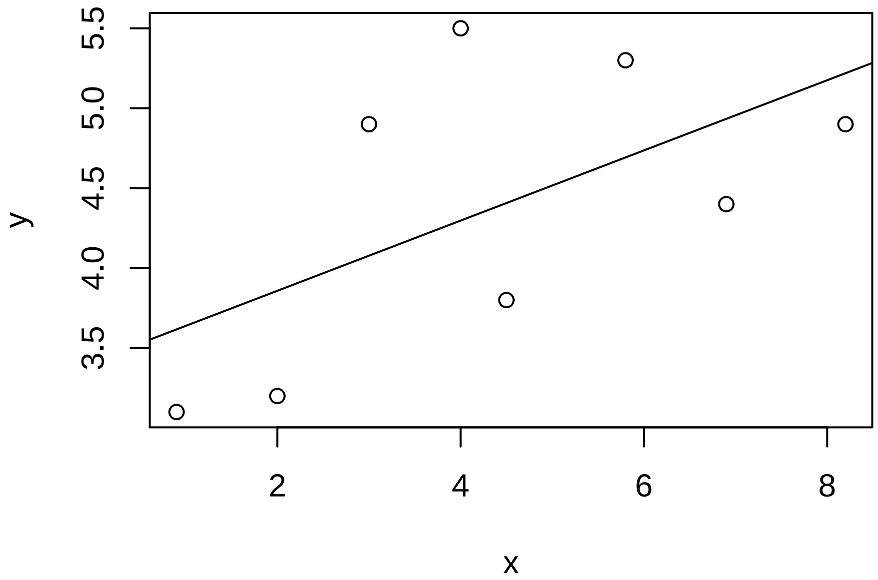
$$\sum_{j=1}^{n} (x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 4.41 \qquad s_x = 2.472$$
$$\bar{y} = 4.39 \qquad s_y = 0.926$$
$$r_{xy} = 0.5857$$

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

$$\begin{aligned} b^* &= 0.219 \\ a^* &= 3.42 \end{aligned}$$

und die Regressionsgerade $y = 3.42 + 0.219 \cdot x$.



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

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

$$y_{()} = (3.1,\ 3.2,\ 3.8,\ 4.4,\ 4.9,\ 4.9,\ 5.3,\ 5.5)$$

c) Berechnen Sie das 0.15-getrimmte Stichprobenmittel $\bar{y}_{0.15}$ von $(y_1, \ldots, y_8)$.
Lösung: Mit $k = [8 \cdot 0.15] = 1$ ergibt sich

$$\bar{y}_{0.15} = \frac{1}{8 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(7)}) = 4.417$$

d) Bestimmen Sie das Stichproben-0.7-Quantil $\tilde{y}_{0.7}$ von $(y_1, \ldots, y_8)$.
Lösung: Da $8 \cdot 0.7 = 5.6$ nicht ganzzahlig ist, ist mit $k = [5.6] = 5$

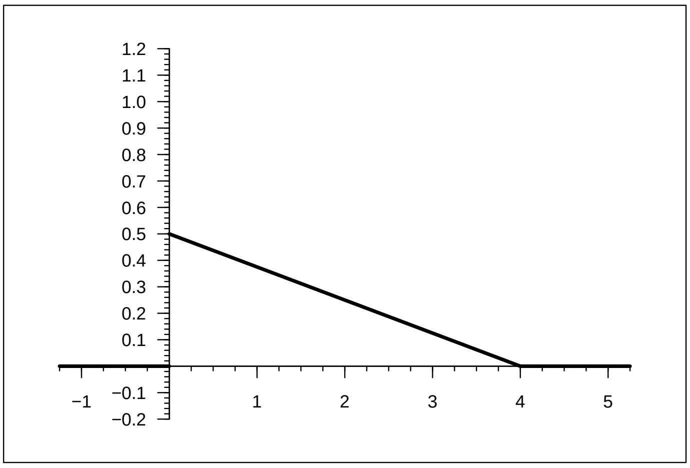
$$\tilde{y}_{0.7} = y_{(k+1)} = y_{(6)} = 4.9$$

Aufgabe B2

Die Dichte f einer stetigen Zufallsvariablen X ist gegeben durch

$$f(x) = \begin{cases} (4-x)/8, & \text{falls } 0 \leq x \leq 4, \\ 0, & \text{sonst.} \end{cases}$$

Das nachstehende Schaubild zeigt f:



a) Bestimmen Sie den Erwartungswert von X.
Lösung: Es gilt

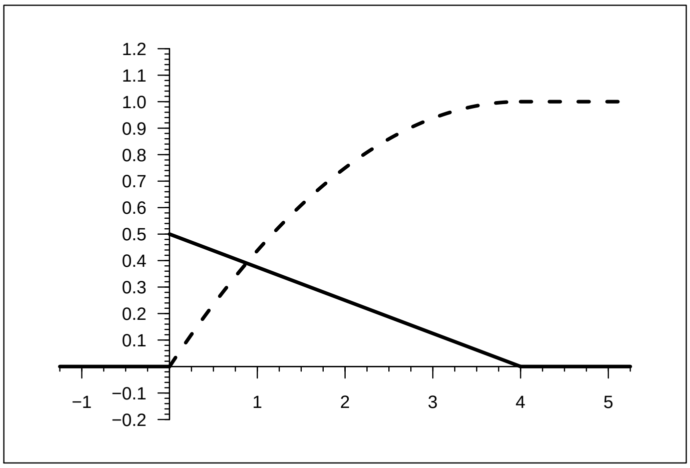
$$\mathbb{E}X = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^4 \left(x/2 - x^2/8\right)\, dx = \left[\frac{x^2}{4} - \frac{x^3}{24}\right]_0^4 = \frac{4}{3}.$$

b) Bestimmen Sie das zweite Moment $\mathbb{E}(X^2)$ und die Varianz $V(X)$ von X.
Lösung: Es ist

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx = \int_0^4 \left(x^2/2 - x^3/8\right)\, dx = \left[\frac{x^3}{6} - \frac{x^4}{32}\right]_0^4 = \frac{8}{3},$$
$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2 = \frac{8}{9}.$$

c) Bestimmen Sie die Verteilungsfunktion F zur Dichte f, und zeichnen Sie den Graphen von F in das Schaubild ein.
Lösung: Es gilt

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 0, & \text{falls } x < 0 \\ \frac{x}{2} - \frac{x^2}{16}, & \text{falls } 0 \leq x \leq 4, \\ 1, & \text{falls } x > 4. \end{cases}$$



d) Bestimmen Sie den Median von F.
Lösung: Es gilt

$$\frac{x}{2} - \frac{x^2}{16} \stackrel{!}{=} \frac{1}{2} \Leftrightarrow x^2 - 8 \cdot x + 8 = 0 \Leftrightarrow x_{1/2} = 4 \pm \sqrt{16-8} = 4 \pm \sqrt{8}.$$

Folglich muss der Median von F gleich $4 - \sqrt{8} = 1.1716$ sein.

Aufgabe B3

In einer Urne befinden sich 3 rote und 2 schwarze Kugeln. Es werden zunächst 2 Kugeln rein zufällig und ohne Zurücklegen gezogen.

Hinweis: Evtl. ist die Einführung einer Zufallsvariablen X, die die zufällige Anzahl der gezogenen roten Kugeln angibt, hilfreich.

a) Mit welcher Wahrscheinlichkeit sind die beiden gezogenen Kugeln rot?
Lösung:

$$\mathbb{P}(\text{beide Kugeln rot}) = \tfrac{3}{5} \cdot \tfrac{2}{4} = \tfrac{3}{10}.$$

Alternativ: Die oben definierte Zufallsvariable X ist $Hyp(2,3,2)$-verteilt. Gesucht ist dann

$$\mathbb{P}(X = 2) = \frac{\binom{3}{2}\cdot\binom{2}{0}}{\binom{5}{2}} = \tfrac{3}{10}.$$

b) Mit welcher Wahrscheinlichkeit haben die beiden gezogenen Kugeln dieselbe Farbe?
Lösung:

$$\mathbb{P}(\text{beide Kugeln gleichfarbig}) = \tfrac{3}{10} + \tfrac{2}{5} \cdot \tfrac{1}{4} = \tfrac{3}{10} + \tfrac{1}{10} = \tfrac{4}{10}.$$

c) Sind die beiden gezogenen Kugeln gleichfarbig, werden zwei rote Kugeln in die Urne gelegt. Haben die beiden gezogenen Kugeln unterschiedliche Farben, werden zwei schwarze Kugeln in die Urne gelegt. Die Zufallsvariable Z bezeichne die Anzahl der roten Kugeln, die danach in der Urne liegen. Bestimmen Sie die Wahrscheinlichkeiten $\mathbb{P}(Z = k)$ für $k = 2, 3, 4, 5$.
Lösung: Die Zufallsvariable Z nimmt die Werte $2, 3, 5$ an: Es gilt $Z = 2$, falls die beiden gezogenen Kugeln unterschiedliche Farben haben, und $Z = 3$ bzw. $Z = 5$, falls die beiden gezogenen Kugeln rot bzw. schwarz sind. Somit gilt

$$\mathbb{P}(Z = 3) = \frac{3}{10},\ \mathbb{P}(Z = 4) = 0,\ \mathbb{P}(Z = 5) = \frac{1}{10},\ \mathbb{P}(Z = 2) = 1 - \frac{4}{10} = \frac{6}{10}.$$

Alternativ: Es gilt

$$\begin{aligned}
\mathbb{P}(Z = 3) &= \mathbb{P}(X = 2) = \frac{3}{10},\\
\mathbb{P}(Z = 5) &= \mathbb{P}(X = 0) = \frac{1}{10},\\
\mathbb{P}(Z = 2) &= \mathbb{P}(X = 1) = 1 - \frac{4}{10} = \frac{6}{10}.
\end{aligned}$$

d) Anschließend wird nach gutem Mischen rein zufällig eine Kugel gezogen. Wie groß ist die Wahrscheinlichkeit, dass diese Kugel rot ist?
Lösung: Mit dem Satz über die totale Wahrscheinlichkeit erhält man

$$\begin{aligned}
\mathbb{P}(\text{es wird eine rote Kugel gezogen}) =& \mathbb{P}(\text{Kugel rot} \mid Z = 2) \cdot \mathbb{P}(Z = 2)\\
&+ \mathbb{P}(\text{Kugel rot} \mid Z = 3) \cdot \mathbb{P}(Z = 3)\\
&+ \mathbb{P}(\text{Kugel rot} \mid Z = 5) \cdot \mathbb{P}(Z = 5)\\
=& \frac{2}{5} \cdot \frac{6}{10} + \frac{3}{5} \cdot \frac{3}{10} + \frac{5}{5} \cdot \frac{1}{10}\\
=& \frac{12 + 9 + 5}{50} = \frac{13}{25}.
\end{aligned}$$

e) Wie groß ist die bedingte Wahrscheinlichkeit p, dass die beiden zuerst gezogenen Kugeln rot sind, unter der Bedingung, dass die zuletzt gezogene Kugel rot ist?
Lösung:

$$\begin{aligned} p &= \mathbb{P}(\text{die beiden zuerst gezogenen Kugeln rot}|\text{die zuletzt gezogene Kugel ist rot}) \\ &= \mathbb{P}(\text{alle drei gezogenen Kugeln rot})/\mathbb{P}(\text{die zuletzt gezogene Kugel ist rot}) \\ &= \frac{\frac{3}{5} \cdot \frac{2}{4} \cdot \frac{3}{5}}{\frac{13}{25}} = \frac{9}{26}. \end{aligned}$$

Aufgabe B4

Auf einer Metallhobelmaschine werden ringförmige Platten mit zufälligem Außendurchmesser X und Innendurchmesser Y hergestellt. X mit Normalverteilung $\mathcal{N}(100, 10)$ und Y mit Normalverteilung $\mathcal{N}(60, 6)$ seien stochastisch unabhängig. Falls die Ringbreite $U = \frac{X-Y}{2}$ in das Intervall $[14, 26]$ fällt, so wird die Platte hergestellt, ansonsten wird der Vorgang abgebrochen.

a) Wie groß ist die Wahrscheinlichkeit, dass X zwischen 98 und 103 liegt?
Lösung:

$$\begin{aligned}
\mathbb{P}(98 \leq X \leq 103) &= \Phi(\frac{103-100}{\sqrt{10}}) - \Phi(\frac{98-100}{\sqrt{10}}) \\
&= \Phi(0.95) - \Phi(-0.63) = \Phi(0.95) + \Phi(0.63) - 1 \\
&= 0.8289 + 0.7357 - 1 = 0.5646.
\end{aligned}$$

b) Geben Sie den Erwartungswert und die Varianz von U an. Welche Verteilung hat U? Wie groß ist die Wahrscheinlichkeit, dass eine Platte hergestellt wird?
Lösung: Es gilt $\mathbb{E}(U) = (100 - 60)/2 = 20$; wegen der Unabhängigkeit von X und Y gilt weiter $V(U) = (10+6)/4 = 4$. Nach der Faltungsformel ist U $\mathcal{N}(20, 4)$-verteilt, es ist also $\sigma := \sqrt{V(U)} = \sqrt{4} = 2$. Somit folgt

$$\begin{aligned}
\mathbb{P}(14 \leq U \leq 26) &= \mathbb{P}(|U - 20| \leq 6) \\
&= \mathbb{P}(|U - 20| \leq 3\sigma) \\
&= 2 \cdot \Phi(3) - 1 = 2 \cdot 0.9987 - 1 = 0.9974.
\end{aligned}$$

c) Sei $W = \frac{X+Y}{2}$. Welche Verteilung besitzt W?
Lösung: Es gilt $W \sim \mathcal{N}(80, 4)$.

d) Berechnen Sie Kovarianz und Korrelationskoeffizient von U und W.
Lösung:

$$\begin{aligned}
C(U, W) &= C\left(\frac{X-Y}{2}, \frac{X+Y}{2}\right) \\
&= \frac{1}{4} \cdot [C(X, X) + C(X, Y) - C(Y, X) - C(Y, Y)] \\
&= \frac{1}{4} \cdot (V(X) - V(Y)) = \frac{1}{4} \cdot (10 - 6) = 1.
\end{aligned}$$

Weiter gilt

$$\rho(U, W) = \frac{C(U, W)}{\sqrt{V(U) \cdot V(W)}} = \frac{1}{4}.$$

e) Die zufällige Fläche der ringförmigen Platte ist $F = \pi \cdot U \cdot W = \frac{\pi}{4}(X^2 - Y^2)$. Bestimmen Sie den Erwartungswert von F.
Lösung:

$$\begin{aligned}
\mathbb{E}F &= \mathbb{E}(\pi \cdot U \cdot W) \\
&= \pi \cdot [C(U, W) + \mathbb{E}U \cdot \mathbb{E}W] \\
&= \pi \cdot (1 + 20 \cdot 80) = 1601\pi.
\end{aligned}$$

Alternativ:

$$\begin{aligned}
\mathbb{E}F &= \frac{\pi}{4} \cdot \mathbb{E}\left(X^2 - Y^2\right) \\
&= \frac{\pi}{4} \cdot \mathbb{E}\left(V(X) + \mathbb{E}(X^2) - V(Y) - \mathbb{E}(Y^2)\right) \\
&= \frac{\pi}{4} \cdot (10 + 10000 - 6 - 3600) = 1601\pi.
\end{aligned}$$

Aufgabe B5

Die zufällige Lebensdauer X eines Produktes besitze die Dichte

$$f_\vartheta(t) = \begin{cases} 2 \cdot \vartheta \cdot t \cdot \exp(-\vartheta \cdot t^2), & t > 0, \\ 0, & \text{sonst}, \end{cases}$$

und die Verteilungsfunktion

$$F_\vartheta(t) = \begin{cases} 1 - \exp(-\vartheta \cdot t^2), & t > 0, \\ 0, & \text{sonst}, \end{cases}$$

mit einem unbekannten Parameter $\vartheta > 0$.

Hinweis: Es gilt $\mathbb{E}X = \frac{1}{2}\sqrt{\frac{\pi}{\vartheta}}$ und $V(X) = \frac{4-\pi}{4\vartheta}$.

a) Bestimmen Sie die Likelihood-Funktion $L_x(\vartheta)$ und die Log-Likelihood-Funktion $M_x(\vartheta)$ zur Stichprobe $x = (x_1, \ldots, x_n)$, wobei $x_i > 0$ für $i = 1, \ldots, n$ gelte.
Lösung: Die Likelihood-Funktion ist

$$L_x(\vartheta) = \prod_{i=1}^n f_\vartheta(x_i) = (2\vartheta)^n \exp\left(-\vartheta \sum_{i=1}^n x_i^2\right) \cdot \prod_{i=1}^n x_i.$$

Weiter gilt

$$\log(f_\vartheta(t)) = \log\vartheta + \log(2t) - \vartheta t^2, \qquad t > 0$$

somit ergibt sich die Loglikelihood-Funktion zu

$$\begin{aligned} M_x(\vartheta) &= \sum_{i=1}^n \log(f_\vartheta((x_i)) = \sum_{i=1}^n \left(\log\vartheta + \log(2x_i) - \vartheta x_i^2\right) \\ &= n\log\vartheta + \sum_{i=1}^n \log(2x_i) - \vartheta \sum_{i=1}^n x_i^2. \end{aligned}$$

b) Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ zur Stichprobe x.
Lösung: Die Ableitung der Loglikelihood-Funktion nach ϑ ist

$$M_x'(\vartheta) = \frac{n}{\vartheta} - \sum_{i=1}^n x_i^2.$$

Damit gilt $M_x'(\vartheta) = 0$ für $\vartheta = n/\sum_{i=1}^n x_i^2$. Da für die zweite Ableitung

$$M_x''(\vartheta) = -\frac{n}{\vartheta^2} < 0$$

gilt, erhält man den Maximum-Likelihood-Schätzer

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

c) Bestimmen Sie den Momentenschätzer $\tilde{\vartheta}(x)$ für ϑ zur Stichprobe $x = (x_1, \ldots, x_n)$.
Lösung: Aus der Gleichung

$$\frac{1}{2}\sqrt{\frac{\pi}{\vartheta}} = \mathbb{E}X = \bar{x}$$

ergibt sich

$$\tilde{\vartheta}(x) = \frac{\pi}{4\bar{x}^2}.$$

d) $X_1, \ldots, X_n$ seien unabhängige Zufallsvariablen mit Dichte f_ϑ. Bestimmen Sie Konstanten a und k so, dass $\frac{1}{n}\sum_{i=1}^n X_i^2$ ein erwartungstreuer Schätzer für $\gamma(\vartheta) = a \cdot \vartheta^k$ ist.
Lösung:

$$\begin{aligned} \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i^2 = \mathbb{E}X_1^2 \\ &= V(X_1) + (\mathbb{E}X_1)^2 = \frac{4-\pi}{4\vartheta} + \frac{\pi}{4\vartheta} = \frac{1}{\vartheta}. \end{aligned}$$

Der Schätzer ist also erwartungstreu für $\gamma(\vartheta) = 1/\vartheta$, es muss also $a = 1$ und $k = -1$ gewählt werden.

e) Bestimmen Sie die Verteilung von $X_1^2/3$.
Lösung: Für $x > 0$ ist

$$\mathbb{P}(X_1^2/3 \le x) = \mathbb{P}(X_1 \le \sqrt{3x}) = 1 - \exp(-\vartheta \cdot 3x),$$

d.h. $X_1^2/3$ ist exponentialverteilt mit Parameter 3ϑ.