INSTITUT FÜR STOCHASTIK
UNIVERSITÄT KARLSRUHE
Priv.-Doz. Dr. D. Kadelka
Dipl.-Math. W. Lao

SS 2007

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 12.9.2007

Musterlösungen

Aufgabe A1:

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

j	1	2	3	4	5	6	7	8	9	10
x_j	0.9	1.8	3	4.1	5.1	6	7.5	8.1	9.2	10.1
y_j	5.5	2.2	2.1	−0.5	−1.4	−2.3	−6.7	−5	−7.3	−11.1

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

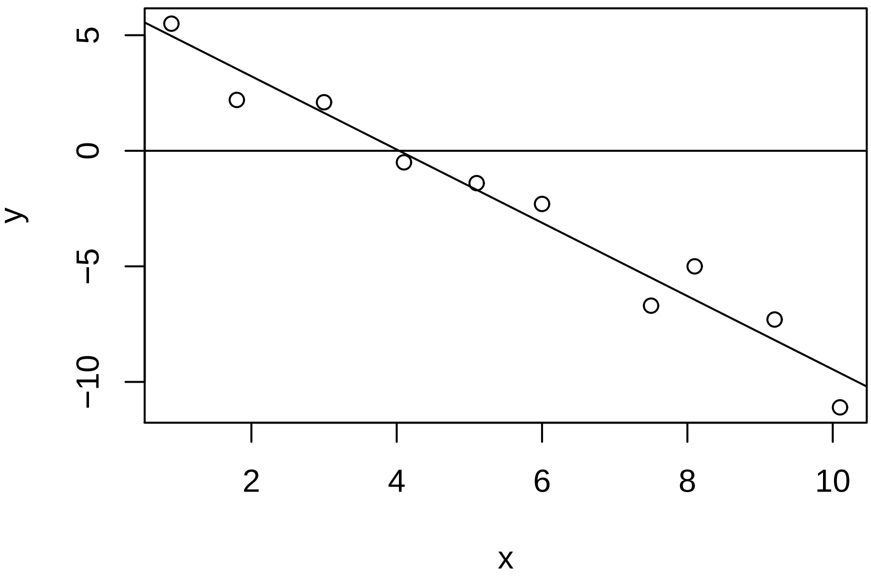
$$\sum_{j=1}^{n} (x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 5.58 \qquad s_x = 3.145$$
$$\bar{y} = -2.45 \qquad s_y = 5.092$$
$$r_{xy} = -0.9789$$

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

$$b^* = -1.585$$
$$a^* = 6.39$$

und die Regressionsgerade $y = 6.39 - 1.585 \cdot x$.



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

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

$$y_{()} = (-11.1,\ -7.3,\ -6.7,\ -5,\ -2.3,\ -1.4,\ -0.5,\ 2.1,\ 2.2,\ 5.5)$$

c) Bestimmen Sie das Stichproben-0.1-Quantil $\tilde{y}_{0.1}$ von $(y_1, \ldots, y_{10})$.
Lösung: Da $10 \cdot 0.1 = 1$ ganzzahlig ist, ist mit $k = [1] = 1$

$$\tilde{y}_{0.1} = \frac{y_{(k)} + y_{(k+1)}}{2} = \frac{y_{(1)} + y_{(2)}}{2} = -9.2$$

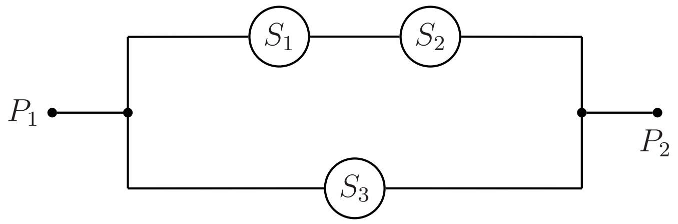
d) Berechnen Sie den Quartilsabstand von $(y_1, \ldots, y_{10})$.
Lösung: Da $0.25 \cdot 10 = 2.5$ und $0.75 \cdot 10 = 7.5$ beide nicht ganzzahlig sind, ergibt sich mit $k_1 = [2.5] = 2$ und $k_2 = [7.5] = 7$

$$\begin{aligned} \tilde{y}_{0.25} &= y_{(k_1+1)} = y_{(3)} = -6.7 \\ \tilde{y}_{0.75} &= y_{(k_2+1)} = y_{(8)} = 2.1 \end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 8.8$.

Aufgabe A2

Zwischen 2 Punkten P_1 und P_2 verläuft folgendes Leitungsnetz:



Dabei sind S_1, S_2, S_3 störanfällige Stellen. Die Zufallsvariablen X_i seien definiert als

$$X_i = \begin{cases} 0, & S_i \text{ ist unterbrochen,} \\ 1, & S_i \text{ ist nicht unterbrochen,} \end{cases} \qquad i = 1, 2, 3.$$

$A = (\{X_1 = 1\} \cap \{X_2 = 1\}) \cup \{X_3 = 1\}$ ist also das Ereignis „P_1 ist mit P_2 verbunden“. Dabei seien X_1, X_2, X_3 stochastisch unabhängig mit $\mathbb{P}(X_1 = 1) = \mathbb{P}(X_2 = 1) = \frac{1}{2}$ und $\mathbb{P}(X_3 = 1) = \frac{1}{4}$.

a) Bestimmen Sie die Wahrscheinlichkeit $\mathbb{P}(\{X_1 = 1\} \cap \{X_2 = 1\})$.

Lösung: Da X_1 und X_2 stochastisch unabhängig sind, gilt

$$\mathbb{P}(\{X_1 = 1\} \cap \{X_2 = 1\}) = \mathbb{P}(X_1 = 1) \cdot \mathbb{P}(X_2 = 1) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

b) Berechnen Sie $\mathbb{P}(A)$.

Lösung: Allgemein gilt $\mathbb{P}(B \cup C) = \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B) \cdot \mathbb{P}(C)$, wegen der Unabhängigkeit von $B := \{X_1 = 1\} \cap \{X_2 = 1\}$ und $C := \{X_3 = 1\}$ und $\mathbb{P}(C) = \mathbb{P}(X_3 = 1) = \frac{1}{4}$ hier also

$$\mathbb{P}(A) = \mathbb{P}(B \cup C) = \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(B) \cdot \mathbb{P}(C) = \frac{1}{4} + \frac{1}{4} - \frac{1}{4} \cdot \frac{1}{4} = \frac{7}{16}.$$

c) Berechnen Sie $\mathbb{P}(A \mid \{X_3 = 1\})$ und $\mathbb{P}(\{X_3 = 1\} \mid A)$.

Lösung: Nach Definition der bedingten Wahrscheinlichkeiten gilt

$$\mathbb{P}(A \mid \{X_3 = 1\}) = \mathbb{P}(B \cup C \mid C) = \frac{\mathbb{P}((B \cup C) \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(C)}{\mathbb{P}(C)} = 1$$

und nach der Formel von Bayes dann

$$\mathbb{P}(\{X_3 = 1\} \mid A) = \frac{\mathbb{P}(X_3 = 1) \cdot \mathbb{P}(A \mid \{X_3 = 1\})}{\mathbb{P}(A)} = \frac{1/4}{7/16} = \frac{4}{7}.$$

d) Bestimmen Sie die Verteilung von $Y := X_1 \cdot X_2$ und von $Z := X_1 \cdot X_2 + X_3$.
Hinweis: Y und Z sind beide binomialverteilt. Bestimmen Sie die Parameter.

Lösung: Y nimmt nur die Werte 0 uns 1 an. Dabei ist gilt genau dann $Y = 1$, wenn $X_1 = 1$ und $X_2 = 1$, andernfalls ist $Y = 0$. Wegen $\mathbb{P}(X_1 = 1, X_2 = 1) = \mathbb{P}(B) = \frac{1}{4}$ gilt $Y \sim Bin(1, 1/4)$. Da Y und X_3 unabhängig und beide $Bin(1, 1/4)$-verteilt sind, gilt wegen der Faltungsformel $Z = Y + X_3 \sim Bin(2, 1/4)$.

Aufgabe A3

Bei einer Platte haben die zufälligen Abweichungen X und Y der Länge von der Solllänge $l = 40$ und der Breite von der Sollbreite $b = 30$ die Normalverteilungen $\mathcal{N}(0, 16)$ bzw. $\mathcal{N}(0, 9)$. Es ist also $L := l + X$ die zufällige Länge und $B := b + Y$ die zufällige Breite der Platte. Die Zufallsvariablen X und Y seien stochastisch unabhängig.

a) Welche Verteilungen haben L, B und der Umfang der Platte $2 \cdot (L + B)$?
Lösung: Es gilt $L \sim \mathcal{N}(40, 16)$, $B \sim \mathcal{N}(30, 9)$. Da L und B unabhängig sind, gilt nach Faltungsformel

$$L + B \sim \mathcal{N}(70, 25),$$

somit nach Satz 9.7

$$2 \cdot (L + B) \sim \mathcal{N}(2 \cdot 70, 4 \cdot 25) = \mathcal{N}(140, 100).$$

b) Die zufällige Fläche der Platte ist $F = L \cdot B$. Bestimmen Sie den Erwartungswert von F.
Lösung: Da L und B unabhängig sind, gilt nach Satz 12.8 c)

$$\mathbb{E}(F) = \mathbb{E}(L \cdot B) = \mathbb{E}(L) \cdot \mathbb{E}(B) = 40 \cdot 30 = 1200.$$

c) Berechnen Sie die Kovarianz von F und B. Sind F und B positiv, null- oder negative korreliert?
Lösung: Es gilt

$$\begin{aligned} C(F, B) &= \mathbb{E}(F \cdot B) - \mathbb{E}(F) \cdot \mathbb{E}(B) \\ &= \mathbb{E}(L \cdot B^2) - \mathbb{E}(L \cdot B) \cdot \mathbb{E}(B) \\ &= \mathbb{E}(L) \cdot \mathbb{E}(B^2) - \mathbb{E}L \cdot (\mathbb{E}B)^2 \\ &= \mathbb{E}L \cdot (\mathbb{E}(B^2) - (\mathbb{E}B)^2) \\ &= \mathbb{E}L \cdot V(B) = 40 \cdot 9 = 360. \end{aligned}$$

Es ist $C(F, B) > 0$ und damit $\rho(F, B) > 0$. F und B sind also positiv korreliert.

d) Berechnen Sie das 0.975-Quantil von L.
Lösung:

$$\mathbb{P}(L \leq t) = \Phi\left(\frac{t - 40}{4}\right) \stackrel{!}{=} 0.975 \quad \Longleftrightarrow \quad \frac{t - 40}{4} \stackrel{!}{=} 1.96$$

Also ist $t_{0.975} = 1.96 \cdot 4 + 40 = 47.84$ das 0.975-Quantil von L.

e) Wie groß ist die Wahrscheinlichkeit, dass L um mehr als 6 von der Solllänge $l = 40$ abweicht? Bestimmen Sie diese Wahrscheinlichkeit auf 4 Nachkommastellen genau.
Lösung: Mit $\sigma = \sqrt{V(L)} = \sqrt{16} = 4$ und Folgerung 9.9 gilt

$$\begin{aligned} \mathbb{P}(|L - 40| > 6) &= 1 - \mathbb{P}(|L - l| \leq 1.5\sigma) \\ &= 1 - (2 \cdot \Phi(1.5) - 1) = 2 \cdot (1 - \Phi(1.5)) \\ &= 2 \cdot (1 - 0.9332) = 0.1336. \end{aligned}$$

Aufgabe A4

In einem Produktionsprozess müssen Bauteile $W := Z \cdot X$ Zeiteinheiten warten, bevor mit der Bearbeitung der Dauer B begonnen wird. Hierbei sind $B > 0, X > 0$ und Z stochastisch unabhängige Zufallsvariable. B hat die Verteilung $Exp(\mu)$, X die Verteilung $Exp(\alpha)$ und Z die Verteilung $Bin(1,p)$, wobei $0 < \alpha < \mu$ und $p \in (0,1)$.
Es ist also $W = 0$, wenn $Z = 0$ ist, und $W > 0$ für den Fall $Z = 1$.

a) Berechnen Sie den Erwartungswert $\mathbb{E}W$, ferner $\mathbb{E}W^2$ und (damit) die Varianz $V(W)$.

b) $S := W + B$ ist der zufällige Zeitpunkt des Endes der Bearbeitung eines Bauteils. Bestimmen Sie $\mathbb{E}S$ und $V(S)$.

c) Bestimmen Sie $\mathbb{P}(W = 0)$.

d) Bestimmen Sie die Verteilungsfunktion F_W von W.
Hinweis: Verwenden Sie die Formel von der totalen Wahrscheinlichkeit mit den Fällen $\{Z = 0\}$ und $\{Z = 1\}$.

Lösung:

a) Wegen der Unabhängigkeit von Z und X gilt

$$\mathbb{E}W = \mathbb{E}(Z \cdot X) = \mathbb{E}Z \cdot \mathbb{E}X = p \cdot \frac{1}{\alpha},$$

ferner gilt

$$\mathbb{E}(W^2) = \mathbb{E}(Z^2) \cdot \mathbb{E}(X^2) = \frac{2p}{\alpha^2}$$

wegen

$$\begin{aligned} \mathbb{E}(Z^2) &= V(Z) + (\mathbb{E}Z)^2 = p \cdot (1-p) + p^2 = p, \\ \mathbb{E}(X^2) &= V(X) + (\mathbb{E}X)^2 = \frac{1}{\alpha^2} + \frac{1}{\alpha^2} = \frac{2}{\alpha^2} \end{aligned}$$

und damit

$$\begin{aligned} V(W) &= \mathbb{E}(W^2) - (\mathbb{E}W)^2 = \mathbb{E}(Z^2 \cdot X^2) - (\mathbb{E}W)^2 \\ &= \mathbb{E}Z^2 \cdot \mathbb{E}X^2 - (\mathbb{E}W)^2 \\ &= \frac{2p}{\alpha^2} - \frac{p^2}{\alpha^2} = \frac{p}{\alpha^2}(2-p). \end{aligned}$$

b) $\mathbb{E}(S) = \mathbb{E}(W + B) = \frac{p}{\alpha} + \frac{1}{\mu}$. Wegen der Unabhängigkeit von W und B gilt zudem

$$V(S) = V(W) + V(B) = \frac{p}{\alpha^2}(2-p) + \frac{1}{\mu^2}.$$

c) Wegen $X > 0$ gilt $W = 0$ genau dann, wenn $Z = 0$. Daher gilt

$$\mathbb{P}(W = 0) = \mathbb{P}(Z = 0) = 1 - p.$$

d) Sei F die Verteilungsfunktion von W. Es gilt für $t \geq 0$

$$\begin{aligned}
F(t) &= \mathbb{P}(W \leq t) \\
&= \mathbb{P}(\underbrace{Z \cdot X}_{=0} \leq t, Z = 0) + \mathbb{P}(\underbrace{Z \cdot X}_{=X} \leq t, Z = 1) \\
&= \mathbb{P}(Z = 0) + \mathbb{P}(X \leq t, Z = 1) \\
&= \underbrace{\mathbb{P}(Z = 0)}_{=1-p} + \mathbb{P}(X \leq t) \cdot \underbrace{\mathbb{P}(Z = 1)}_{=p} \qquad (X, Z \text{ unabhängig}) \\
&= 1 - p + p \cdot (1 - e^{-\alpha t}) \\
&= 1 - p \cdot e^{-\alpha t}.
\end{aligned}$$

Für $t < 0$ gilt $F(t) = 0$ wegen $W \geq 0$.

Aufgabe A5

Es soll der unbekannte Parameter $\vartheta \in \Theta := (0, 1)$ für die Verteilung mit der Zähldichte

$$f_\vartheta(k) := k\ \vartheta^2\ (1 - \vartheta)^{k-1}, \qquad k \in \mathbb{N},$$

bestimmt werden.

a) Berechnen Sie zur Stichprobe $x = (x_1, x_2, \ldots, x_n)$ mit $x_i \in \mathbb{N}_0$, $1 \leq i \leq n$, die Loglikelihood-Funktion $M_x(\vartheta) := \ln L_x(\vartheta)$.

b) Berechnen Sie den Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ.
Hinweis: $\bar{x} > 1$ darf vorausgesetzt werden.

Lösung:

a) Wegen $\ln f_\vartheta(k) = \ln(k) + 2 \cdot \ln(\vartheta) + (k - 1) \cdot \ln(1 - \vartheta)$ gilt

$$M_x(\vartheta) = \sum_{i=1}^{n} \ln f_\vartheta(x_i) = 2n\ \ln(\vartheta) + \sum_{i=1}^{n} (x_i - 1) \cdot \ln(1 - \vartheta) + \sum_{i=1}^{n} \ln(x_i)$$

b) Man berechnet

$$\begin{aligned}
M_x'(\vartheta) &= \frac{2n}{\vartheta} - \frac{1}{1 - \vartheta} \sum_{i=1}^{n} (x_i - 1) \\
&= \frac{1}{\vartheta(1 - \vartheta)} \left[2n(1 - \vartheta) - \vartheta \underbrace{\sum_{i=1}^{n} (x_i - 1)}_{n\bar{x} - n} \right] \\
&= \frac{n}{\vartheta(1 - \vartheta)} \left[2 - (1 + \bar{x})\vartheta \right].
\end{aligned}$$

Damit gilt $M_x'(\vartheta) \gtreqless 0$ für $\vartheta \lesseqgtr \hat{\vartheta}(x) := \dfrac{2}{1 + \bar{x}}$. Also steigt $M_x(\vartheta)$ zuerst in ϑ, hat dann eine Maximumstelle in $\hat{\vartheta}(x) = \dfrac{2}{1 + \bar{x}}$, und fällt dann. Somit ist $\hat{\vartheta}(x)$ ein ML-Schätzer für ϑ.

Aufgabe A6

Eine Firma liefert 20 Waagen an einen Supermarkt. Die Waagen werden unabhängig voneinander und unter gleichen Bedingungen eingesetzt. Der TÜV kontrolliert die 20 Waagen nach einem Monat. Danach müssen 12 der Waagen neu kalibriert werden, da sie zu große Abweichungen von den tatsächlichen Gewichten anzeigen.

a) Es sei ϑ die unbekannte Wahrscheinlichkeit, dass eine Waage nach einem Monat neu kalibriert werden muss. Geben Sie ein Konfidenzintervall für ϑ zum Konfidenzniveau 0.95 an.

Lösung: Nach Voraussetzung kann wie in Beispiel 18.5 ein ideales Zufallsexperiment mit den zwei möglichen Ergebnissen „Kalibrierung erforderlich" (Treffer) und „Kalibrierung nicht erforderlich" (Niete) und der Trefferwahrscheinlichkeit ϑ angesehen werden. Nach diesem Beispiel ist mit $n = 20$ und $x = 12$ das gesuchte Konfidenzintervall $[l(x), L(x)]$, wobei die Konfidenzgrenzen $l(x)$ und $L(x)$ für $x = 12$ und $n - x = 8$ und $1 - \alpha = 0.95$ aus Tabelle A.4 entnommen werden. Das gesuchte Konfidenzintervall zum Konfidenzniveau 0.95 für ϑ ist daher

$$\mathcal{C}_1(x) = [0.361\ ,\ 0.809]$$

mit der Länge $0.809 - 0.361 = 0.448$.

b) Es sei jetzt p die unbekannte Wahrscheinlichkeit, dass eine Waage nach einem Monat nicht neu kalibriert werden muss. Geben Sie ein Konfidenzintervall für die Wahrscheinlichkeit p zum Konfidenzniveau 0.90 an. Ist dieses Konfidenzintervall kürzer oder länger als das zum Konfidenzniveau 0.95?

Lösung: Es ist $p = 1 - \vartheta$. Genauso wie oben ergibt sich als Konfidenzintervall zum Konfidenzniveau 0.9 für ϑ diesmal

$$\mathcal{C}_2(x) = [0.394\ ,\ 0.783]$$

und damit das gesuchte Konfidenzintervall

$$[1 - 0.783\ ,\ 1 - 0.394] = [0.217\ ,\ 0.606]$$

der Länge $0.606 - 0.217 = 0.389$. Dieses Intervall ist also kürzer als das Intervall aus a). (Dies gilt allgemein: Das gesuchte Konfidenzintervall und $\mathcal{C}_2(x)$ haben die gleiche Länge. Da wir gegenüber a) die „Sicherheitswahrscheinlichkeit" erniedrigt haben, sollte auch das zugehörige Konfidenzintervall kürzer werden.)