INSTITUT FÜR MATHEMATISCHE STOCHASTIK WS 2005/06
UNIVERSITÄT KARLSRUHE
Priv.-Doz. Dr. D. Kadelka

Klausur (Maschineningenieure)

Wahrscheinlichkeitstheorie und Statistik
vom 9.3.2006

Musterlösungen

Aufgabe 1:

Gegeben sei eine Urliste mit den Paaren $(x_1, y_1), \ldots, (x_{12}, y_{12})$

j	1	2	3	4	5	6	7	8	9	10	11	12
x_j	-2	-1.1	-0.1	1.2	2	2.8	3.9	5.1	6.3	7	8.2	8.9
y_j	9.9	6.3	10.3	5.6	3.7	4.2	-0.2	0.1	-1	-5	-1.5	-7.6

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

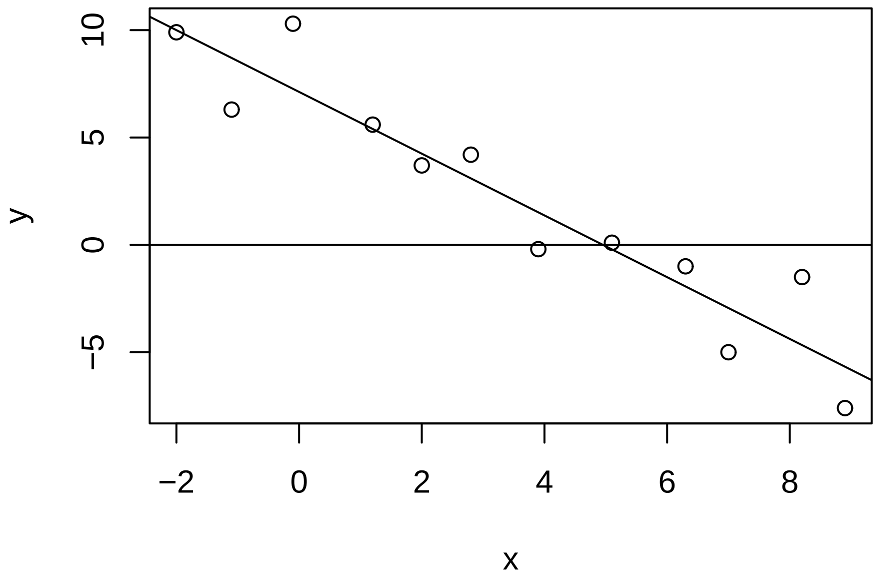
$$\sum_{j=1}^{n} (x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 3.52 \qquad s_x = 3.649$$
$$\bar{y} = 2.07 \qquad s_y = 5.57$$
$$r_{xy} = -0.942$$

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

$$\begin{aligned} b^* &= -1.438 \\ a^* &= 7.12 \end{aligned}$$

und die Regressionsgerade $y = 7.12 - 1.438 \cdot x$.



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

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

$$y_{()} = (-7.6,\ -5,\ -1.5,\ -1,\ -0.2,\ 0.1,\ 3.7,\ 4.2,\ 5.6,\ 6.3,\ 9.9,\ 10.3)$$

c) Berechnen Sie das 0.15-getrimmte Stichprobenmittel $\bar{y}_{0.15}$ von $(y_1, \ldots, y_{12})$.
Lösung: Mit $k = [12 \cdot 0.15] = 1$ ergibt sich

$$\bar{y}_{0.15} = \frac{1}{12 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(11)}) = 2.21$$

d) Bestimmen Sie das Stichproben-0.2-Quantil $\tilde{y}_{0.2}$ von $(y_1, \ldots, y_{12})$.
Lösung: Da $12 \cdot 0.2 = 2.4$ nicht ganzzahlig ist, ist mit $k = [2.4] = 2$

$$\tilde{y}_{0.2} = y_{(k+1)} = y_{(3)} = -1.5$$

e) Berechnen Sie den Quartilsabstand von $(y_1, \ldots, y_{12})$.
Lösung: Da $0.25 \cdot 12 = 3$ und $0.75 \cdot 12 = 9$ beide ganzzahlig sind, ergibt sich mit $k_1 = 3$ und $k_2 = 9$

$$\begin{aligned} \tilde{y}_{0.25} &= \frac{y_{(k_1)} + y_{(k_1+1)}}{2} = \frac{y_{(3)} + y_{(4)}}{2} = -1.25 \\ \tilde{y}_{0.75} &= \frac{y_{(k_2)} + y_{(k_2+1)}}{2} = \frac{y_{(9)} + y_{(10)}}{2} = 5.95 \end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 7.2$.

Aufgabe 2

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 $L_0 = 700\ mm$ eine $\mathcal{N}(0, 25)$-verteilte Zufallsvariable ist.

a) Welche Verteilung besitzt L?

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

c) Ein Eisenstab ist unbrauchbar, wenn seine Länge kleiner als 686 mm ist. Wie groß ist die Wahrscheinlichkeit, dass ein Eisenstab unbrauchbar ist?

d) Der Hersteller der Eisenstäbe will garantieren, dass die tatsächliche Länge L zwischen $700 - c\ mm$ und $700 + c\ mm$ liegt, wobei c eine der Zahlen $1, 2, 3, \ldots$ ist. Wie groß muss c mindestens sein, damit weniger als 3% der Produktion diese Grenze überschreiten?

Lösung:

a) Wegen $L = L_0 + (L - L_0)$ und $L - L_0 \sim \mathcal{N}(0, 25)$ gilt nach Satz 9.7 $L \sim \mathcal{N}(L_0, 25) = \mathcal{N}(700, 25)$.

b) Gesucht ist $\mathbb{P}(|L - L_0| > 10)$. Wegen der $k \cdot \sigma$-Regel (Folgerung 9.9) gilt mit $\sigma = 5$

$$\mathbb{P}(|L - L_0| \leq 10) = \mathbb{P}(|L - L_0| \leq 2 \cdot \sigma) = 2 \cdot \Phi(2) - 1 = 0.9545$$

und damit $\mathbb{P}(|L - L_0| > 10) = 1 - \mathbb{P}(|L - L_0| \leq 10) = 1 - 0.9545 = 0.0455$.

c) Der Eisenstab ist unbrauchbar, wenn $L < 686$. Damit gilt wegen Satz 9.6

$$\begin{aligned}\mathbb{P}(\text{„Eisenstab unbrauchbar“}) = \mathbb{P}(L < 686) &= \Phi\left(\frac{686 - 700}{5}\right) \\ &= \Phi(-2.80) = 1 - \Phi(2.80) = 1 - 0.9974 = 0.0026.\end{aligned}$$

d) Es ist $L_0 = 700$ (in mm). Gesucht ist ein $c = 1, 2, 3, \ldots$ mit

$$\mathbb{P}(L < L_0 - c \text{ oder } L > L_0 + c) \leq 0.03.$$

Dies ist gleichbedeutend mit

$$\mathbb{P}(|L - L_0| > c) \leq 0.03 \quad \text{bzw.} \quad \mathbb{P}(|L - L_0| \leq c) \geq 0.97.$$

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

$$\mathbb{P}(|L - L_0| \leq c) = \mathbb{P}(|L - L_0| \leq \left(\frac{c}{\sigma}\right) \cdot \sigma) = 2 \cdot \Phi\left(\frac{c}{\sigma}\right) - 1 = 2 \cdot \Phi\left(\frac{c}{5}\right) - 1$$

und $2 \cdot \Phi\left(\frac{c}{5}\right) - 1 \geq 0.97$ genau dann wenn $\Phi\left(\frac{c}{5}\right) \geq \frac{1+0.97}{2} = 0.985$. Nach Tabelle A.1 ist $\Phi(2.17) = 0.9850$. Wir erhalten die Bedingung $\frac{c}{5} > 2.17$ bzw. $c > 5 \cdot 2.17 = 10.85$. Damit ergibt sich

$$c = 11.$$

Aufgabe 3

In einer Druckerei befinden sich 4 unabhängig voneinander arbeitende, gleichartige Maschinen, von denen jede mit der Wahrscheinlichkeit 0.8 in einer bestimmten Zeitspanne nicht ausfällt.

a) Gesucht sind die Wahrscheinlichkeiten, dass innerhalb dieser Zeitspanne

- a_1) wenigstens eine Maschine nicht ausfällt,
- a_2) genau eine Maschine nicht ausfällt,
- a_3) genau zwei Maschinen nicht ausfallen,
- a_4) 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, 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 0.8 nicht ausfällt, hat die zufällige Anzahl N von Maschinen, die innerhalb der Zeitspanne nicht ausfallen, die Binomialverteilung $Bin(4, 0.8)$.

a) Wir können die gesuchten Wahrscheinlichkeiten berechnen, indem wir die angegebenen Ereignisse mit Hilfe von N bestimmen. Da N die Binomialverteilung $Bin(4, 0.8)$ mit der Zähldichte

$$f_N(k) = \mathbb{P}(N = k) = \binom{4}{k}\ 0.8^k\ (1 - 0.8)^{4-k} \text{ für } k = 0, 1, 2, 3, 4$$

besitzt, gilt

- a_1) $\mathbb{P}(\text{„Wenigstens eine Maschine fällt nicht aus“}) = \mathbb{P}(N \geq 1) = 1 - \mathbb{P}(N = 0) = 1 - \binom{4}{0}\ 0.8^0\ (1 - 0.8)^4 = 1 - 0.2^4 = 0.9984$
- a_2) $\mathbb{P}(\text{„genau eine Maschine fällt nicht aus“}) = \mathbb{P}(N = 1) = \binom{4}{1}\ 0.8^1\ (1 - 0.8)^3 = 0.0256$
- a_3) $\mathbb{P}(\text{„genau zwei Maschinen fallen nicht aus“}) = \mathbb{P}(N = 2) = \binom{4}{2}\ 0.8^2\ 0.2^2 = 0.1536$
- a_4) $\mathbb{P}(\text{„keine der Maschinen fällt aus“}) = \mathbb{P}(N = 4) = \binom{4}{4}\ 0.8^4\ 0.2^0 = 0.4096.$

b) Da N die Binomialverteilung $Bin(4, 0.8)$ besitzt, gilt nach den Tabellen auf S. 124 bzw. 128 im Skriptum

- $\mathbb{E}N = 4 \cdot 0.8 = 3.2$ und
- $V(N) = 4 \cdot 0.8 \cdot (1 - 0.8) = 0.64.$

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

$$\begin{aligned}
\mathbb{P}(N \geq 2 \mid N \geq 1) &= \frac{\mathbb{P}(\{N \geq 2\} \cap \{N \geq 1\})}{\mathbb{P}(N \geq 1)} = \frac{\mathbb{P}(N \geq 2)}{\mathbb{P}(N \geq 1)} = \frac{1 - \mathbb{P}(N \leq 1)}{1 - \mathbb{P}(N = 0)} \\
&= \frac{1 - (\mathbb{P}(N = 0) + \mathbb{P}(N = 1))}{1 - \mathbb{P}(N = 0)} = \frac{1 - \left(\binom{4}{0} \cdot 0.2^4 + \binom{4}{1} \cdot 0.8 \cdot 0.2^3\right)}{1 - \binom{4}{0} \cdot 0.2^4} \\
&= \frac{1 - (0.0016 + 0.0256)}{1 - 0.0016} = 0.9744
\end{aligned}$$

Aufgabe 4

Seien X, Y und Z drei unabhängige Zufallsvariablen mit den Dichten

$$f(t) := f_X(t) = f_Y(t) = f_Z(t) = 2e^{-2t}, \qquad t \geq 0.$$

a) X, Y und Z besitzen die Verteilung $Exp(\lambda)$. Bestimmen Sie den Parameter λ.

b) Welche Verteilung besitzt $X+Y$ und welche Verteilung besitzt $X+Y+Z$?

c) Berechnen Sie $\mathbb{P}(X+Y>1)$.

d) Berechnen Sie $\mathbb{E}[2 \cdot X \cdot (X+Y+Z)]$.

e) Bestimmen Sie die gemeinsame Dichte $f_{X,Y}(s,t)$ von X und Y für $s>0$ und $t>0$.

Lösung:

a) Allgemein besitzt $Exp(\lambda)$ die Dichte $f_1(x) = \lambda \cdot e^{-\lambda \cdot x}$ für $x>0$ (S.88). Ein direkter Vergleich mit f ergibt, dass

$$X \sim Y \sim Z \sim Exp(2).$$

Es ist also $\lambda = 2$.

b) Da die Zufallsvariablen X, Y und Z unabhängig sind, lassen sich die Faltungsformeln in der Tabelle auf S. 114 anwenden. Es ergibt sich

$$X+Y \sim \Gamma(2,2)$$

und damit

$$X+Y+Z = (X+Y)+Z \sim \Gamma(2+1,2) = \Gamma(3,2).$$

c) Da $\Gamma(2,2)$ die Verteilungsfunktion (vergl. (9.1))

$$F_2(t) = 1 - e^{-2t} - e^{-2t} \cdot \frac{2t}{1!} = 1 - e^{-2t} \cdot (1+2t), \quad t \geq 0$$

besitzt, gilt

$$\mathbb{P}(X+Y>1) = 1 - \mathbb{P}(X+Y \leq 1) = 1 - F_2(1) = e^{-2} \cdot (1+2) = 3e^{-2} = 0.4060.$$

d) Es gilt nach den Eigenschaften des Erwartungswertes (Satz 12.6)

$$\begin{aligned}\mathbb{E}[2 \cdot X \cdot (X+Y+Z)] = 2 \cdot \mathbb{E}[X \cdot (X+Y+Z)] &= 2 \cdot \mathbb{E}[X^2 + X \cdot Y + X \cdot Z] \\ &= 2 \cdot (\mathbb{E}X^2 + \mathbb{E}(X \cdot Y) + \mathbb{E}(X \cdot Z))\end{aligned}$$

Da X und Y bzw. X und Z unabhängig sind, gilt $\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ und genauso $\mathbb{E}(X \cdot Z) = \frac{1}{4}$. (Diese Regeln darf man natürlich nicht auf die Berechnung von $\mathbb{E}(X^2) = \mathbb{E}(X \cdot X)$ anwenden!) Hier verwenden wir Satz 12.11 und erhalten $\mathbb{E}X^2 = V(X) + (\mathbb{E}X)^2 = \frac{1}{4} + \left(\frac{1}{2}\right)^2 = \frac{1}{2}$ (B.12.16 b) liefert das gleiche Ergebnis). Insgesamt ergibt sich

$$\mathbb{E}[2 \cdot X \cdot (X+Y+Z)] = 2.$$

e) Da X und Y stochastisch unabhängig sind, ist nach Satz 11.8

$$f_{X,Y}(s,t) = f_X(s) \cdot f_Y(t), \quad s,t \in \mathbb{R}$$

eine gemeinsame Dichte von X und Y. Einsetzen der Dichten von X und Y ergibt für $s, t > 0$

$$f_{X,Y}(s,t) = 2e^{-2s} \cdot 2e^{-2t} = 4e^{-2(s+t)}.$$

Aufgabe 5

Für einen Katalysator werden Kugeln mit dem zufälligen Radius R produzierten. Die Zufallsvariable R besitze (annähernd) die Gleichverteilung $U(0,10)$ (in Mikrometer gemessen) und damit die Dichte

$$f_R(x) = \frac{1}{10}, \quad 0 < x < 10.$$

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

b) Bestimmen Sie das 0.8-Quantil 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.
Hinweis: $\mathbb{E}R^2$ kann ohne Auswertung von Integralen ermittelt werden.

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

e) Berechnen Sie die Kovarianz $C(R,F)$ und den Korrelationskoeffizienten $\rho(R,F)$ von R und F.
Hinweis: Sie dürfen ohne Nachweis benützen, dass $V(R^2) = \frac{8000}{9}$. Wie groß ist also $V(F)$?

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 $b = 10$

$$\mathbb{E}R = 5 \quad \text{und} \quad V(R) = \frac{100}{12} = \frac{25}{3}.$$

b) Da $U(0,10)$ die Verteilungsfunktion $F(t) = \frac{t}{10}$, $0 \le t \le 10$ besitzt (B.8.9), ist nach Definition 12.19 die Gleichung

$$F(t_{0.8}) = \frac{t_{0.8}}{10} = 0.8$$

zu lösen. Wir erhalten, dass $t_{0.8} = 8.0$ das 0.8-Quantil von R ist.

c) Wegen Satz 12.6 gilt

$$\mathbb{E}R^2 = V(R) + (\mathbb{E}R)^2 \stackrel{a)}{=} \frac{25}{3} + 5^2 = \frac{100}{3} = 33.33$$

und damit

$$\mathbb{E}F = \mathbb{E}(4\pi \cdot R^2) = 4\pi \cdot \mathbb{E}R^2 = \frac{400\pi}{3} = 418.9.$$

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

$$\mathbb{E}R^3 = \int_{-\infty}^{\infty} g(x)\ f_R(x)\ dx = \int_0^{10} x^3 \cdot \frac{1}{10}\ dx = \frac{x^4}{4} \cdot \frac{1}{10}\bigg|_0^{10} = \frac{10^4}{40} = 250$$

und damit

$$\mathbb{E}V = \mathbb{E}\left(\frac{4\pi}{3}R^3\right) = \frac{4\pi}{3}\mathbb{E}R^3 = \frac{1000\pi}{3} = 1047.2.$$

e) Wegen Satz 12.23 gilt

$$\begin{aligned} C(R,F) = C(F,R) = C(4\pi R^2, R) = 4\pi\ C(R^2,R) &= 4\pi \cdot (E(R^2 \cdot R) - \mathbb{E}R^2 \cdot \mathbb{E}R) \\ = 4\pi \cdot (ER^3 - \mathbb{E}R^2 \cdot \mathbb{E}R) \overset{b),c)}{=} 4\pi\left(250 - \frac{100}{3}\cdot 5\right) &= \frac{1000\pi}{3} = 1047.2. \end{aligned}$$

Wir weisen zuerst die Gültigkeit des Hinweises nach. Es ist

$$\begin{aligned} V(R^2) = \mathbb{E}[(R^2)^2] - (\mathbb{E}R^2)^2 = \mathbb{E}R^4 - (\mathbb{E}R^2)^2 = \int_0^{10} \frac{1}{10}x^4\ dx - \left(\frac{100}{3}\right)^2 \\ = 2000 - \frac{10000}{9} = \frac{8000}{9} \end{aligned}$$

und damit $V(F) = V(4\pi\ R^2) = (4\pi)^2 \cdot \frac{8000}{9}$. Wir erhalten

$$\rho(R,F) = \frac{C(R,F)}{\sqrt{V(R)\cdot V(F)}} = \frac{4\pi\frac{250}{3}}{\sqrt{\frac{25}{3}\cdot(4\pi)^2\cdot\frac{8000}{9}}} = \frac{\frac{250}{3}}{\sqrt{\frac{25}{3}\cdot\frac{8000}{9}}} = \frac{5\sqrt{3}}{\sqrt{80}} = 0.9682.$$

Ein wenig vereinfachen könnte man die Rechnung mit $\rho(R,F) = \rho(R, 4\pi \cdot R^2) = \rho(R,R^2)$, indem man Satz 12.23 c) ausnützt.

Aufgabe 6

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

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

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

a) Zeigen Sie, dass

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

ein Maximum-Likelihood-Schätzer für ϑ zur Stichprobe $x = (x_1,\dots,x_n)$ ist. Vorausgesetzt werden darf dabei, dass $\sum_{i=1}^n x_i > 0$.

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

c) Berechnen Sie $\mathbb{E}\hat\vartheta(X_1,\dots,X_n) = \mathbb{E}\left(\frac{1}{n}\cdot\sum_{i=1}^n X_i\right)^2$.
Hinweis: Allgemein gilt $\mathbb{E}Y^2 = V(Y) + (\mathbb{E}Y)^2$.

d) Ist $\hat{\vartheta}(x)$ ein erwartungstreuer Schätzer für $\gamma(\vartheta) = \vartheta$?

e) Ist $\hat{\vartheta}(x)$ ein asymptotisch erwartungstreuer Schätzer für $\gamma(\vartheta) = \vartheta$?
Hinweis: Begründen Sie in d) und e) Ihre Antwort!

Lösung:

a) Zur Verteilung $Po(\alpha)$ gehört die Zähldichte $f(x) = e^{-\alpha} \cdot \frac{\alpha^k}{k!}$, $k \in \mathbb{N}_0$, also zur $Po(\vartheta^{1/2})$-Verteilung die Zähldichte

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

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

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

mit der Ableitung

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

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 $M_x'(\vartheta) \left\{\begin{array}{c} > \\ = \\ < \end{array}\right\} 0$ genau dann wenn $\vartheta^{1/2} \left\{\begin{array}{c} < \\ = \\ > \end{array}\right\} \dfrac{1}{n}\displaystyle\sum_{i=1}^n x_i$, bzw. wenn $\vartheta \left\{\begin{array}{c} < \\ = \\ > \end{array}\right\} \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^n x_i\right)^2$. 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.

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

c) Wir schreiben zur Abkürzung $Y := \sum_{i=1}^{n} X_i$. Es gilt

$$\mathbb{E}_\vartheta(\frac{1}{n}Y)^2 = \mathbb{E}_\vartheta \frac{1}{n^2}Y^2 = \frac{1}{n^2}\mathbb{E}_\vartheta Y^2 = \frac{1}{n^2}(V_\vartheta(Y) + (\mathbb{E}_\vartheta Y)^2)$$
$$= \frac{1}{n^2}(n\vartheta^{1/2} + n^2\vartheta) = \frac{\vartheta^{1/2}}{n} + \vartheta$$

und damit

$$\mathbb{E}_\vartheta\ \hat{\vartheta}(X) = \frac{\vartheta^{1/2}}{n} + \vartheta.$$

d) Nach Definition der Erwartungstreue wäre $\hat{\vartheta}(x)$ genau dann ein erwartungstreuer Schätzer für ϑ, wenn $\mathbb{E}_\vartheta\ \hat{\vartheta}(X) = \vartheta$. Wegen c) ist dies nicht erfüllt und damit $\hat{\vartheta}(x)$ kein erwartungstreuer Schätzer für ϑ.

e) Es gilt $\lim_{n\to\infty} \mathbb{E}_\vartheta\ \hat{\vartheta}(X_1, \ldots, X_n) = \lim_{n\to\infty}\left(\frac{\vartheta^{1/2}}{n} + \vartheta\right) = \vartheta$ für alle $\vartheta > 0$. Damit ist $\hat{\vartheta}(x)$ ein asymptotisch erwartungstreuer Schätzer für ϑ.

Aufgabe 7

Eine Firma baut $n = 15$ gleichartige Maschinen, von denen $x = 7$ einwandfrei sind und ohne Nachbesserung vom Käufer übernommen werden, während die übrigen nicht einwandfrei sind.

a) Es sei p die unbekannte Wahrscheinlichkeit, dass eine Maschine bei Auslieferung einwandfrei ist. Geben Sie ein zweiseitiges Konfidenzintervall

$$\mathcal{C}_1(x) = [l_1(x)\ , L_1(x)]$$

zum Konfidenzniveau 0.90 für die Wahrscheinlichkeit p an, dass eine Maschine bei Auslieferung einwandfrei ist.

Lösung: Nach Voraussetzung kann wie in Beispiel 18.5 ein ideales Zufallsexperiment mit den zwei möglichen Ergebnissen „einwandfrei" (Treffer) und „nicht einwandfrei" (Niete) und der Trefferwahrscheinlichkeit $\vartheta := p$ angesehen werden. Nach diesem Beispiel und wegen 18.6 ist das gesuchte Konfidenzintervall $[l_1(x), L_1(x)]$, wobei die Konfidenzgrenzen $l_1(x)$ und $L_1(x)$ für $x = 7$ und $n - x = 8$ und $1 - \alpha = 0.90$ aus Tabelle A.4 entnommen werden. Dies ergibt das Konfidenzintervall

$$\mathcal{C}_1(x) = [0.244\ ,\ 0.700]\ .$$

b) Bestimmen Sie mit den gleichen Daten ein zweiseitiges Konfidenzintervall

$$\mathcal{C}_2(x) = [l_2(x)\ , L_2(x)]$$

zum Konfidenzniveau 0.95 diesmal für die Wahrscheinlichkeit $1-p$ an, dass eine Maschine bei der Auslieferung nicht einwandfrei ist.

Lösung: Gefragt ist jetzt nach der Wahrscheinlichkeit $\vartheta := 1 - p$. Ein Treffer liegt nun vor, wenn die Maschine „nicht einwandfrei" ist, andernfalls eine Niete. Die Anzahl dieser Treffer ist daher jetzt 8 und wir erhalten das gesuchte Konfidenzintervall $[l_2(x), L_2(x)]$, wenn wir die Konfidenzgrenzen $l_2(x)$ und $L_2(x)$ aus Tabelle A.4 entnehmen, allerdings mit den Werten $x = 8$, $n - x = 7$ und $1 - \alpha = 0.95$. Dies ergibt das Konfidenzintervall

$$\mathcal{C}_2(x) = [0.266\ ,\ 0.787]\ .$$