

Klausur
Wahrscheinlichkeitstheorie und Statistik
vom 2.3.2007
Musterlösungen

Aufgabe B1:

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	0.9	2.1	3	3.7	5	6.1	7	7.9	8.9	9.8
y_j	8.4	7	6.9	5.2	2.5	2.9	3.9	3.8	0.8	-3.1

- 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

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

$$\bar{x} = 5.44$$

$$s_x = 2.997$$

$$\bar{y} = 3.83$$

$$s_y = 3.368$$

$$r_{xy} = -0.9016$$

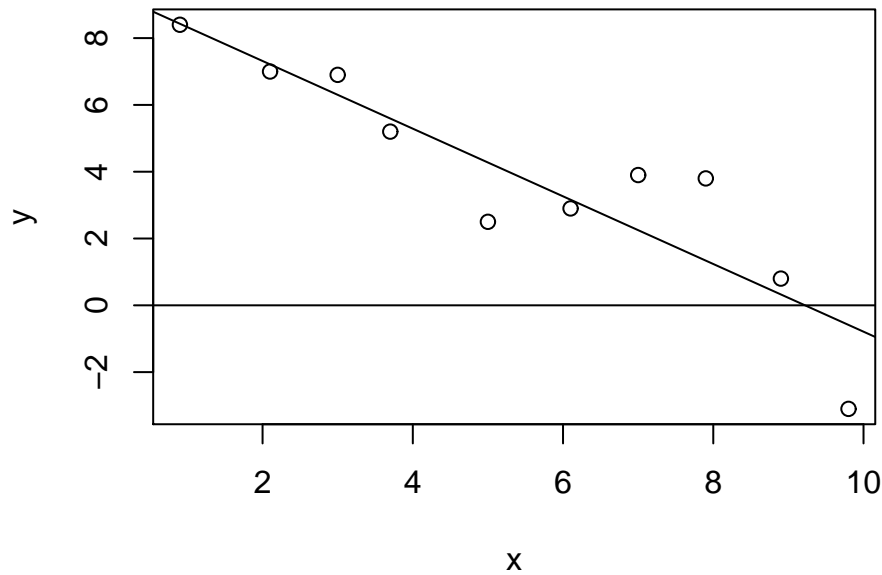
- b) Bestimmen Sie die zugehörige Regressionsgerade $y = a^* + b^* \cdot x$ von y auf x .

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

$$b^* = -1.013$$

$$a^* = 9.34$$

und die Regressionsgerade $y = 9.34 - 1.013 \cdot x$.



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

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

$$y_{(j)} = (-3.1, 0.8, 2.5, 2.9, 3.8, 3.9, 5.2, 6.9, 7, 8.4)$$

- c) Berechnen Sie das 0.2-getrimmte Stichprobenmittel $\bar{y}_{0.2}$ von (y_1, \dots, y_{10}) .
Lösung: Mit $k = [10 \cdot 0.2] = 2$ ergibt sich

$$\bar{y}_{0.2} = \frac{1}{10 - 2 \cdot 2} \cdot (y_{(3)} + \dots + y_{(8)}) = 4.2$$

- d) Bestimmen Sie das Stichproben-0.2-Quantil $\tilde{y}_{0.2}$ von (y_1, \dots, y_{10}) .
Lösung: Da $10 \cdot 0.2 = 2$ ganzzahlig ist, ist mit $k = [2] = 2$

$$\tilde{y}_{0.2} = \frac{y_{(k)} + y_{(k+1)}}{2} = \frac{y_{(2)} + y_{(3)}}{2} = 1.65$$

Aufgabe B2:

Die Zufallsvariable X besitze die Verteilung $\mathcal{N}(3, 16)$ und die Zufallsvariable Y die Verteilung $\mathcal{N}(-8, 9)$. X und Y seien stochastisch unabhängig.

- Welche Verteilung besitzt $Z := X + Y$ und welche Verteilung besitzt $U := X - Y$?
- Bestimmen Sie die Varianz $V(U)$ von U .
- Berechnen Sie die Kovarianz $C(U, Y)$ von U und Y . Sind U und Y positiv korreliert, negativ korreliert oder unkorreliert?
- Bestimmen Sie $\mathbb{P}(X \geq 7)$.
- Berechnen Sie das 0.975-Quantil von Y .

Lösung:

- a) Es ist $-Y \sim \mathcal{N}(8, 9)$, also nach der Faltungsformel 11.16

$$\begin{aligned} Z &\sim \mathcal{N}(3 - 8, 9 + 16) = \mathcal{N}(-5, 25), \\ U = X + (-Y) &\sim \mathcal{N}(3 + 8, 9 + 16) = \mathcal{N}(11, 25). \end{aligned}$$

- b) Der zweite Parameter von $\mathcal{N}(\mu, \sigma^2)$ ist die Varianz dieser Verteilung, also wegen b) $V(Z) = 25$.

- c) Wegen Satz 12.23 und der Unabhängigkeit von X und Y gilt

$$C(U, Y) = C(X - Y, Y) = C(X, Y) - C(Y, Y) = 0 - V(Y) = -9.$$

Hieraus erhält man ohne Rechnung, dass $\rho(U, Y) < 0$ ist. Daher sind U und Y negativ korreliert.

- d) Wegen der Stetigkeit von $\mathcal{N}(3, 16)$ gilt nach (9.6) und Tabelle A.1

$$\begin{aligned} \mathbb{P}(X \geq 7) &= \mathbb{P}(X > 7) = 1 - \mathbb{P}(X \leq 7) = 1 - \Phi\left(\frac{7 - 3}{4}\right) \\ &= 1 - \Phi(1) = 1 - 0.8413 = \underline{0.1587}. \end{aligned}$$

- e) Nach Definition 12.19 gilt, dass $t_{0.975}$ die Lösung ist von

$$\Phi_{-3,16}(t_{0.975}) = \Phi\left(\frac{t_{0.975} - (-8)}{3}\right) = 0.975.$$

Wegen $\Phi(1.96) = 0.975$ gilt also

$$\frac{t_{0.975} + 8}{3} = 1.96$$

und daraus $t_{0.975} = 3 \cdot 1.96 - 8 = \underline{-2.12}$. (Man kann hierfür auch 12.20 c) benützen.)

Aufgabe B3:

Ein Arbeiter überwacht 3 Aggregate, die unabhängig voneinander arbeiten. Sei A_i das Ereignis, dass Aggregat i einer Überprüfung bedarf und $p_i = \mathbb{P}(A_i)$ die zugehörige Wahrscheinlichkeit, $i = 1, 2, 3$, mit $p_1 = 0.3$, $p_2 = 0.5$ und $p_3 = 0.7$. Die Ereignisse A_1, A_2 und A_3 seien stochastisch unabhängig.

B sei das Ereignis „Mindestens ein Aggregat bedarf einer Überprüfung“.

- a) Berechnen Sie $\mathbb{P}(A_1 \cap A_2 \cap A_3)$ und $\mathbb{P}(A_1 \cup A_2)$.
- b) Stellen Sie B formelmäßig mit Hilfe von A_1, A_2, A_3 dar.
- c) Berechnen Sie $\mathbb{P}(B)$.
- d) Berechnen Sie die (bedingte) Wahrscheinlichkeit $\mathbb{P}(A_1 | B)$, dass Aggregat 1 einer Überprüfung bedarf, wenn mindestens ein Aggregat einer Überprüfung bedarf.

Lösung: a) Da A_1, A_2 und A_3 unabhängig sind, gilt

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) = 0.105.$$

Ferner gilt

$$\begin{aligned} \mathbb{P}(A_1 \cup A_2) &= \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \\ &\stackrel{A_1, A_2 \text{ unabhängig}}{=} p_1 + p_2 - \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) = p_1 + p_2 - p_1 p_2 = 0.65 \end{aligned}$$

b) Dass mindestens ein Aggregat der Überprüfung bedarf, ist gleichbedeutend damit, dass Aggregat 1 **oder** 2 **oder** 3 einer Überprüfung bedarf. Daher gilt

$$B = A_1 \cup A_2 \cup A_3.$$

c) Wegen der Formel des Ein- und Ausschließens gilt

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_2 \cap A_3) + \mathbb{P}(A_1 \cap A_2 \cap A_3) \\ &= p_1 + p_2 + p_3 - p_1 p_2 - p_1 p_3 - p_2 p_3 + p_1 p_2 p_3 = 0.895 \end{aligned}$$

d) Nach Definition der bedingten Wahrscheinlichkeit gilt wegen $A_1 \subset B$ (wenn 1 einer Überprüfung bedarf, bedarf insbesondere mindestens ein Aggregat der Überprüfung)

$$\mathbb{P}(A_1 | B) = \frac{\mathbb{P}(A_1 \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_1)}{\mathbb{P}(B)} = \frac{0.3}{0.895} = 0.3352$$

Aufgabe B4:

Eine Messeinrichtung habe eine zufällige Lebensdauer mit Verteilung $Exp(\alpha)$, $\alpha > 0$. Immer wenn eine Messeinrichtung ausfällt, wird sie durch eine gleichartige neue ersetzt. Das Ersetzen dauert genau 2 Zeiteinheiten. Die Lebensdauern X_1, X_2, \dots der einzelnen Messeinrichtungen seien unabhängig voneinander und identisch verteilt.

$$Z_n = \sum_{i=1}^n X_i + 2n$$

ist damit der Zeitpunkt, zu dem das $(n+1)$ -te Messgerät mit den Messungen beginnt. Damit hat $Z_n - 2 \cdot n$ die Verteilung $\Gamma(n, \alpha)$.

- a) Bestimmen Sie $\mathbb{P}(Z_1 > t)$ für $t > 2$.
- b) Bestimmen Sie die Verteilungsfunktion $F_{Z_2}(t)$ und eine Dichte $f_{Z_2}(t)$ von Z_2 für $t > 4$.
Hinweis: Die $\Gamma(2, \alpha)$ -Verteilung hat die Verteilungsfunktion

$$t \mapsto 1 - e^{-\alpha t} \cdot (1 + \alpha t), \quad t \geq 0.$$

c) Die Ersetzung einer Messeinrichtung kostet a Euro, ferner fallen pro Zeiteinheit $c > 0$ Euro laufende Kosten an (auch während der Zeit, wenn eine Messeinrichtung ersetzt wird).

c₁) Die (zufälligen) Kosten K_n bis zum Zeitpunkt Z_n sind von der Form

$$K_n = d_n + b_n \cdot \sum_{i=1}^n X_i$$

mit geeigneten Konstanten b_n, d_n . Bestimmen Sie b_n und d_n .

c₂) Berechnen Sie $\mathbb{E}K_n$ und $V(K_n)$.

Lösung: $Z_n - 2n = \sum_{i=1}^n X_i$ ist die n -fache Faltung der $Exp(\alpha) = \Gamma(1, \alpha)$ -verteilten Zufallsvariablen X_1, \dots, X_n . Wegen 11.16 gilt daher

$$Z_n - 2n \sim \Gamma(n, \alpha).$$

a) $\mathbb{P}(Z_1 > t) = \mathbb{P}(X_1 + 2 > t) = \mathbb{P}(X_1 > t - 2) = 1 - (1 - e^{-\alpha(t-2)}) = e^{-\alpha(t-2)}$ für $t > 2$, also $t - 2 > 0$.

b) Wegen a) hat $Y := Z_2 - 4$ die Verteilung $\Gamma(2, \alpha)$. Insbesondere gilt $Y > 0$ und damit

$$F_{Z_2}(t) = \mathbb{P}(Z_2 \leq t) = \mathbb{P}(Y + 4 \leq t) = \mathbb{P}(Y \leq t - 4) = 0$$

für $t \leq 4$. Für $t > 4$ gilt dagegen

$$F_{Z_2}(t) = \mathbb{P}(Y \leq t - 4) = 1 - e^{-\alpha(t-4)} \cdot (1 + \alpha(t-4)).$$

Diese Verteilungsfunktion ist stetig und bis auf eventuell die Stelle $t = 4$ stetig differenzierbar (ob F_{Z_2} auch an der einzelnen Stelle $t = 4$ differenzierbar ist, muss nicht überprüft werden). Wegen Satz 8.12 ist daher

$$f_{Z_2}(t) = \begin{cases} 0 & , t \leq 4 \\ F'_{Z_2}(t) = \alpha e^{-\alpha(t-4)} \cdot (1 + \alpha(t-4)) - \alpha e^{-\alpha(t-4)} = \alpha^2(t-4)e^{-\alpha(t-4)} & , t > 4 \end{cases}$$

eine Dichte von Z_2 .

c) Bis zum Zeitpunkt Z_n liegen n Ersetzungen vor, daher sind die Kosten

$$K_n = n \cdot a + c \cdot Z_n = n \cdot (a + 2 \cdot c) + c \cdot \sum_{i=1}^n X_i,$$

wobei $\sum_{i=1}^n X_i \sim \Gamma(n, \alpha)$. Damit gilt c₁) $b_n = c$ und $d_n = n \cdot (a + 2 \cdot c)$.

c₂) Mit den Tabellen auf S. 124 und S. 128 ergibt sich

$$\mathbb{E}K_n = \mathbb{E} \left(n \cdot (a + 2 \cdot c) + c \cdot \sum_{i=1}^n X_i \right) = n \cdot (a + 2 \cdot c) + c \cdot \mathbb{E} \sum_{i=1}^n X_i = n \cdot (a + 2 \cdot c) + c \frac{n}{\alpha}$$

und wegen Satz 12.11 c)

$$V(K_n) = V \left(n \cdot (a + 2 \cdot c) + c \cdot \sum_{i=1}^n X_i \right) = V \left(c \cdot \sum_{i=1}^n X_i \right) = c^2 \cdot V \left(\sum_{i=1}^n X_i \right) = \frac{c^2 \cdot n}{\alpha^2}.$$

Aufgabe B5:

Bei $n = 200$ Kontrollen wurde die zufällige Fehlerzahl x_i bei jeweils 20 Vorgängen ermittelt, $i = 1, \dots, n$. Man nimmt an, dass die zufälligen Zahlen X_i der Fehler stochastisch unabhängig sind mit der Zähldichte

$$f_{\vartheta}(k) = \binom{20}{k} \cdot \left(\frac{\vartheta}{20}\right)^k \cdot \left(1 - \frac{\vartheta}{20}\right)^{20-k}, \quad k = 0, \dots, 20,$$

ϑ ein unbekannter Parameter mit $0 < \vartheta < 20$.

- Welche Verteilung besitzen die zufälligen Zahlen der Fehler X_i , d.h. welche Verteilung hat gerade die Zähldichte f_{ϑ} ? (Geben Sie auch die Parameter an!)
- Bestimmen Sie die Loglikelihood-Funktion $M_x(\vartheta)$ zur Stichprobe $x = (x_1, \dots, x_n)$.
- Berechnen Sie einen Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ .
- Insgesamt stellte man
 - 38 mal 0 Fehler,
 - 60 mal 1 Fehler,
 - 67 mal 2 Fehler,
 - 24 mal 3 Fehler,
 - 8 mal 4 Fehler und
 - 3 mal 5 Fehler fest.

Berechnen Sie das Stichprobenmittel \bar{x} und die Stichprobenvarianz s_x^2 .

Lösung: a) Es ist $f_{\vartheta}(k) = \binom{20}{k} \cdot p^k \cdot (1-p)^{20-k}$ mit $p := \frac{\vartheta}{20}$ und dies ist gerade die Zähldichte von $Bin(20, p)$. Daher besitzen die Zufallsvariablen X_i die Verteilung $Bin(20, \frac{\vartheta}{20})$.

b) Wir berechnen die Loglikelihood-Funktion $M_x(\vartheta)$. Es gilt $\ln(f_{\vartheta}(k)) = \ln \binom{20}{k} + k \cdot \ln(\vartheta/20) + (20-k) \cdot \ln(1 - \vartheta/20)$ und damit

$$\begin{aligned} M_x(\vartheta) &= \sum_{i=1}^n \ln(f_{\vartheta}(x_i)) = \sum_{i=1}^n \left[\ln \binom{20}{x_i} + x_i \cdot \ln(\vartheta/20) + (20 - x_i) \cdot \ln(1 - \vartheta/20) \right] \\ &= \sum_{i=1}^n \ln \binom{20}{x_i} + \ln(\vartheta/20) \cdot \sum_{i=1}^n x_i + \ln(1 - \vartheta/20) \cdot \sum_{i=1}^n (20 - x_i) \end{aligned}$$

c) M_x besitzt die Ableitung

$$\begin{aligned} M'_x(\vartheta) &= \frac{1}{\vartheta} \cdot \sum_{i=1}^n x_i - \frac{1}{20 - \vartheta} \cdot \sum_{i=1}^n (20 - x_i) = \frac{n \cdot \bar{x}}{\vartheta} - \frac{n \cdot (20 - \bar{x})}{20 - \vartheta} \\ &= \frac{n \cdot \bar{x} \cdot (20 - \vartheta) - n \cdot (20 - \bar{x}) \cdot \vartheta}{\vartheta \cdot (20 - \vartheta)} = \frac{n \cdot \bar{x} \cdot 20 - n \cdot 20 \cdot \vartheta}{\vartheta \cdot (20 - \vartheta)} = \frac{20 \cdot n}{\vartheta \cdot (20 - \vartheta)} \cdot (\bar{x} - \vartheta) \end{aligned}$$

wobei wir $\sum_{i=1}^n x_i = n \cdot \bar{x}$ benützt haben. Hierbei ist $\frac{20-n}{\vartheta \cdot (20-\vartheta)} > 0$ und damit

$$M'_x(\vartheta) \begin{cases} > \\ = \\ < \end{cases} 0 \quad \text{genau dann, wenn} \quad \vartheta \begin{cases} < \\ = \\ > \end{cases} \bar{x}.$$

Dies bedeutet, dass die Funktion M_x unterhalb von $\hat{\vartheta}(x) = \bar{x}$ steigt und oberhalb von $\hat{\vartheta}(x) = \bar{x}$ fällt. Damit ist $\hat{\vartheta}(x)$ die (einzige) Maximumstelle von M_x und $\hat{\vartheta}(x) = \bar{x}$ der gesuchte Maximum-Likelihood-Schätzer.

d) Zur Berechnung von \bar{x} und s_x^2 kann Satz 1.10 verwendet werden. Alternativ berechnet man direkt $\sum_{i=1}^{200} x_i = 38 \cdot 0 + 60 \cdot 1 + \dots + 3 \cdot 5 = 313$ und $\sum_{i=1}^{200} x_i^2 = 38 \cdot 0^2 + 60 \cdot 1^2 + \dots + 3 \cdot 5^2 = 747$ und damit

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{1}{200} \cdot \sum_{i=1}^{200} x_i = \frac{313}{200} = \underline{1.565} \quad \text{und}$$

$$s_x^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right] = \frac{1}{199} \left(747 - \frac{313^2}{200} \right) = \underline{1.2922}.$$

Aufgabe B6:

Eine Firma besitzt 16 gleichartige Messgeräte, welche unabhängig voneinander und unter gleichen Bedingungen 2 Jahre lang ohne Wartung eingesetzt wurden. Bei der Überprüfung dieser Messgeräte stellt sich heraus, dass 10 dieser Messgeräte keine korrekten Messergebnisse mehr liefern.

- a) Es sei p die unbekannte Wahrscheinlichkeit, dass ein Messgerät auch nach 2 Jahren noch korrekte Messergebnisse liefert. Geben Sie ein Konfidenzintervall für die Wahrscheinlichkeit p zur Konfidenzwahrscheinlichkeit 0.95 an.

Lösung: Nach Voraussetzung kann wie in Beispiel 18.5 ein ideales Zufallsexperiment mit den zwei möglichen Ergebnissen „Keine korrekten Messergebnisse“ (Treffer) und „Korrekte Messergebnisse“ (Niete) und der Trefferwahrscheinlichkeit $\vartheta := p$ angesehen werden. Nach diesem Beispiel und wegen 18.6 ist mit $n = 16$ und $x = 10$ das gesuchte Konfidenzintervall $[l(x), L(x)]$, wobei die Konfidenzgrenzen $l(x)$ und $L(x)$ für $x = 10$ und $n - x = 6$ und $1 - \alpha = 0.95$ aus Tabelle A.4 entnommen werden. Dies ergibt das Konfidenzintervall

$$\mathcal{C}_2(x) = [0.354, 0.848].$$

- b) Sie suchen jetzt ein Konfidenzintervall für p zum Konfidenzniveau 0.90. Ist dieses kleiner, gleich oder größer als das zur Konfidenzwahrscheinlichkeit 0.95?

Lösung: Für das Konfidenzintervall zur Konfidenzwahrscheinlichkeit 0.90 erhält man aus Tabelle A.4

$$\mathcal{C}_1(x) = [0.391, 0.822].$$

Dieses ist kleiner als $\mathcal{C}_1(x)$. (Dies gilt allgemein: Wird die Konfidenzwahrscheinlichkeit (die Sicherheit) erniedrigt, so muss auch das Konfidenzintervall kleiner werden.)