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

WS 2010/11

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 7.2.2011

Musterlösungen

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	1	2.3	3.1	4	5.1	6.3	6.8	8.6
y_j	9.5	5.9	6.4	4.4	3.7	5.6	0.5	−2.6

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

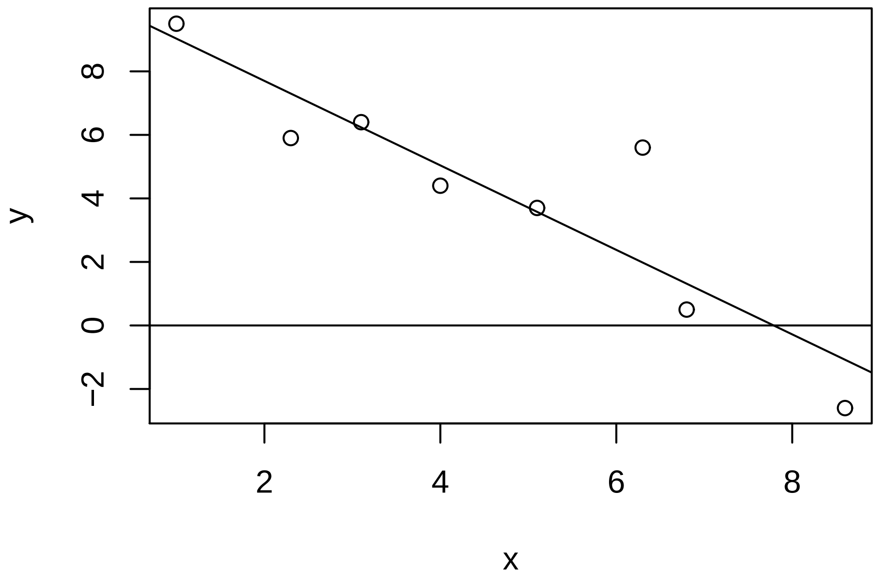
$$\sum_{j=1}^{n}(x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\begin{aligned} \bar{x} &= 4.65 & \qquad s_x &= 2.53 \\ \bar{y} &= 4.17 & \qquad s_y &= 3.742 \\ r_{xy} &= -0.8995 \end{aligned}$$

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

$$\begin{aligned} b^* &= -1.33 \\ a^* &= 10.36 \end{aligned}$$

und die Regressionsgerade $y = 10.36 - 1.33 \cdot x$.



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

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

$$y_{()} = (-2.6,\ 0.5,\ 3.7,\ 4.4,\ 5.6,\ 5.9,\ 6.4,\ 9.5)$$

c) Berechnen Sie das 0.2-getrimmte Stichprobenmittel $\bar{y}_{0.2}$ von $(y_1, \ldots, y_8)$.
Lösung: Mit $k = [8 \cdot 0.2] = 1$ ergibt sich

$$\bar{y}_{0.2} = \frac{1}{8 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(7)}) = 4.417$$

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 $k = [1.6] = 1$

$$\tilde{y}_{0.2} = y_{(k+1)} = y_{(2)} = 0.5$$

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 $k_2 = 6$

$$\begin{aligned}
\tilde{y}_{0.25} &= \frac{y_{(k_1)} + y_{(k_1+1)}}{2} = \frac{y_{(2)} + y_{(3)}}{2} = 2.1 \\
\tilde{y}_{0.75} &= \frac{y_{(k_2)} + y_{(k_2+1)}}{2} = \frac{y_{(6)} + y_{(7)}}{2} = 6.15
\end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 4.05$.

Aufgabe B2

Ein Transformatorkern habe die zufällige Dicke X (in mm) und eine Spule den zufällige Innendurchmesser Y. X und Y seien stochastisch unabhängige Zufallsvariable mit den Normalverteilungen $\mathcal{N}(50, 0.34)$ bzw. $\mathcal{N}(51, 0.15)$.

a) Welche Verteilung besitzt $Z := Y - X$?

b) Berechnen Sie die Kovarianz $C(Z, Y)$ und den Korrelationskoeffizienten $\rho(Z, Y)$ von Z und Y.

c) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(Z > 0)$, dass ein zufällig ausgewählter Kern in eine zufällig ausgewählte Spule passt.

d) Bestimmen Sie $t \in \mathbb{R}$ so, dass $\mathbb{P}(X > t) = 0.05$ gilt.

Lösung:

a) Es gilt $Z = Y + (-X)$. Wegen Satz 9.7 gilt $-X \sim \mathcal{N}(-50, 0.34)$, also nach der Faltungsformel

$$Z \sim \mathcal{N}(51 - 50, 0.34 + 0.15) = \mathcal{N}(1, 0.49)$$

(Die Varianzen dürfen hier **nicht** subtrahiert werden!)

b) Anwenden von Satz 12.23 ergibt

$$C(Z, Y) = C(Y + (-X), Y) = C(Y, Y) + C(-X, Y) = V(Y) + 0 = 0.15,$$

da $-X$ und Y stochastisch unabhängig sind. Damit ergibt sich der Korrelationskoeffizient zu

$$\rho(Z, Y) = \frac{C(Z, Y)}{\sqrt{V(Z) \cdot V(Y)}} = \frac{0.15}{\sqrt{0.49 \cdot 0.15}} = 0.5533$$

c) Wegen Satz 9.6 und (9.5) gilt

$$\begin{aligned}\mathbb{P}(Z > 0) = 1 - \mathbb{P}(Z \leq 0) = 1 - \Phi_{1,0.49}(0) = 1 - \Phi\left(\frac{0 - 1}{0.7}\right) \\ = 1 - \Phi(-1.43) = \Phi(1.43) = 0.9236.\end{aligned}$$

d) Die Gleichung

$$0.05 = \mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t) = 1 - \Phi_{50,0.34}(t) = 1 - \Phi\left(\frac{t - 50}{\sqrt{0.34}}\right)$$

ist nach t aufzulösen, also

$$\Phi\left(\frac{t - 50}{\sqrt{0.34}}\right) = 1 - 0.05 = 0.95$$

Da $\Phi(1.6449) = 0.95$ gilt (vergl. Skriptum, 12.20), muss also

$$\frac{t - 50}{\sqrt{0.34}} = 1.6449$$

gelten, also $t = 1.6449 \cdot 0.5831 + 50 = 50.9591$.

Aufgabe B3

Die Zufallsvariablen X und Y nehmen die Werte $0, 1$ und 2 an. Dabei seien die folgenden Wahrscheinlichkeiten bekannt:

$$\mathbb{P}(X=0) = \mathbb{P}(Y=0) = 0.2,\ \mathbb{P}(X=1) = \mathbb{P}(Y=1) = 0.4,\ \mathbb{P}(X=2) = \mathbb{P}(Y=2) = 0.4$$

und

$$\mathbb{P}(X=0, Y=0) = 0.04, \quad \mathbb{P}(X=0, Y=1) = 0.06,$$
$$\mathbb{P}(X=1, Y=1) = 0.20, \quad \mathbb{P}(X=2, Y=2) = 0.10.$$

a) Ergänzen Sie die folgende Tabelle von $\mathbb{P}(X = i, Y = j)$:

j \ i	0	1	2	$\mathbb{P}(Y = j)$
0	0.04		0.10	0.20
1	0.06	0.20		0.40
2				0.40
$\mathbb{P}(X = i)$	0.20	0.40	0.40	

b) Berechnen Sie $\mathbb{E}X$, $\mathbb{E}Y$, $V(X)$, $V(Y)$, die Kovarianz $C(X, Y)$ von X und Y und den Korrelationskoeffizienten $\rho(X, Y)$.

c) Berechnen Sie die bedingte Wahrscheinlichkeit $\mathbb{P}(X = 1 \mid Y = 1)$.

d) Sind X und Y unabhängig?

Lösung:

a) In der rechten Spalte der Tafel stehen die Zeilensummen, in der unteren Zeile die Spaltensumme.

Für die erste Zeile gilt $0.04+\mathbb{P}(X = 1, Y = 0)+0.10 = 0.20$, also $\mathbb{P}(X = 1, Y = 0) = 0.06$. Für die zweite Zeile gilt $0.06 + 0.20 + \mathbb{P}(X = 2, Y = 1) = 0.40$, also $\mathbb{P}(X = 2, Y = 1) = 0.14$. Die Werte in der dritten Zeile erhalten wir dann dadurch, dass sich die Werte in den Spalten zu den Spaltensummen, die in der untersten Zeile stehen, aufsummieren müssen. Wir erhalten

j \ i	0	1	2	$\mathbb{P}(Y = j)$
0	0.04	0.06	0.10	0.20
1	0.06	0.20	0.14	0.40
2	0.10	0.14	0.16	0.40
$\mathbb{P}(X = i)$	0.20	0.40	0.40	1.00

b) Da X und Y dieselbe Verteilung besitzen, erkennt man schon ohne Rechnung $\mathbb{E}X = \mathbb{E}Y$ und $V(X) = V(Y)$. Es gilt

$$\begin{aligned} \mathbb{E}X &= \mathbb{E}Y = 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) + 2 \cdot \mathbb{P}(X = 2) = 0.4 + 2 \cdot 0.4 = 1.2, \\ \mathbb{E}X^2 &= \mathbb{E}Y^2 = 0 \cdot \mathbb{P}(X = 0) + 1^2 \cdot \mathbb{P}(X = 1) + 2^2 \cdot \mathbb{P}(X = 2) = 0.4 + 2^2 \cdot 0.4 = 2.0 \end{aligned}$$

und daraus $V(X) = V(Y) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 2.0 - 1.2^2 = 0.56$. Schließlich gilt

$$\begin{aligned} \mathbb{E}(X \cdot Y) = 1 \cdot \mathbb{P}(X = 1, Y = 1) + 1 \cdot 2 \cdot \mathbb{P}(X = 1, Y = 2) \\ + 2 \cdot 1 \cdot \mathbb{P}(X = 2, Y = 2) + 2 \cdot 2 \cdot \mathbb{P}(X = 2, Y = 2) = 1.40 \end{aligned}$$

und damit

$$C(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y = 1.40 - 1.2 \cdot 1.2 = -0.04$$

und

$$\rho(X, Y) = \frac{C(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{-0.04}{0.56} = -\frac{1}{14} = -0.0714.$$

c) $\mathbb{P}(X = 1 \mid Y = 1) = \dfrac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(Y = 1)} = \dfrac{0.20}{0.40} = \dfrac{1}{2} = 0.5.$

d) Wären X und Y unabhängig, so müsste

$$0.5 = \mathbb{P}(X = 1 \mid Y = 1) = \mathbb{P}(X = 1) = 0.4$$

gelten. Da dies falsch ist, können X und Y nicht unabhängig sein. Alternativ: Wegen $\rho(X, Y) \neq 0$ können X und Y nicht unabhängig sein.

Aufgabe B4

Seien X und Y stochastisch unabhängige Zufallsvariablen, wobei X die Binomialverteilung $Bin(1, \alpha)$ und Y die Poisson–Verteilung $Po(\beta)$ besitzt mit $\alpha \in (0, 1)$ und $\beta \in (0, \infty)$. $Z := X \cdot (Y + 1)$ sei die zufällige Lebensdauer eines elektronischen Bauteils. Es ist also $\{Z = 0\} = \{X = 0\}$ und $\{Z = k\} = \{X = 1, Y = k - 1\}$ für $k = 1, 2, \ldots$ (keine Begründung erforderlich).

a) Bestimmen Sie $\mathbb{E}(Y + 1)$, $V(Y + 1)$ und damit $\mathbb{E}(Y + 1)^2$.

b) Bestimmen Sie die Zähldichte f_Z von Z. (Begründen Sie Ihre Antwort!)

$$f_Z(0) = \boxed{} \qquad f_Z(k) = \boxed{}, \ k = 1, 2, 3, \ldots$$

c) Berechnen Sie $\mathbb{P}(Z = k \mid Z \geq 1)$ für $k = 1, 2, 3, \ldots$.

d) Berechnen Sie $\mathbb{E}Z$, $\mathbb{E}Z^2$ und daraus $V(Z)$.

Lösung:

a) Nach den Tabellen in 12.7 und 12.13 gilt

$$\mathbb{E}Y = \beta \quad \text{und} \quad V(Y) = \beta$$

und daher wegen Satz 12.6 und 12.11

$$\begin{aligned}
\mathbb{E}(Y+1) &= \mathbb{E}Y + 1 = \beta + 1 \\
V(Y+1) &= V(Y) = \beta \\
\mathbb{E}(Y+1)^2 &= V(Y+1) + (\mathbb{E}(Y+1))^2 = \beta + (\beta+1)^2 = \beta^2 + 3\cdot\beta + 1.
\end{aligned}$$

b) (Wegen $Y+1>0$ sind die Ereignisse $\{Z=0\}$ und $\{X=0\}$ identisch und daher auch die Ereignisse $\{Z>0\}$ und $\{X=1\}$ identisch und es ist $\{Z=k\} = \{X=1,\ Y+1=k\}$.) Nach dem Hinweis gilt

$$f_Z(0) = \mathbb{P}(Z=0) = \mathbb{P}(X=0) = 1-\alpha$$

und wegen der Unabhängigkeit von X und Y und wegen (7.16) für $k = 1, 2, 3, \ldots$

$$\begin{aligned}
f_Z(k) &= \mathbb{P}(Z=k) = \mathbb{P}(X=1,\ Y=k-1) = \mathbb{P}(X=1)\cdot\mathbb{P}(Y=k-1) \\
&= \alpha\cdot e^{-\beta}\frac{\beta^{k-1}}{(k-1)!}
\end{aligned}$$

c) Wegen $\{Z=k\} = \{X=1,\ Y=k-1\}$ für $k=1,2,3,\ldots$ gilt nach b)

$$\begin{aligned}
\mathbb{P}(Z=k \mid Z\geq 1) &= \frac{\mathbb{P}(Z=k,\ Z\geq 1)}{\mathbb{P}(Z\geq 1)} = \frac{\mathbb{P}(Z=k)}{1-\mathbb{P}(Z=0)} \\
&= \frac{\alpha\cdot e^{-\beta}\frac{\beta^{k-1}}{(k-1)!}}{\alpha} = e^{-\beta}\frac{\beta^{k-1}}{(k-1)!}.
\end{aligned}$$

d) Wegen der Unabhängigkeit von X und Y (und damit auch von X^k und $(Y+1)^k$) und wegen $X = X^2 \sim Bin(1,\alpha)$ gilt

$$\begin{aligned}
\mathbb{E}Z &= \mathbb{E}(X\cdot(Y+1)) = \mathbb{E}X\cdot\mathbb{E}(Y+1) = \alpha\cdot(\beta+1) \\
\mathbb{E}Z^2 &= \mathbb{E}(X^2\cdot(Y+1)^2) = \mathbb{E}X^2\cdot\mathbb{E}(Y+1)^2 = \alpha\cdot(\beta^2+3\cdot\beta+1) \\
V(Z) &= \mathbb{E}Z^2 - (\mathbb{E}Z)^2 = \alpha\cdot(\beta^2+3\cdot\beta+1) - \alpha^2\cdot(\beta+1)^2 \\
&= \alpha\beta + \alpha(1-\alpha)(1+\beta)^2.
\end{aligned}$$

Aufgabe B5

Ein Merkmal X besitze eine Verteilung mit der Dichte

$$f_\vartheta(x) = \frac{\vartheta e^x}{(1+e^x)^{\vartheta+1}}, \quad x\in\mathbb{R},$$

wobei $\vartheta > 0$ ein unbekannter Parameter ist, und die Verteilungsfunktion

$$F_\vartheta(t) = 1-(1+e^t)^{-\vartheta}, \quad t\in\mathbb{R}.$$

Der Parameter $\vartheta > 0$ soll aufgrund einer unabhängigen Stichprobe $x = (x_1,\ldots,x_n)\in\mathbb{R}^n$ geschätzt werden.

a) Bestimmen Sie die Loglikelihood-Funktion zur Stichprobe $x = (x_1, \ldots, x_n)$.

b) Bestimmen Sie den Maximum-Likelihood-Schätzer $\widehat{\vartheta}(x_1, \ldots, x_n)$ für ϑ.

c) Bestimmen Sie die Verteilungsfunktion F_Y der Zufallsvariablen

$$Y := \ln(1 + e^X).$$

d) Welche Verteilung besitzt Y?

Lösung:

a) Es gilt

$$\ln(f_\vartheta(x)) = \ln(\vartheta) + x - (\vartheta + 1) \cdot \ln(1 + e^x).$$

Damit ist die Log-Likelihoodfunktion

$$\begin{aligned} M_x(\vartheta) = \sum_{i=1}^{n} \ln(f_\vartheta(x_i)) = \sum_{i=1}^{n} \ln(\vartheta) + \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} (\vartheta + 1) \cdot \ln(1 + e^{x_i}) \\ = n \ln(\vartheta) + \sum_{i=1}^{n} x_i - (\vartheta + 1) \cdot \sum_{i=1}^{n} \ln(1 + e^{x_i}) \end{aligned}$$

b) Sie hat die Ableitung (nach ϑ)

$$M_x'(\vartheta) = \frac{n}{\vartheta} - \sum_{i=1}^{n} \ln(1 + e^{x_i})$$

und die zweite Ableitung

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

Aus $M_x'(\vartheta) = 0$ erhält man daher die gesuchte Maximumstelle $\hat{\vartheta}(x)$ von $M_x(\vartheta)$. Auflösen dieser Gleichung nach ϑ ergibt

$$\hat{\vartheta}(x_1, \ldots, x_n) = \frac{n}{\sum_{i=1}^{n} \ln(1 + e^{x_i})}.$$

c) Es ist $1 + e^X > 1$ und damit $Y = \ln(1 + e^X) > 0$. Damit gilt $F_Y(t) := \mathbb{P}(Y \leq 0) = 0$ für $t \leq 0$. Sei $t > 0$. Dann gilt

$$\begin{aligned} F_Y(t) &= \mathbb{P}(Y \leq t) = \mathbb{P}(\ln(1 + e^X) \leq t) = \mathbb{P}(1 + e^X \leq e^t) = \mathbb{P}(e^X \leq e^t - 1) \\ &= \mathbb{P}(X \leq \ln(e^t - 1)) = 1 - (1 + e^{\ln(e^t - 1)})^{-\vartheta} = 1 - (1 + e^t - 1)^{-\vartheta}) \\ &= 1 - (e^t)^{-\vartheta} = 1 - e^{-\vartheta t} \end{aligned}$$

Insgesamt erhalten wir

$$F_Y(t) = \begin{cases} 0, & \text{falls } t \leq 0, \\ 1 - e^{-\vartheta t}, & \text{falls } t > 0. \end{cases}$$

d) Da dies gerade die Verteilungsfunktion von $Exp(\vartheta)$ ist, gilt also $Y \sim Exp(\vartheta)$.