INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT KARLSRUHE
Priv.-Doz. Dr. D. Kadelka

WS 2005/06

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 9.3.2006

Musterlösungen

Aufgabe B1:

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	-3	-2.1	-0.9	0	1.1	1.9	2.8	3.8	4.6	5.7
y_j	7	9.6	10.9	1.8	5.8	8.5	3	2.7	5.8	-0.3

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

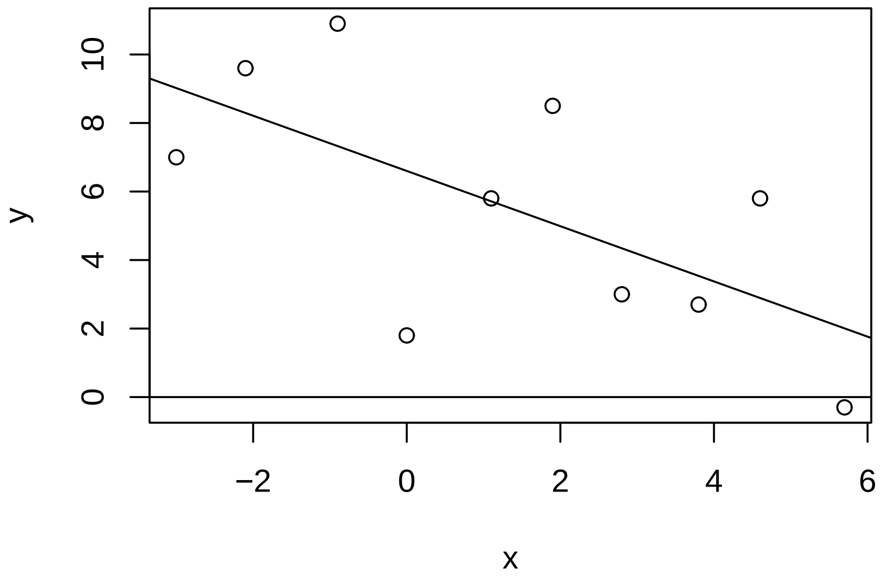
$$\sum_{j=1}^{n} (x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 1.39 \qquad s_x = 2.899$$
$$\bar{y} = 5.48 \qquad s_y = 3.633$$
$$r_{xy} = -0.6432$$

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

$$\begin{aligned} b^* &= -0.806 \\ a^* &= 6.6 \end{aligned}$$

und die Regressionsgerade $y = 6.6 - 0.806 \cdot x$.



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

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

$$y_{()} = (-0.3,\ 1.8,\ 2.7,\ 3,\ 5.8,\ 5.8,\ 7,\ 8.5,\ 9.6,\ 10.9)$$

c) Berechnen Sie das 0.15-getrimmte Stichprobenmittel $\bar{y}_{0.15}$ von $(y_1, \ldots, y_{10})$.
Lösung: Mit $k = [10 \cdot 0.15] = 1$ ergibt sich

$$\bar{y}_{0.15} = \frac{1}{10 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(9)}) = 5.525$$

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

$$\tilde{y}_{0.2} = \frac{y_{(k)} + y_{(k+1)}}{2} = \frac{y_{(2)} + y_{(3)}}{2} = 2.25$$

Aufgabe B2

Bei der Produktionskontrolle von Eisenstäben ergibt sich, dass deren Länge eine zufällige Größe L darstellt, wobei die Differenz $L - L_0$ von L zum Sollwert $L_0 = 1200\ mm$ eine $\mathcal{N}(0, 100)$-verteilte Zufallsvariable ist.

a) Welche Verteilung besitzt L?

b) Bestimmen Sie die Wahrscheinlichkeit p_1, dass die tatsächliche Länge um mehr als $10\ mm$ vom Sollwert abweicht.

c) Ein Eisenstab ist unbrauchbar, wenn seine Länge kleiner als $1182\ mm$ ist. Wie groß ist die Wahrscheinlichkeit p_2, dass ein Eisenstab unbrauchbar ist?

d) Der Hersteller der Eisenstäbe will garantieren, dass die tatsächliche Länge L zwischen $1200-c\ mm$ und $1200+c\ mm$ liegt, wobei c eine der Zahlen $1, 2, 3, \ldots$ ist. Wie groß muss c mindestens sein, damit weniger als 5% der Produktion diese Grenze überschreiten?

Lösung:

a) Wegen $L = L_0 + (L - L_0)$ und $L - L_0 \sim \mathcal{N}(0, 100)$ gilt nach Satz 9.7 $L \sim \mathcal{N}(L_0, 100) = \mathcal{N}(1200, 100)$.

b) Gesucht ist $\mathbb{P}(|L - L_0| > 10)$. Wegen der $k \cdot \sigma$-Regel (Folgerung 9.9) gilt mit $\sigma = 10$

$$\mathbb{P}(|L - L_0| \leq 10) = \mathbb{P}(|L - L_0| \leq 1 \cdot \sigma) = 2 \cdot \Phi(1) - 1 = 0.6826$$

und damit $\mathbb{P}(|L - L_0| > 10) = 1 - \mathbb{P}(|L - L_0| \leq 10) = 1 - 0.6826 = 0.3174$.

c) Der Eisenstab ist unbrauchbar, wenn $L < 1182$. Damit gilt wegen Satz 9.6

$$\begin{aligned}\mathbb{P}(\text{„Eisenstab unbrauchbar“}) = \mathbb{P}(L < 1182) &= \Phi\left(\frac{1182 - 2000}{10}\right) \\ &= \Phi(-1.80) = 1 - \Phi(1.80) = 1 - 0.9641 = 0.0359.\end{aligned}$$

d) Es ist $L_0 = 1200$ (in mm). Gesucht ist ein $c = 1, 2, 3, \ldots$ mit

$$\mathbb{P}(L < L_0 - c \text{ oder } L > L_0 + c) \leq 0.05.$$

Dies ist gleichbedeutend mit

$$\mathbb{P}(|L - L_0| > c) \leq 0.05 \quad \text{bzw.} \quad \mathbb{P}(|L - L_0| \leq c) \geq 0.95.$$

Anwendung der $k \cdot \sigma$-Regel ergibt

$$\mathbb{P}(|L - L_0| \leq c) = \mathbb{P}\left(|L - L_0| \leq \left(\frac{c}{\sigma}\right) \cdot \sigma\right) = 2 \cdot \Phi\left(\frac{c}{\sigma}\right) - 1 = 2 \cdot \Phi\left(\frac{c}{10}\right) - 1$$

und $2 \cdot \Phi\left(\frac{c}{10}\right) - 1 \geq 0.95$ genau dann wenn $\Phi\left(\frac{c}{10}\right) \geq \frac{1+0.95}{2} = 0.975$. Nach Tabelle A.1 ist $\Phi(1.96) = 0.9750$. Wir erhalten die Bedingung $\frac{c}{10} > 1.96$ bzw. $c > 10 \cdot 1.96 = 19.6$. Damit ergibt sich

$$c = 20.$$

Aufgabe B3

Für einen Katalysator werden Kugeln mit dem zufälligen Radius R produzierten. Die Zufallsvariable R besitze (annähernd) die Gleichverteilung $U(0, 20)$ (in Mikrometer gemessen) und damit die Dichte

$$f_R(x) = \frac{1}{20}, \quad 0 < x < 20.$$

a) Bestimmen Sie den Erwartungswert $\mathbb{E}R$ und die Varianz $V(R)$ von R.

b) Bestimmen Sie das 0.9-Quantil q von R.

c) Berechnen Sie $\mathbb{E}R^2$ und damit den Erwartungswert $\mathbb{E}F$ für die zufällige Oberfläche $F := 4\pi\ R^2$ der Kugeln.

d) Berechnen Sie $\mathbb{E}R^3$ und damit den Erwartungswert $\mathbb{E}V$ des zufälligen Volumens $V = \frac{4\pi}{3}R^3$ einer Kugel.

Lösung:

a) Da eine $U(a, b)$-verteilte Zufallsvariable X den Erwartungswert $\mathbb{E}X = \frac{a+b}{2}$ und die Varianz $V(X) = \frac{(b-a)^2}{12}$ besitzt, gilt hier mit $a = 0$ und $b = 20$

$$\mathbb{E}R = 10 \quad \text{und} \quad V(R) = \frac{400}{12} = \frac{100}{3}.$$

b) Da $U(0, 20)$ die Verteilungsfunktion $F(t) = \frac{t}{20}$, $0 \leq t \leq 20$ besitzt (B.8.9), ist nach Definition 12.19 die Gleichung

$$F(t_{0.9}) = \frac{t_{0.9}}{20} = 0.9$$

zu lösen. Wir erhalten, dass $q = t_{0.9} = 18.0$ das 0.9-Quantil von R ist.

c) Wegen Satz 12.6 gilt

$$\mathbb{E}R^2 = V(R) + (\mathbb{E}R)^2 \overset{a)}{=} \frac{100}{3} + 10^2 = \frac{400}{3} = 133.33$$

und damit

$$\mathbb{E}F = \mathbb{E}(4\pi \cdot R^2) = 4\pi \cdot \mathbb{E}R^2 = \frac{1600\pi}{3} = 1675.5$$

d) Wegen Satz 12.8 a) mit $g(x) = x^3$ gilt

$$\mathbb{E}R^3 = \int_{-\infty}^{\infty} g(x)\ f_R(x)\ dx = \int_0^{20} x^3 \cdot \frac{1}{20}\ dx = \frac{x^4}{4} \cdot \frac{1}{20}\bigg|_0^{20} = \frac{20^4}{80} = 2000$$

und damit

$$\mathbb{E}V = \mathbb{E}\left(\frac{4\pi}{3}R^3\right) = \frac{4\pi}{3}\mathbb{E}R^3 = \frac{8000\pi}{3} = 8377.6.$$

Aufgabe B4

In einer Druckerei befinden sich 4 unabhängig voneinander arbeitende, gleichartige Maschinen, von denen jede mit der Wahrscheinlichkeit 0.7 in einer bestimmten Zeitspanne nicht ausfällt.

a) Gesucht sind die Wahrscheinlichkeiten

- p_1, dass innerhalb dieser Zeitspanne wenigstens eine Maschine nicht ausfällt,
- p_2, dass innerhalb dieser Zeitspanne genau eine Maschine nicht ausfällt,
- p_3, dass innerhalb dieser Zeitspanne genau zwei Maschinen nicht ausfallen,
- p_4, dass innerhalb dieser Zeitspanne keine der Maschinen ausfällt.

b) Sei N die zufällige Anzahl von Maschinen, die innerhalb der Zeitspanne nicht ausfallen. Bestimmen Sie den Erwartungswert $\mathbb{E}N$ und die Varianz $V(N)$ von N.

c) Berechnen Sie die bedingte Wahrscheinlichkeit p, dass mindestens zwei Maschinen nicht ausfallen unter der Bedingung, dass mindestens eine Maschine nicht ausfällt.

Lösung: Da jede der vier Maschinen unabhängig von den anderen mit Wahrscheinlichkeit 0.7 nicht ausfällt, hat die zufällige Anzahl N von Maschinen, die innerhalb der Zeitspanne nicht ausfallen, die Binomialverteilung $Bin(4, 0.7)$.

a) Wir können die gesuchten Wahrscheinlichkeiten berechnen, indem wir die angegebenen Ereignisse mit Hilfe von N bestimmen. Da N die Binomialverteilung $Bin(4, 0.7)$ mit der Zähldichte

$$f_N(k) = \mathbb{P}(N = k) = \binom{4}{k} \, 0.7^k \, (1 - 0.7)^{4-k} \text{ für } k = 0, 1, 2, 3, 4$$

besitzt, gilt

- $p_1 = \mathbb{P}(\text{„Wenigstens eine Maschine fällt nicht aus“}) = \mathbb{P}(N \geq 1) = 1 - \mathbb{P}(N = 0) = 1 - \binom{4}{0} \, 0.7^0 \, (1 - 0.7)^4 = 1 - 0.3^4 = 0.9919$
- $p_2 = \mathbb{P}(\text{„genau eine Maschine fällt nicht aus“}) = \mathbb{P}(N = 1) = \binom{4}{1} \, 0.7^1 \, (1 - 0.7)^3 = 0.0756$
- $p_3 = \mathbb{P}(\text{„genau zwei Maschinen fallen nicht aus“}) = \mathbb{P}(N = 2) = \binom{4}{2} \, 0.7^2 \, 0.3^2 = 0.2646$
- $p_4 = \mathbb{P}(\text{„keine der Maschinen fällt aus“}) = \mathbb{P}(N = 4) = \binom{4}{4} \, 0.7^4 \, 0.3^0 = 0.2401.$

b) Da N die Binomialverteilung $Bin(4, 0.7)$ besitzt, gilt nach den Tabellen auf S. 124 bzw. 128 im Skriptum

- $\mathbb{E}N = 4 \cdot 0.7 = 2.8$ und
- $V(N) = 4 \cdot 0.7 \cdot (1 - 0.7) = 0.84.$

c) Gesucht ist die bedingte Wahrscheinlichkeit $p = \mathbb{P}(N \geq 2 \mid N \geq 1)$. Nach Definition der bedingten Wahrscheinlichkeit gilt

$$\begin{aligned}
p = \mathbb{P}(N \geq 2 \mid N \geq 1) &= \frac{\mathbb{P}(\{N \geq 2\} \cap \{N \geq 1\})}{\mathbb{P}(N \geq 1)} = \frac{\mathbb{P}(N \geq 2)}{\mathbb{P}(N \geq 1)} = \frac{1 - \mathbb{P}(N \leq 1)}{1 - \mathbb{P}(N = 0)} \\
&= \frac{1 - (\mathbb{P}(N = 0) + \mathbb{P}(N = 1))}{1 - \mathbb{P}(N = 0)} = \frac{1 - \left(\binom{4}{0} \cdot 0.3^4 + \binom{4}{1} \cdot 0.7 \cdot 0.3^3\right)}{1 - \binom{4}{0} \cdot 0.3^4} \\
&= \frac{1 - (0.0081 + 0.0756)}{1 - 0.0081} = 0.9238
\end{aligned}$$

Aufgabe B5

Ein Merkmal besitze die $Po(\vartheta^{1/2})$-Verteilung mit unbekanntem Parameter $\vartheta > 0$. Diese Verteilung besitzt die Zähldichte

$$f_\vartheta(k) = e^{-\vartheta^{1/2}} \cdot \frac{\vartheta^{k/2}}{k!}, \quad k = 0, 1, 2, \ldots$$

$X_1, \ldots, X_n$ seien die stochastisch unabhängigen, $Po(\vartheta^{1/2})$-verteilten Stichprobenvariablen.

a) Bestimmen Sie die Verteilung von $\sum_{i=1}^n X_i$.

b) Beobachtet sei die Stichprobe $x = (x_1, \ldots, x_n)$, wobei $x_1, \ldots, x_n \in \mathbb{N}_0 = \{0, 1, \ldots\}$. Bestimmen Sie die Loglikelihood-Funktion $M_x(\vartheta)$ zur Stichprobe x.

c) Berechnen Sie die Ableitung $M_x'(\vartheta)$.

d) Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ zur Stichprobe x.
Hinweis: Sie dürfen $x_1 + \ldots + x_n > 0$ voraussetzen.

e) Bestimmen Sie den Momentenschätzer $\hat{\vartheta}_m(x)$ für ϑ zur Stichprobe x.

Lösung:

a) Da jede der unabhängigen Zufallsvariablen X_i, $i = 1, \ldots, n$ $Po(\vartheta^{1/2})$-verteilt ist, gilt nach der Faltungsformel $\displaystyle\sum_{i=1}^n X_i \sim Po(n\vartheta^{1/2})$.

b) Zur Berechnung der Loglikelihood-Funktion benötigen wir $\ln(f_\vartheta(x))$. Mit

$$\ln(f_\vartheta(x)) = -\vartheta^{1/2} + \frac{k}{2} \cdot \ln(\vartheta) - \ln(k!), k = 0, 1, 2, \ldots$$

ergibt sich die Loglikelihood-Funktion zu

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

c) Die Ableitung ist

$$M_x'(\vartheta) = -\frac{n}{2} \cdot \vartheta^{-1/2} + \frac{1}{2\vartheta} \cdot \sum_{i=1}^n x_i = \frac{n}{2\vartheta}\left(-\vartheta^{1/2} + \frac{1}{n}\sum_{i=1}^n x_i\right)$$

d) Wegen $\frac{n}{2\vartheta} > 0$ gilt hier $M_x'(\vartheta) = 0$ genau dann, wenn $-\vartheta^{1/2} + \frac{1}{n}\sum_{i=1}^n x_i = 0$ bzw. aufgelöst nach ϑ

$$\vartheta = \hat{\vartheta}(x) := \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2$$

Zu zeigen bleibt, dass $\hat{\vartheta}(x)$ tatsächlich eine Maximumstelle der Loglikelihood-Funktion ist. Wegen $\dfrac{n}{2\vartheta} > 0$ gilt $M_x'(\vartheta) \left\{\begin{matrix} > \\ = \\ < \end{matrix}\right\} 0$ genau dann wenn $\vartheta^{1/2} \left\{\begin{matrix} < \\ = \\ > \end{matrix}\right\} \dfrac{1}{n}\displaystyle\sum_{i=1}^n x_i$, bzw.

wenn $\vartheta \left\{ \begin{array}{c} < \\ = \\ > \end{array} \right\} \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2$. Dies bedeutet, dass die Loglikelihood-Funktion bis zur Stelle $\hat{\vartheta}(x) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2$ steigt und danach fällt. Damit ist $\hat{\vartheta}(x)$ die gesuchte Maximumstelle.

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 $Po(\vartheta^{1/2})$ besitzt, gilt nach der Tabelle auf S. 124 im Skriptum $\mathbb{E}_\vartheta(X_1) = \vartheta^{1/2}$ und damit

$$\begin{aligned} \vartheta^{1/2} &= \bar{x} \\ \vartheta &= (\bar{x})^2 = \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2 = \hat{\vartheta}(x) \end{aligned}$$

Der gesuchte Momentenschätzer ist also $\hat{\vartheta}_m(x) := \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2$ und stimmt für diesen Fall mit dem Maximum-Likelihood-Schätzer überein.