INSTITUT FÜR STOCHASTIK
Karlsruher Institut für Technologie
Priv.-Doz. Dr. D. Kadelka

SS 2010

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 14.9.2010

Musterlösungen

Aufgabe 1:

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

j	1	2	3	4	5	6	7	8
x_j	0.5	2	3.1	4	5.4	5.8	7	7.9
y_j	10.9	11.8	6.8	8.5	5.2	6.3	5	5.3

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

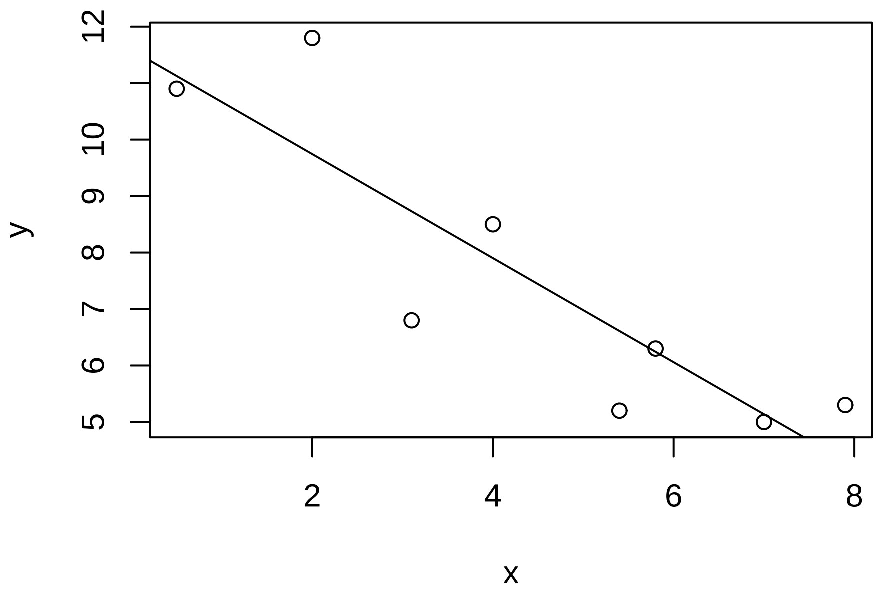
$$\sum_{j=1}^{n} (x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 4.46 \qquad s_x = 2.529$$
$$\bar{y} = 7.47 \qquad s_y = 2.655$$
$$r_{xy} = -0.8774$$

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

$$\begin{aligned} b^* &= -0.921 \\ a^* &= 11.59 \end{aligned}$$

und die Regressionsgerade $y = 11.59 - 0.921 \cdot x$.



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

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

$$y_{()} = (5,\ 5.2,\ 5.3,\ 6.3,\ 6.8,\ 8.5,\ 10.9,\ 11.8)$$

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

$$\bar{y}_{0.15} = \frac{1}{8 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(7)}) = 7.167$$

d) Bestimmen Sie das Stichproben-0.2-Quantil $\tilde{y}_{0.2}$ von $(y_1, \ldots, y_8)$.
Lösung: Da $8 \cdot 0.2 = 1.6$ nicht ganzzahlig ist, ist mit $k = [1.6] = 1$

$$\tilde{y}_{0.2} = y_{(k+1)} = y_{(2)} = 5.2$$

e) Berechnen Sie den Quartilsabstand von $(y_1, \ldots, y_8)$.
Lösung: Da $0.25 \cdot 8 = 2$ und $0.75 \cdot 8 = 6$ beide ganzzahlig sind, ergibt sich mit $k_1 = 2$ und $k_2 = 6$

$$\begin{aligned}
\tilde{y}_{0.25} &= \frac{y_{(k_1)} + y_{(k_1+1)}}{2} = \frac{y_{(2)} + y_{(3)}}{2} = 5.25 \\
\tilde{y}_{0.75} &= \frac{y_{(k_2)} + y_{(k_2+1)}}{2} = \frac{y_{(6)} + y_{(7)}}{2} = 9.7
\end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 4.45$.

Aufgabe 2

In einem Rohr soll Wasser nur in einer Richtung fließen. Dazu baut man hintereinander in das Rohr zwei Ventile ein, die das Wasser jeweils nur in ein und derselben Richtung durchlassen sollen.

Ventil 1 bzw. 2 ist in Zustand $\left\{\begin{array}{c} 0 \\ 1 \\ 2 \end{array}\right\}$, falls es das Wasser $\left\{\begin{array}{c} \text{in keiner Richtung} \\ \text{in beide Richtungen} \\ \text{in der gewünschten Richtung} \end{array}\right\}$ durchlässt.

Sei X der zufällige Zustand des ersten, Y der des zweiten Ventils. Es ist also

$$\begin{aligned} A &:= \text{„Die Ventilkombination funktioniert richtig“} \\ &= [X=1, Y=2] + [X=2, Y=1] + [X=2, Y=2]. \end{aligned}$$

a) Kreuzen Sie an, welche der nachfolgenden Aussagen richtig oder falsch sind. (Falsche Antworten ergeben einen Punkteabzug in dieser Teilaufgabe!)

	richtig	falsch
$A = [X = 2] \cap [Y = 2]$	☐	☐
$A = [X \geq 1] \cap [Y \geq 1]$	☐	☐
$A = [X + Y \geq 3]$	☐	☐
$A = [X \leq 1]^c \cup [Y \leq 1]^c$	☐	☐

Es seien X und Y unabhängig mit

$$\mathbb{P}(X = 1) = \mathbb{P}(Y = 1) = 0.09 \text{ und } \mathbb{P}(X = 2) = \mathbb{P}(Y = 2) = 0.9.$$

b) Bestimmen Sie $\mathbb{P}(Y = 0)$ und $\mathbb{P}(Y = 0 \mid X = 1)$.

c) Berechnen Sie $\mathbb{P}(X = 2, Y = 2)$.

d) Bestimmen Sie die Wahrscheinlichkeit, dass die Ventilkombination richtig funktioniert.

e) Berechnen Sie $\mathbb{P}(X + Y \leq 2)$.

Lösung:

a) $A = [X = 2] \cap [Y = 2]$ kann nicht richtig sein. Denn $[X = 2] \cap [Y = 2]$ ist gleichbedeutend damit, dass beide Ventile das Wasser in der gewünschten Richtung durchlassen. Aber die Ventilkombination funktioniert z.B. auch richtig, wenn $X = 1$ und $Y = 2$.

$A = [X \geq 1] \cap [Y \geq 1]$ kann nicht richtig sein. Denn falls $X = 1$ und $Y = 1$ gilt, so gilt auch $X \geq 1$ und $Y \geq 1$. Die Ventilkombination funktioniert dann aber dennoch nicht richtig.

Für die dritte Teilaufgabe überlegt man sich, wann genau $X + Y \geq 3$ möglich ist. Dies ist genau dann der Fall, wenn $X \geq 1$ und $Y \geq 1$ gilt, aber wenn nicht gleichzeitig $X = 1$ und $Y = 1$ gilt. Die Aussage ist also richtig.

$[X \leq 1]^c \cup [Y \leq 1]^c$ ist gleichbedeutend mit $[X = 2] \cup [Y = 2]$. Dies ist z.B. erfüllt, wenn Ventil 1 das Wasser in der richtigen Richtung durchlässt ($X = 2$), Ventil 2 das Wasser in keiner Richtung durchlässt ($Y = 0$). Die Aussage ist also falsch.

b) Da Y nur die Werte 0, 1 und 2 annimmt, gilt

$$\mathbb{P}(Y = 0) = 1 - \mathbb{P}(Y \neq 0) = 1 - \mathbb{P}(Y = 1) - \mathbb{P}(Y = 2) = 0.01$$

Da ferner X und Y stochastisch unabhängig sind, gilt

$$\mathbb{P}(Y = 0 \mid X = 1) = \mathbb{P}(Y = 0) = 0.01.$$

c) Wegen der Unabhängigkeit von X und Y gilt

$$\mathbb{P}(X = 2, Y = 2) = \mathbb{P}(X = 2) \cdot \mathbb{P}(Y = 2) = 0.9 \cdot 0.9 = 0.81.$$

d) Wegen der Unabhängigkeit von X und Y gilt diesmal

$$\begin{aligned}\mathbb{P}(A) = \mathbb{P}(X = 1, Y = 2) + \mathbb{P}(X = 2, Y = 1) + \mathbb{P}(X = 2, Y = 2) = \\ \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = 2) + \mathbb{P}(X = 2) \cdot \mathbb{P}(Y = 1) + \mathbb{P}(X = 2) \cdot \mathbb{P}(Y = 2) = \\ 0.09 \cdot 0.9 + 0.9 \cdot 0.09 + 0.9\colon 0.9 = 0.972.\end{aligned}$$

e) Wegen $A^c = [X+Y \geq 3]^c = [X+Y \leq 2]$ gilt $\mathbb{P}(X+Y \leq 2) = \mathbb{P}(A^c) = 1-\mathbb{P}(A) = 0.028.$

Aufgabe 3

Ein Unternehmen beschäftigt zwei Gruppen von Verkäufern. Die Umsätze $X_1, \ldots, X_{10}$ (alle in €) der 10 Verkäufer in Gruppe 1 seien normalverteilt mit Mittelwert 800 und Standardabweichung 240, während die Umsätze $Y_1, \ldots, Y_{40}$ der 40 Verkäufer in Gruppe 2 normalverteilt mit Mittelwert 1000 und Standardabweichung 300 sind. Die Zufallsvariablen $X_1, \ldots, X_{10}, Y_1, \ldots, Y_{40}$ seien stochastisch unabhängig. $Z := X_1 + \ldots + X_{10} + Y_1 + \ldots + Y_{40}$ sei der tägliche Gesamtumsatz aller Verkäufer.

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

b) Welche Verteilung besitzt der Gesamtumsatz Z und welche Verteilung besitzt $\bar{Z} = \frac{Z}{50}$, der tägliche mittlere Umsatz aller Verkäufer?

c) Mit welcher Wahrscheinlichkeit ist der tägliche Gesamtumsatz Z größer als 50000 €?

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

Lösung:

a) Nach Voraussetzung gilt $X_1, \ldots, X_{10} \sim \mathcal{N}(800, 240^2)$ und $Y_1, \ldots, Y_{40} \sim \mathcal{N}(1000, 300)$. Dann gilt für den Gesamtumsatz Z

$$\begin{aligned}\mathbb{E}Z = \mathbb{E}(X_1 + \ldots + X_{10} + Y_1 + \ldots + Y_{40}) = \mathbb{E}X_1 + \ldots + \mathbb{E}X_{10} + \mathbb{E}Y_1 + \ldots + \mathbb{E}Y_{40} \\ = 10 \cdot 800 + 40 \cdot 1000 = 48000\end{aligned}$$

und wegen der Unabhängigkeit der Einzelumsätze nach Satz 12.23 f)

$$\begin{aligned}V(Z) = V(X_1 + \ldots + X_{10} + Y_1 + \ldots + Y_{40}) = V(X_1) + \ldots + V(X_{10} + V(Y_1) + \ldots + V(Y_{40}) \\ = 10 \cdot 240^2 + 40 \cdot 300^2 = 4176000.\end{aligned}$$

b) Da die Zufallsvariablen $X_1, \ldots, X_{10}, Y_1, \ldots, Y_{40}$ stochastisch unabhängig und normalverteilt sind, ist die Summe $Z = X_1 + \ldots + X_{10} + Y_1 + \ldots + Y_{40}$ nach der Tabelle auf S. 114 wieder normalverteilt, also $Z \sim \mathcal{N}(\mu, \sigma^2)$ mit den Parametern $\mu = \mathbb{E}Z$ und $\sigma^2 = V(Z)$. Wegen a) gilt also

$$Z \sim \mathcal{N}(48000, 4176000).$$

Nach Satz 9.7 gilt dann

$$\bar{Z} = \frac{1}{50} \cdot Z \sim \mathcal{N}\left(\frac{48000}{50}, \frac{4176000}{50^2}\right) = \mathcal{N}(960, 1670.4)$$

c) Zu berechnen ist $\mathbb{P}(Z > 50000)$.

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

$$\begin{aligned} \mathbb{P}(Z > 50000) &= 1 - \mathbb{P}(Z \leq 50000) = 1 - \Phi\left(\frac{50000 - 48000}{\sqrt{4176000}}\right) \\ &= 1 - \Phi(0.9787) = \approx 1 - \Phi(0.98) = 1 - 0.8365 = 0.1635. \end{aligned}$$

nach Anhang A.1 im Skriptum.

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

$$0.975 = \mathbb{P}(Z \leq t) = \Phi\left(\frac{t - 48000}{\sqrt{4176000}}\right)$$

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

$$1.96 = \frac{t - 48000}{\sqrt{4176000}}$$

also

$$t = 48000 + 1.96 \cdot \sqrt{4176000} = 52005.$$

Aufgabe 4

Die stochastisch unabhängigen Zufallsvariablen X, Y und Z besitzen alle die Exponentialverteilung $Exp(0.5)$.

a) Bestimmen Sie den Erwartungswert und die Varianz von X.

b) Bestimmen Sie die Verteilung der Zufallsvariablen $U := X + Y + Z$ (einschließlich der Parameter).

c) Bestimmen Sie das 0.05-Quantil $q_{0.05}$ der Zufallsvariablen X (Genauigkeit: 4 Nachkommastellen)

d) Berechnen Sie die Kovarianz $C(U, X)$ und den Korrelationskoeffizienten $\rho(U, X)$ von U und X.

e) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(-1 \leq X \leq 2,\ 1 \leq Y \leq 3)$.

Lösung:

a) Allgemein besitzt $Exp(\alpha)$ den Erwartungswert $\frac{1}{\alpha}$ und die Varianz $\frac{1}{\alpha^2}$. Damit gilt

$$\mathbb{E}X = \frac{1}{0.5} = 2 \quad \text{und} \quad V(X) = \frac{1}{0.5^2} = 4.$$

b) Wegen $Exp(0.5) = \Gamma(1, 0.5)$ und der Faltungsformel in 11.16 gilt

$$U = X + Y + Z \sim \Gamma(1 + 1 + 1, 0.5) = \Gamma(3, 0.5).$$

c) Wegen Beispiel 12.21 ist $q_{0.05} = -\frac{1}{\alpha} \cdot \ln(1 - 0.05)$ das 0.05-Quantil von $Exp(\alpha)$, hier also mit $\alpha = 0.5$

$$q_{0.05} = -\frac{1}{0.05} \cdot \ln(0.95) = 0.1026.$$

d) Wegen Satz 12.23 und der Unabhängigkeit von X und $Y + Z$, also $C(Y + Z, X) = 0$, gilt

$$C(U, X) = C(X + (Y + Z), X) = C(X, X) + C(Y + Z, X) = V(X) = 4,$$

ferner wegen $U \sim \Gamma(3, 0.5)$

$$V(U) = \frac{3}{0.5^2} = 12$$

und damit

$$\rho(U, X) = \frac{C(U, X)}{\sqrt{V(U) \cdot V(X)}} = \frac{4}{\sqrt{12 \cdot 4}} = \frac{1}{\sqrt{3}} = 0.5774.$$

e) Wegen der Unabhängigkeit von $X \sim Exp(0.5)$ und $Y \sim Exp(0.5)$ und da X und Y stetig sind gilt

$$\begin{aligned} \mathbb{P}(-1 \leq X \leq 2, 1 \leq Y \leq 3) &= \mathbb{P}(-1 \leq X \leq 2) \cdot \mathbb{P}(1 \leq Y \leq 3) \\ &= (F_X(2) - F_X(-1)) \cdot (F_Y(3) - F_Y(1)) \\ &= ((1 - e^{-0.5 \cdot 2}) - 0) \cdot ((1 - e^{-0.5 \cdot 3}) - (1 - e^{-0.5 \cdot 1})) \\ &\qquad 0.6321 \cdot (0.7769 - 0.3935) = 0.2424. \end{aligned}$$

Aufgabe 5

Ein Merkmal besitze eine logarithmische Normalverteilung $\mathcal{LN}(\vartheta, 4)$ mit der Dichte

$$f_\vartheta(x) = \begin{cases} 0 & , x \leq 0 \\ \frac{1}{\sqrt{8\pi}x} \cdot e^{-\frac{1}{8}(\ln(x) - \vartheta)^2} & , x > 0, \end{cases}$$

wobei der Parameter $\vartheta \in \mathbb{R}$ unbekannt ist.

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 zur Stichprobe $x = (x_1, \ldots, x_n)$ ist, wobei $x_i > 0$ für $i = 1, \ldots, n$.

c) Bestimmen Sie den Momentenschätzer für ϑ zur Stichprobe $x = (x_1, \ldots, x_n)$.

d) Bestimmen Sie den Momentenschätzer $\hat{\vartheta}_1(x_1, \ldots, x_n)$ für ϑ zur Stichprobe $x = (x_1, \ldots, x_n)$, wobei $x_i > 0$ für $i = 1, \ldots, n$.

e) Sei Y eine beliebige $\mathcal{LN}(\vartheta, 4)$-verteilte Zufallsvariable. Bestimmen Sie den Erwartungswert $\mathbb{E}(\ln(Y))$ und die Varianz $V(\ln(Y))$ von $\ln(Y)$.

<u>Hinweis:</u> Sie können diese Werte auch ohne Integration bestimmen.

Lösung:

a) Mit

$$\ln(f_\vartheta(x)) = -\ln(\sqrt{8\pi}x) - \frac{1}{8}(\ln(x) - \vartheta)^2$$

ergibt sich die Loglikelihood-Funktion zu

$$\begin{aligned} M_x(\vartheta) &= \sum_{i=1}^{n} \ln(f_\vartheta(x_i)) = \sum_{i=1}^{n} \left(-\ln(\sqrt{8\pi}x_i) - \frac{1}{8}(\ln(x_i) - \vartheta)^2 \right) \\ &= -\sum_{i=1}^{n} \ln(\sqrt{8\pi}x_i) - \frac{1}{8}\sum_{i=1}^{n}(\ln(x_i) - \vartheta)^2. \end{aligned}$$

b) M_x hat die Ableitungen

$$\begin{aligned} M_x'(\vartheta) &= -\frac{1}{8}\sum_{i=1}^{n} -2 \cdot (\ln(x_i) - \vartheta) = \frac{1}{4}\sum_{i=1}^{n}(\ln(x_i - \vartheta) = \sum_{i=1}^{n} \ln(x_i) - n \cdot \vartheta \\ M_x''(\vartheta) &= -\frac{n}{4} < 0. \end{aligned}$$

Da die zweite Ableitung 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 damit

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

c) 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 $\mathcal{LN}(\vartheta, 4)$ besitzt, gilt nach der Tabelle auf S. 124 im Skriptum $\mathbb{E}_\vartheta(X_1) = e^{\vartheta+2}$ und damit

$$\begin{aligned} e^{\vartheta+2} &= \bar{x} \\ \vartheta + 2 &= \ln(\bar{x}) \\ \vartheta &= \ln(\bar{x}) - 2 \end{aligned}$$

Der gesuchte Momentenschätzer ist also $\hat{\vartheta}_m(x) := \ln(\bar{x}) - 2$.

d) Hat eine Zufallsvariable X die Verteilung $\mathcal{N}(\vartheta, 4)$, so hat nach Skript 9.10 die Zufallsvariable $Y = e^X$ die Verteilung $\mathcal{LN}(\vartheta, 4)$. Hat also umgekehrt Y die Verteilung $\mathcal{LN}(\vartheta, 4)$, so hat $X := \ln(Y)$ die Verteilung $\mathcal{N}(\vartheta, 4)$. Damit gilt

$$\mathbb{E}(\ln(Y)) = \vartheta \text{ und } V(\ln(Y)) = 4.$$