

Klausur
Wahrscheinlichkeitstheorie und Statistik
vom 8.3.2005
Musterlösungen

Aufgabe A1:

Gegeben sei eine Urliste mit den Paaren $(x_1, y_1), \dots, (x_{10}, y_{10})$

j	1	2	3	4	5	6	7	8	9	10
x_j	0.6	2	3.6	4.1	4.9	6	6.5	7.8	9.2	10
y_j	14.7	10.4	7.8	7.8	8.1	4.2	2.6	-0.5	-3	-1.9

- 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

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

$$\bar{x} = 5.47$$

$$s_x = 3.031$$

$$\bar{y} = 5.02$$

$$s_y = 5.743$$

$$r_{xy} = -0.9808$$

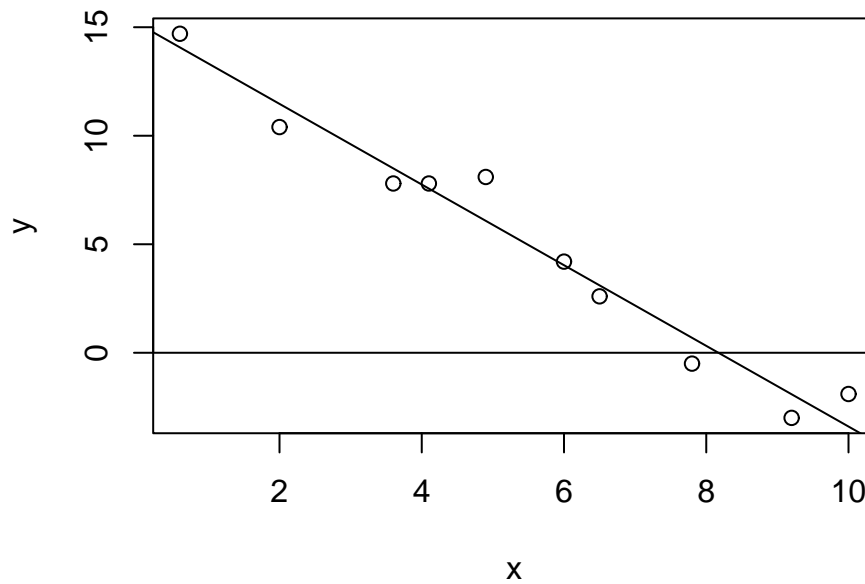
- b) Bestimmen Sie die zugehörige Regressionsgerade $y = a^* + b^* \cdot x$ von y auf x .

Lösung: Nach Paragraph 1.5 ist $b^* = r_{xy} \cdot \frac{s_y}{s_x}$ und $a^* = \bar{y} - b^* \cdot \bar{x}$, also

$$b^* = -1.858$$

$$a^* = 15.19$$

und die Regressionsgerade $y = 15.19 - 1.858 \cdot x$.



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

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

$$y_{(j)} = (-3, -1.9, -0.5, 2.6, 4.2, 7.8, 7.8, 8.1, 10.4, 14.7)$$

- c) Berechnen Sie das 0.14-getrimmte Stichprobenmittel $\bar{y}_{0.14}$ von (y_1, \dots, y_{10}) .
Lösung: Mit $k = [10 \cdot 0.14] = 1$ ergibt sich

$$\bar{y}_{0.14} = \frac{1}{10 - 2 \cdot 1} \cdot (y_{(2)} + \dots + y_{(9)}) = 4.812$$

- d) Berechnen Sie den Quartilsabstand von (y_1, \dots, 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 $k_2 = [7.5] = 7$

$$\begin{aligned} \tilde{y}_{0.25} &= y_{(k_1+1)} = y_{(3)} = -0.5 \\ \tilde{y}_{0.75} &= y_{(k_2+1)} = y_{(8)} = 8.1 \end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 8.6$.

Aufgabe A2

Ein Autohändler verkauft je Tag X Autos des Typs **1** und Y Autos des Typs **2**. Die Tabelle zeigt die gemeinsame Zähldichte $f_{X,Y}(i, j) = \mathbb{P}(X = i, Y = j)$ von X und Y für $i, j = 0, 1, 2$.

		j		
		0	1	2
i	0	0.1	0.1	0.0
	1	0.1	0.3	0.1
	2	0.0	0.2	0.1

Es ist also z.B. $\mathbb{P}(X = 2, Y = 1) = f_{X,Y}(2, 1) = 0.2$.

- Bestimmen Sie die Zähldichten f_X von X und f_Y von Y .
- Berechnen Sie die bedingte Wahrscheinlichkeit $\mathbb{P}(X \geq 1 \mid Y = 1)$.
- Berechnen Sie die Erwartungswerte $\mathbb{E}X$ und $\mathbb{E}Y$ und die Varianz $V(X)$.
- Berechnen Sie die Kovarianz $C(X, Y)$ von X und Y . Sind X und Y positiv, negativ oder unkorreliert?

Lösung:

- Die Bestimmung der Zähldichte von X bzw. Y geschieht durch Summation der Wahrscheinlichkeiten der gemeinsamen Verteilung über die möglichen Werte von Y bzw. X (vergl. Skriptum Beispiel 6.7 und Tabelle 6.1)

		j			Σ	$f_X(i)$
		0	1	2		
i	0	0.1	0.1	0.0	0.2	
	1	0.1	0.3	0.1	0.5	
	2	0.0	0.2	0.1	0.3	
Σ	0.2	0.6	0.2	1.0		
		$f_Y(j)$				

Damit ergibt sich für die Zähldichten f_X und f_Y

i	0	1	2
$f_X(i)$	0.2	0.5	0.3
$f_Y(i)$	0.2	0.6	0.2

- Nach Definition gilt $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, falls $\mathbb{P}(B) > 0$. Da $A \rightarrow \mathbb{P}(A|B)$ ein Wahrscheinlichkeitsmaß ist (Satz 10.11), gilt

$$\begin{aligned} \mathbb{P}(X \geq 1 \mid Y = 1) &= \mathbb{P}(X = 1 \mid Y = 1) + \mathbb{P}(X = 2 \mid Y = 1) \\ &= \frac{\mathbb{P}(X = 1, Y = 1)}{\mathbb{P}(Y = 1)} + \frac{\mathbb{P}(X = 2, Y = 1)}{\mathbb{P}(Y = 1)} = \frac{0.3}{0.6} + \frac{0.2}{0.6} = \frac{5}{6} = 0.8333 \end{aligned}$$

c) Es gilt

$$\begin{aligned}\mathbb{E}X &= 0 \cdot f_X(0) + 1 \cdot f_X(1) + 2 \cdot f_X(2) = 0 \cdot 0.2 + 1 \cdot 0.5 + 2 \cdot 0.3 = 1.1 \\ \mathbb{E}X^2 &= 0^2 \cdot f_X(0) + 1^2 \cdot f_X(1) + 2^2 \cdot f_X(2) = 0 \cdot 0.2 + 1 \cdot 0.5 + 4 \cdot 0.3 = 1.7\end{aligned}$$

und damit dann $V(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 1.7 - 1.1^2 = 0.49$.

Analog

$$\mathbb{E}Y = 0 \cdot f_Y(0) + 1 \cdot f_Y(1) + 2 \cdot f_Y(2) = 0 \cdot 0.2 + 1 \cdot 0.6 + 2 \cdot 0.2 = 1.0$$

d) Es gilt $C(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}X \cdot \mathbb{E}Y$. Aus der Tabelle ergibt sich

$$\begin{aligned}\mathbb{E}(X \cdot Y) &= 1 \cdot 1 \cdot f_{X,Y}(1, 1) + 1 \cdot 2 \cdot f_{X,Y}(1, 2) + 2 \cdot 1 \cdot f_{X,Y}(2, 1) + 2 \cdot 2 \cdot f_{X,Y}(2, 2) \\ &= 1 \cdot 0.3 + 2 \cdot 0.1 + 2 \cdot 0.2 + 4 \cdot 0.1 = 1.3\end{aligned}$$

und damit $C(X, Y) = 1.3 - 1.1 \cdot 1.0 = 0.20$, damit ist dann auch $\rho(X, Y) > 0$ und X und Y sind positiv korreliert.

Aufgabe A3

Ein Unternehmen beschäftigt 24 Verkäufer, deren tägliche Umsätze voneinander unabhängig und normalverteilt seien. 20 dieser Verkäufer erzielen dabei einen mittleren täglichen Umsatz von 800 € bei einer Standardabweichung von 240 €, während die restlichen 4 einen mittleren Umsatz von 1000 € bei einer Standardabweichung von 300 € erreichen. Sei Z der tägliche Gesamtumsatz aller Verkäufer.

- Man berechne den Erwartungswert und die Varianz von Z .
- Welche Verteilung besitzt $\bar{Z} = \frac{Z}{24}$, der tägliche mittlere Umsatz aller Verkäufer?
- Mit welcher Wahrscheinlichkeit p ist der tägliche Gesamtumsatz Z größer als 19600 €?
- Bestimmen Sie denjenigen Gesamtumsatz t , so dass $\mathbb{P}(Z > t) = 0.025$ gilt.

Lösung:

- a) Seien $Z_1, \dots, Z_{20} \sim \mathcal{N}(800, 240^2)$ die täglichen Einzelumsätze der 20 Verkäufer und $Z_{21}, \dots, Z_{24} \sim \mathcal{N}(1000, 300^2)$ die täglichen Einzelumsätze der restlichen 4 Verkäufer. Dann gilt

$$\mathbb{E}Z = \mathbb{E}(Z_1 + \dots + Z_{24}) = \mathbb{E}Z_1 + \dots + \mathbb{E}Z_{24} = 20 \cdot 800 + 4 \cdot 1000 = 20000$$

und wegen der Unabhängigkeit der Einzelumsätze nach Satz 12.23 f)

$$V(Z) = V(Z_1 + \dots + Z_{24}) = V(Z_1) + \dots + V(Z_{24}) = 20 \cdot 240^2 + 4 \cdot 300^2 = 1512000.$$

Damit gilt $\mathbb{E}Z = 20000$ und $V(Z) = 1512000$.

- b) Wegen der Unabhängigkeit der Einzelumsätze gilt nach der Tabelle auf S. 114

$$\begin{aligned} Z &= Z_1 + \dots + Z_{24} \\ &\sim \mathcal{N}(800, 240^2) * \dots * \mathcal{N}(800, 240^2) * \mathcal{N}(1000, 300^2) * \dots * \mathcal{N}(1000, 300^2) \\ &\sim \mathcal{N}(20 \cdot 800 + 4 \cdot 1000, 20 \cdot 240^2 + 4 \cdot 300^2) = \mathcal{N}(20000, 1512000) \end{aligned}$$

Nach Satz 7. gilt dann

$$\bar{Z} = \frac{1}{24} \cdot Z \sim \mathcal{N}\left(\frac{20000}{24}, \frac{1512000}{24^2}\right) = \mathcal{N}(833.33, 2625)$$

- c) Zu berechnen ist $p = \mathbb{P}(Z > 19600)$.

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

$$\begin{aligned} \mathbb{P}(Z > 19600) &= 1 - \mathbb{P}(Z \leq 19600) = 1 - \Phi\left(\frac{19600 - 20000}{\sqrt{1512000}}\right) \\ &= 1 - \Phi(-0.3253) = \Phi(0.3253) \approx \Phi(0.33) = 0.6293 \end{aligned}$$

nach Anhang A.1 im Skriptum. Die gesuchte Wahrscheinlichkeit ist also

$$p = 0.6293$$

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

$$0.975 = \mathbb{P}(Z \leq t) = \Phi\left(\frac{t - 20000}{\sqrt{1512000}}\right)$$

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

$$1.96 = \frac{t - 20000}{\sqrt{1512000}}$$

also

$$t = 20000 + 1.96 \cdot \sqrt{1512000} = 22410.$$

Aufgabe A4

Ein Radiobastler hört an drei Tagen seinen Lieblingshit von 4-minütiger Dauer auf einem alten Radioempfänger. Die zufällige Anzahl von Knacklauten, die der Empfänger während einer Zeit von t Minuten produziert, genüge einer Poissonverteilung $Po(\lambda \cdot t)$ mit dem festen Parameter $\lambda = 0.2$. Die Anzahlen an den drei Tagen seien stochastisch unabhängig voneinander.

- Welche Verteilung besitzt die zufällige Anzahl von Knacklauten während einer Übertragung des Lieblingshits?
- Mit welcher Wahrscheinlichkeit p_1 wird während einer 4-minütigen Übertragung des Lieblingshits eines Hörers kein Knacken das Vergnügen beeinträchtigen?
- Bestimmen Sie die Wahrscheinlichkeit p_2 , dass der Radiobastler an allen drei Tagen zusammen mindestens 3 Knacklaute während seiner Lieblingshits hört.
- Berechnen Sie die Wahrscheinlichkeit p_3 , dass der Radiobastler an keinem der drei Tage seinen Lieblingshit ohne Knacklaut hören kann.

Lösung:

- a) Sei X_1 bzw. X_2, X_3 die zufällige Anzahl von Knacklauten während des Lieblingshits am ersten, bzw. zweiten und dritten Tag. Da der Lieblingshit 4 Minuten dauert, gilt

$$X_1 \sim Po(0.8).$$

(Analog gilt $X_2 \sim Po(0.8)$ und $X_3 \sim Po(0.8)$.)

- b) Gesucht ist $p_1 = \mathbb{P}(X_1 = 0)$. Wegen a) gilt

$$p_1 = \mathbb{P}(X_1 = 0) = e^{-0.8} = 0.4493$$

- c) Nach Voraussetzung sind X_1, X_2 und X_3 stochastisch unabhängig und daher gilt nach der Faltungsformel für die Gesamtzahl der Knacklaute

$$X = X_1 + X_2 + X_3 \sim Po(0.8 + 0.8 + 0.8) = Po(2.4)$$

Gesucht ist $p_2 = \mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X \leq 2)$. Da X die Zähldichte

$$f_X(k) = e^{-2.4} \cdot \frac{2.4^k}{k!}, k = 0, 1, \dots$$

besitzt, gilt

$$\begin{aligned} \mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = f_X(0) + f_X(1) + f_X(2) \\ &= e^{-2.4} + e^{-2.4} \cdot \frac{2.4}{1} + e^{-2.4} \cdot \frac{2.4^2}{2} = 0.5697 \end{aligned}$$

und damit $p_2 = \mathbb{P}(X \geq 3) = 1 - 0.5697 = 0.4303$.

- d) Jetzt ist $p_3 = \mathbb{P}(X_1 \geq 1, X_2 \geq 1, X_3 \geq 1)$ gesucht. Wegen der Unabhängigkeit von X_1, X_2 und X_3 gilt

$$\begin{aligned} p_3 &= \mathbb{P}(X_1 \geq 1, X_2 \geq 1, X_3 \geq 1) = \mathbb{P}(X_1 \geq 1) \cdot \mathbb{P}(X_2 \geq 1) \cdot \mathbb{P}(X_3 \geq 1) \\ &= \mathbb{P}(X_1 \geq 1)^3 = (1 - \mathbb{P}(X_1 = 0))^3 \stackrel{b)}{=} (1 - 0.4493)^3 = 0.1670 \end{aligned}$$

da X_1, X_2 und X_3 alle dieselbe Verteilung $Po(0.8)$ besitzen.

Aufgabe A5

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

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

und unbekanntem Parameter $\vartheta \in \mathbb{R}$.

- Bestimmen Sie die Loglikelihood-Funktion $M_x(\vartheta)$ zur Stichprobe $x = (x_1, \dots, x_n)$, wobei $x_i > 0$ für $i = 1, \dots, n$.
- Berechnen Sie die Ableitung $M'_x(\vartheta)$.
- Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ zur Stichprobe x .
- Sei Y eine Zufallsvariable mit der Verteilung $\mathcal{LN}(\vartheta, 4)$. Welche Verteilung besitzt die Zufallsvariable $Z := \ln(Y)$, ihr natürlicher Logarithmus?
- Bestimmen Sie den Momentenschätzer $\hat{\vartheta}_m(x)$ für ϑ zur Stichprobe $x = (x_1, \dots, x_n)$.

Lösung:

- a) Mit

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

ergibt sich die Loglikelihood-Funktion zu

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

- b) Ihre Ableitung ist

$$M'_x(\vartheta) = -\frac{1}{8} \sum_{i=1}^n -2 \cdot (\ln(x_i) - \vartheta) = \frac{1}{4} \sum_{i=1}^n (\ln(x_i) - \vartheta) = \frac{1}{4} \left(\sum_{i=1}^n \ln(x_i) - n \cdot \vartheta \right)$$

- c) Da die zweite Ableitung

$$M''_x(\vartheta) = -\frac{n}{4} < 0.$$

von M_x überall kleiner als 0 ist, ergibt sich die Maximumstelle $\hat{\vartheta}(x)$ von $\vartheta \rightarrow M_x(\vartheta)$ als Lösung von $M'_x(\vartheta) = 0$, also $\sum_{i=1}^n \ln(x_i) = n \cdot \vartheta$ und

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

- d) Bekannt ist (Skriptum 9.10), dass für eine $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariable Z die Zufallsvariable $Y = e^Z$ die logarithmische Normalverteilung $\mathcal{LN}(\mu, \sigma^2)$ besitzt. Besitzt dann umgekehrt Y die Verteilung $\mathcal{LN}(\mu, \sigma^2)$, so besitzt dann $Z = \ln(Y)$ die Verteilung $\mathcal{N}(\mu, \sigma^2)$. Hier ist $\mu = \vartheta$ und $\sigma^2 = 4$. Also besitzt Z die Verteilung $\mathcal{N}(\vartheta, 4)$.

e) 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 $\mathbb{E}_\vartheta(X_1) = e^{\vartheta+1/2 \cdot 4} = e^{\vartheta+2}$ 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 $\hat{\vartheta}_m(x) := \ln(\bar{x}) - 2$.