UNIVERSITÄT KARLSRUHE (TH)

Lehrstuhl für Statistik und Ökonometrie

S T A T I S T I K 2

Lösungsblätter zu den Übungen

Wintersemester 2006/2007

Prof. Dr. G. Bol, Dr. M. Höchstötter
Dipl.-Math. Sebastian Kring
Institut für Statistik
und Mathematische Wirtschaftstheorie
D-76128 Karlsruhe

Aufgabe 28

Die Lebensdauern T_1 und T_2 zweier elektrischer Bauteile B_1 und B_2 seien exponentialverteilt mit den Parametern $\lambda_1 = \frac{1}{500}$ und $\lambda_2 = \frac{1}{300}$ und unabhängig.

(a) Berechnen Sie die Wahrscheinlichkeit dafür, dass B_1 bzw. B_2 den Zeitpunkt $t_0 = 200$ überlebt, wenn das jeweilige Bauteil zur Zeit $t = 0$ eingesetzt wurde. Wie groß ist die Wahrscheinlichkeit dafür, dass das jeweilige Bauteil nach Erreichen von t_0 noch weitere 200 Stunden arbeitet?

(b) Bestimmen Sie die Lebensdauerverteilung eines aus B_1 und B_2 bestehenden Reihen- bzw. Parallelsystems.

(c) Bestimmen Sie die Wahrscheinlichkeiten dafür, dass das Reihen- bzw. das Parallelsystem den Zeitpunkt $t = 200$ überlebt.

Lösung: (Wahrscheinlichkeitstheorie, S.190f., S.189f.)

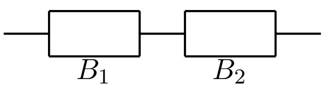
(a) Verteilungsfunktion der Exponentialverteilung ist:

$$F(t) = \begin{cases} 0 & \text{für } t < 0 \\ 1 - e^{-\lambda \cdot t} & \text{für } t \geq 0 \end{cases}$$

$$\begin{aligned} P(T_1 \geq 200) &= 1 - F(200) = e^{-\frac{200}{500}} \\ &= 0.6703 \quad (= P(T_1 \geq 400 \mid T_1 > 200)) \\ P(T_2 \geq 200) &= e^{-\frac{200}{300}} \\ &= 0.5134 \quad (= P(T_2 \geq 400 \mid T_2 > 200)) \end{aligned}$$

Die Exponentialverteilung ist eine Verteilung ohne Gedächtnis, vgl. Aufgabe 10.

(b) Das Reihensystem

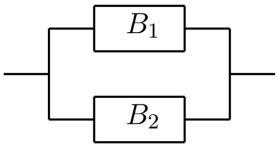


ist intakt, wenn beide Bauteile intakt sind, d.h. $T_s = \min\{T_1, T_2\}$ ist der Zeitpunkt des Systemausfalls.

$$\begin{aligned} P(T_s \leq t) &= P(\min\{T_1, T_2\} \leq t) = 1 - P(\min\{T_1, T_2\} > t) \\ &= 1 - P(T_1 > t \text{ und } T_2 > t) = 1 - P(T_1 > t) \cdot P(T_2 > t) \quad \text{(Unabhängigkeit!)} \\ &= 1 - (1 - F_{T_1}(t)) \cdot (1 - F_{T_2}(t)) = 1 - e^{-\lambda_1 \cdot t} e^{-\lambda_2 \cdot t} \\ &= 1 - e^{-(\lambda_1 + \lambda_2) \cdot t} = F_{T_s}(t) \end{aligned}$$

d.h. das Minimum und damit die Lebensdauer des Reihensystems ist exponentialverteilt mit dem Parameter $(\lambda_1 + \lambda_2)$.

Das Parallellsystem



ist intakt, solange mindestens eines der beiden Bauteile intakt ist, d.h. $T_s = \max\{T_1, T_2\}$ ist der Zeitpunkt des Systemausfalls.

$$
\begin{aligned}
P(T_s \leq t) &= P(\max\{T_1, T_2\} \leq t) = P(T_1 < t \text{ und } T_2 < t) \\
&= P(T_1 \leq t) \cdot P(T_2 \leq t) = F_{T_1}(t) \cdot F_{T_2}(t) \\
&= (1 - e^{-\lambda_1 \cdot t}) \cdot (1 - e^{-\lambda_2 \cdot t}) = 1 - e^{-\lambda_1 \cdot t} - e^{-\lambda_2 \cdot t} + e^{-(\lambda_1 + \lambda_2) \cdot t} \\
&= F_{T_s}(t)
\end{aligned}
$$

d.h. das Maximum und damit die Lebensdauer des Parallelsystems ist nicht exponentialverteilt.

(c) Die Überlebenswahrscheinlichkeit des Reihensystem ist:

$$
\begin{aligned}
P(T_s \geq 200) &= 1 - P(T_s < 200) = e^{-(\frac{1}{500} + \frac{1}{300}) \cdot 200} = e^{-\frac{800}{150000} \cdot 200} = e^{-\frac{16}{15}} \\
&= 0.3442.
\end{aligned}
$$

Die Überlebenswahrscheinlichkeit des Parallellsystem ist:

$$
\begin{aligned}
P(T_s \geq 200) &= 1 - P(T_s < 200) = e^{-\frac{1}{500} \cdot 200} + e^{-\frac{1}{300} \cdot 200} - e^{-(\frac{1}{500} + \frac{1}{300}) \cdot 200} \\
&= 0.8396.
\end{aligned}
$$

Aufgabe 29

Die Zufallsvariable X sei gleichverteilt auf dem Intervall $[1,3]$. $Y = (Y_1, \ldots, Y_n)$ sei eine Stichprobe mit Zurücklegen zu X.

(a) Bestimmen Sie die Verteilung von Y und die Randverteilung von Y_i, $i = 1, \ldots, n$.

(b) Berechnen Sie die Dichte der Spannweite R von Y.

(c) Geben Sie den Erwartungswert von R an.

Lösung: (Wahrscheinlichkeitstheorie, S.161ff., S.139ff., S.191ff., S.79ff.)

(a) Für die gemeinsame Verteilungsfunktion F_Y gilt:

$$F_Y(y_1, \ldots, y_n) = \prod_{i=1}^{n} F_X(y_i)$$

Da X gleichverteilt ist, gilt:

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{für} \quad x \in [1,3] \\ 0 & \quad\text{sonst} \end{cases}$$

$$F_X(\alpha) = \begin{cases} 0 & \text{für} \quad \alpha \in [-\infty, 1] \\ \frac{1}{2}(\alpha - 1) & \text{für} \quad \alpha \in [1,3] \\ 1 & \quad \text{sonst} \end{cases}$$

Damit ergibt sich:

$$F_Y(y_1, \ldots, y_n) = \begin{cases} 0 & \text{für} \quad y_i < 1 \text{ für mindestens ein } i \in \{1, \ldots, n\} \\ \frac{1}{2^n} \prod_{i=1}^{n} (y_i - 1) & \text{für} \quad y_i \in [1,3] \text{ für alle } i = 1, \ldots, n \\ \frac{1}{2^{\#I}} \prod_{i \in I} (y_i - 1) & \text{für} \quad y_i \in [1,3] \text{ für } i \in I \subseteq \{1, \ldots, n\} \text{ und } y_i > 3 \\ & \text{für} \quad \text{alle } i \in \{1, \ldots, n\} \setminus I \\ 1 & \quad \text{sonst} \end{cases}$$

und als Randverteilung von Y_i, $i = 1, \ldots, n$:

$$
\begin{aligned}
F_{Y_i}(y_i) &= \lim_{y_j \to \infty; j \neq i} F_Y(y_1, \ldots, y_n) = F_X(y_i) \\
&= \begin{cases} 0 & \text{falls } y_i < 1 \\ \frac{1}{2}(y_i - 1) & \text{für } y_i \in [1, 3] \\ 1 & y_i \in [3, \infty) \end{cases}
\end{aligned}
$$

(b) $Y = (Y_1, \ldots, Y_n)$ ist nach Voraussetzung eine Stichprobe mit Zurücklegen zu X. Damit ist die gemeinsame Dichte von $\max\{Y_i\}$ und $\min\{Y_i\}$: (s. Wahrscheinlichkeitstheorie, S. 192)

$$
\begin{aligned}
f(x, y) &= \begin{cases} n(n-1)(\frac{1}{2}(x-y))^{n-2}\frac{1}{2}\frac{1}{2} & \text{für} \quad 1 \leq y < x < 3 \\ 0 & \text{sonst} \end{cases} \\
&= \begin{cases} (\frac{1}{2})^n n(n-1)(x-y)^{n-2} & \text{für} \quad 1 \leq y < x < 3 \\ 0 & \text{sonst} \end{cases}
\end{aligned}
$$

Die Dichte der Spannweite ergibt sich daraus zu:

$$
\begin{aligned}
f_R(z) &= \int_{-\infty}^{\infty} f(x, x-z) dx \\
&= \begin{cases} 0 & \text{für} \quad z \leq 0 \text{ (da sonst } y = x - z \geq x) \\ \int_{1+z(*)}^{3} (\frac{1}{2})^n n(n-1)(x-(x-z))^{n-2}\, dx & \text{für} \quad 0 < z < 2 \\ 0 & \text{für} \quad z \geq 2 \text{ (da sonst mit x} < 3 \text{ gilt,} \\ & \qquad \text{dass } x - z < 1 \text{ ist.)} \end{cases}
\end{aligned}
$$

Bemerkung zu $(*)$: Aus $1 \leq y = x - z$ folgt $x \geq z + 1$.

Das Integral im zweiten Funktionsast ergibt:

$$
\begin{aligned}
\int_{z+1}^{3} (\frac{1}{2})^n n(n-1)(x-(x-z))^{n-2}\, dx &= (\frac{1}{2})^n n(n-1) \int_{z+1}^{3} z^{n-2}\, dx \\
&= (\frac{1}{2})^n n(n-1) \left(xz^{n-2}\right) \Big|_{z+1}^{3} \\
&= (\frac{1}{2})^n n(n-1)(2-z)z^{n-2}
\end{aligned}
$$

(c) Der Erwartungswert berechnet sich gemäß:

$$
\begin{aligned}
E(R) &= \int\limits_{-\infty}^{+\infty} z f_R(z)\,dz = \int\limits_{0}^{2} z(\frac{1}{2})^n n(n-1)(2-z)z^{n-2}\,dz \\
&= (\frac{1}{2})^n n(n-1) \int\limits_{0}^{2} (2z - z^2) z^{n-2}\,dz = (\frac{1}{2})^n n(n-1) \int\limits_{0}^{2} (2z^{n-1} - z^n)\,dz \\
&= (\frac{1}{2})^n n(n-1) \left(\frac{2}{n} z^n - \frac{1}{n+1} z^{n+1} \right) \Big|_0^2 = (\frac{1}{2})^n n(n-1) \left(\frac{2}{n} 2^n - \frac{1}{n+1} 2^{n+1} \right) \\
&= 2n(n-1) \left(\frac{1}{n} - \frac{1}{n-1} \right) = 2\frac{n-1}{n+1}
\end{aligned}
$$

Aufgabe 30

(a) Gegeben seien zwei unabhängige diskrete Zufallsvariablen X und Y mit den in der Aufgabe 26 definierten Wahrscheinlichkeitsverteilungen. Bestimmen Sie die Wahrscheinlichkeitsverteilung von $Z = XY$.

(b) Betrachten Sie nun die beiden unabhängigen stetigen Zufallsvariablen X und Y mit den Dichtefunktionen

$$f_X(x) = \begin{cases} \frac{2}{15}x & \text{für} \quad 1 \leq x \leq 4 \\ 0 & \text{sonst} \end{cases},$$

$$f_Y(y) = \begin{cases} \frac{1}{12}y & \text{für} \quad 1 \leq y \leq 5 \\ 0 & \text{sonst} \end{cases}.$$

(i) Bestimmen Sie die Dichtefunktion des Produkts $Z = XY$.

(ii) Überlegen Sie sich, wie Sie die Dichte von $Z = \frac{X}{Y}$ berechnen können.

Lösung: (Wahrscheinlichkeitstheorie, S.194ff.)

(a) Für das Produkt erhält man analog zur Summe im diskreten Fall:

$$P(\prod_{i=1}^{n} Y_i = z) = \sum_{(y_1,\ldots,y_n): \quad y_i = z} P((Y_1 \ldots Y_n) = (y_1 \ldots y_n))$$

Die Werte der Zufallsvariable XY mit X und Y aus Aufgabe 14(a) ergeben sich gemäß folgender Tabelle

$y \setminus x$	1	2	3
-1	-1	-2	-3
0	0	0	0
1	1	2	3

zu: -3,-2,-1,0,1,2,3.

Die Wahrscheinlichkeitsverteilung ergibt sich nach der Tabelle zu

$$P(Z = -3) = P(X \cdot Y = -3) = P((X, Y) = (3, -1)) = 0.15$$

und analog für die Werte -2,-1,1,2,3 sowie

$$P(Z = 0) = P(X, Y) \in \{(1, 0), (2, 0), (3, 0)\} = 0.4.$$

Insgesamt lautet die Wahrscheinlichkeitsverteilung damit:

z	-3	-2	-1	0	1	2	3
$P(Z=z)$	0.15	0.06	0.09	0.4	0.09	0.06	0.15

(b) (i) Dichtefunktion für $Z = X \cdot Y$ (s. Wahrscheinlichkeitstheorie, S. 195) ist:

$$f_Z(z) \quad = \quad \int\limits_{-\infty}^{\infty} f_{(X,Y)}(x, \frac{z}{x}) \frac{1}{|x|}\, dx$$

bzw. da X und Y unabhängig sind, ist die Dichtefunktion:

$$f_Z(z) \quad = \quad \int\limits_{-\infty}^{\infty} f_X(x) f_Y(\frac{z}{x}) \frac{1}{|x|}\, dx$$

Für welche Werte von x zu vorgegebenem z ist der Integrand von Null verschieden? Dies ist genau dann der Fall, wenn $f_X(x) \neq 0$ ist (also für $x \in [1,4]$) und wenn $f_Y(\frac{z}{x}) \neq 0$ ist (also für $\frac{z}{x} \in [1,5]$). Damit muß $x \geq 1$ und $\frac{z}{x} \leq 5$ d.h. $x \geq \frac{z}{5}$ sein, also $x \geq \max\{1, \frac{z}{5}\}$. Ferner folgt aus $x \leq 4$ und $\frac{z}{x} \geq 1$, d.h. $x \leq z$, $x \leq \min\{4, z\}$. Damit ergibt sich für die Dichtefunktion:

$$\begin{aligned}
f_Z(z) \quad &= \quad \int\limits_{\max\{1,\frac{z}{5}\}}^{\min\{4,z\}} \frac{2}{15}x \cdot \frac{1}{12}\frac{z}{x}\frac{1}{x}\, dx \text{ für } z = x \cdot y \in [1 \cdot 1 = 1,\ 4 \cdot 5 = 20] \\
&= \quad \int\limits_{\max\{1;\frac{z}{5}\}}^{\min\{4,z\}} \frac{1}{90}\frac{z}{x} dx \\
&= \quad \frac{1}{90} z (\ln x)\Big|_{\max\{1,\frac{z}{5}\}}^{\min\{4,z\}} \\
&= \quad z(\ln(min\{4,z\}) - \ln(max\{1, \frac{z}{5}\}))
\end{aligned}$$

$$\Rightarrow f_Z(z) \quad = \quad \begin{cases}
0 & \text{für} \quad z < 1 \quad (\max\{1,\frac{z}{5}\} = 1 > \min\{4,z\} = z) \\
\frac{1}{90} z(\ln z - 0) & \text{für} \quad z \in [1,4) \quad (\max\{1,\frac{z}{5}\} = 1, \min\{4,z\} = z) \\
\frac{1}{90} z(\ln 4 - 0) & \text{für} \quad z \in [4,5) \quad (\max\{1,\frac{z}{5}\} = 1, \min\{4,z\} = 4) \\
\frac{1}{90} z(\ln 4 - \ln\frac{z}{5}) & \text{für} \quad z \in [5,20] \quad (\max\{1,\frac{z}{5}\} = \frac{z}{5}, \min\{4,z\} = 4) \\
0 & \text{für} \quad z > 20 \quad (\max\{1,\frac{z}{5}\} = \frac{z}{5} > \min\{4,z\} = 4)
\end{cases}$$

(ii) Für den Quotienten Z zweier Zufallsvariablen X und Y gilt: $Z = X \cdot (\frac{1}{Y})$, d.h. im wesentlichen kann wie beim Produkt vorgegangen werden, wobei für die Zufallsvariable Y der Wert y=0 ausgeschlossen werden muss.

Die Dichte ergibt sich also zu (vgl. Wahrscheinlichkeitstheorie S. 203):

$$f_Z(z) = \int\limits_{-\infty}^{\infty} f(zy, y)|y|\, dy$$

Bei Unabhängigkeit von X und Y also $f_Z(z) = \int\limits_{-\infty}^{\infty} f_X(zy) f_Y(y) |y| \, dy$

Hier sind:

$$f_X(x) \neq 0 \text{ für } x \in [1,4] \text{ und } f_Y(y) \neq 0 \text{ für } y \in [1,5].$$

Der Integrand ist genau dann positiv, falls $zy \in [1,4]$ und $y \in [1,5]$ d.h. $y \in [\frac{1}{z}, \frac{4}{z}]$ und $y \in [1,5]$ ist. Damit gilt:

$$\begin{aligned} f_Z(z) &= \int\limits_{\max\{1;\frac{1}{z}\}}^{\min\{\frac{4}{z},5\}} f_X(zy) f_Y(y) |y| \, dy \\ &= \int\limits_{\max\{1,\frac{1}{z}\}}^{\min\{\frac{4}{z},5\}} \frac{1}{90} z y^3 \, dy \\ &= \frac{1}{360} z y^4 \Big|_{\max\{1,\frac{1}{z}\}}^{\min\{\frac{4}{z},5\}} \end{aligned}$$

und somit

$$f_Z(z) = \begin{cases} 0 & , z < \frac{1}{5} \text{ (dann ist } \frac{1}{z} = \max\{1, \frac{1}{z}\} > \min\{\frac{4}{z}, 5\} = 5) \\ \frac{1}{360} z \left(5^4 - (\frac{1}{z})^4\right) & , z \in [\frac{1}{5}, \frac{4}{5}] \ (\max\{1, \frac{1}{z}\} = \frac{1}{z}, \min\{\frac{4}{z}, 5\} = 5) \\ \frac{1}{360} z \left((\frac{4}{z})^4 - (\frac{1}{z})^4\right) & , z \in [\frac{4}{5}, 1] \ (\max\{1, \frac{1}{z}\} = \frac{1}{z}, \min\{\frac{4}{z}, 5\} = \frac{4}{z}) \\ \frac{1}{360} z \left((\frac{4}{z})^4 - (2)^4\right) & , z \in [1,4] \ (\max\{1, \frac{1}{z}\} = 1, \min\{\frac{4}{z}, 5\} = \frac{4}{z}) \\ 0 & , z > 4 \text{ (dann ist } \max\{1, \frac{1}{z}\} = 1 > \min\{\frac{4}{z}, 5\} = \frac{4}{z}) \end{cases}$$

Aufgabe 31

(a) Gegeben sei eine diskrete Zufallsvariable $X = (X_1, X_2)$, deren Komponenten unabhängig sind mit den im folgenden angegebenen Randverteilungen. Bestimmen Sie die Wahrscheinlichkeitsverteilung einer Zufallsvariablen $Y = (Y_1, Y_2)$ unter der Annahme, dass $Y_1 = X_1$ und $Y_2 = X_1 + X_2$ gilt.

x_1	-1	0	1
$P(X_1 = x_1)$	0.3	0.4	0.3

x_2	-1	0	1
$P(X_2 = x_2)$	0.3	0.4	0.3

(b) Sei X_1 eine auf dem Intervall $[1, 4]$ gleichverteilte Zufallsvariable und X_2 eine Zufallsvariable mit folgender Dichtefunktion:

$$f_{X_2}(x_2) = \begin{cases} 2 - 0.5 \cdot x_2 & \text{für} \quad 2 \leq x_2 \leq 4 \\ \\ 0 & \text{sonst} \end{cases} .$$

(i) Bestimmen Sie die Dichtefunktion der zweidimensionalen Zufallsvariablen $Y = (Y_1, Y_2)$ unter der Annahme, dass X_1 und X_2 unabhängig sind und $Y_1 = X_1$ sowie $Y_2 = X_1 + X_2$ gilt.

(ii) Bestimmen Sie den Erwartungswert von Y.

Lösung: (Wahrscheinlichkeitstheorie, S.199ff.)

(a) $Y = (Y_1, Y_2)$ entsteht durch die Transformation $g(x_1, x_2) = (x_1, x_1 + x_2)$ aus der Zufallsvariable $X =(X_1, X_2)$. Die Funktion g ist umkehrbar eindeutig, d.h. die Wahrscheinlichkeitsverteilung von $g(X)$ kann wie folgt berechnet werden:

$$P(g(X) = y) = P(X = g^{-1}(y))$$

Werte von Y_2 sind:

$Y_2 =$	-2	-1	0	1	2
entsteht aus $(X_1, X_2) =$	(-1,-1)	(-1,0) (0,-1)	(-1,1) (1,-1) (0;0)	(1,0) (0,1)	(1,1)

Damit ergibt sich unter der Berücksichtigung von

$$P((X_1, X_2) = (x_1, x_2)) = P(X_1 = x_1)P(X_2 = x_2)$$

folgende Wahrscheinlichkeitsverteilung für Y:

(y_1, y_2)	(-1,-2)	(-1,-1)	(-1,0)	(0,-1)	(0,0)	(0,1)	(1,0)	(1,1)	(1,2)
$g^{-1}(y_1, y_2)$	(-1,-1)	(-1,0)	(-1,1)	(0,-1)	(0,0)	(0,1)	(1,-1)	(1,0)	(1,1)
$P(X = g^{-1}(y))$	0.3^2	$0.3 \cdot 0.4$	0.3^2	$0.4 \cdot 0.3$	0.4^2	$0.4 \cdot 0.3$	0.3^2	$0.3 \cdot 0.4$	0.3^2

bzw. in anderer Darstellung:

y_2 / y_1	-2	-1	0	1	2
-1	0.09	0.12	0.09	0	0
0	0	0.12	0.16	0.12	0
1	0	0	0.09	0.12	0.09

Die einzelnen Komponenten berechnen sich z.B. wie folgt:

$$\begin{aligned}((Y_1, Y_2) = (1, 0)) &= P(Y_1 = 1 \text{ und } Y_2 = 0) = P(X_1 = 1 \text{ und } X_1 + X_2 = 0)\\ &= P(X_1 = 1 \text{ und } X_2 = -1) = (0.3)^2\\ P((Y_1, Y_2) = (1, -2)) &= P(Y_1 = 1 \text{ und } Y_2 = -2) = P(X_1 = 1 \text{ und } X_1 + X_2 = -2)\\ &= P(X_1 = 1 \text{ und } X_2 = -3) = 0.3 \cdot 0 = 0.\end{aligned}$$

(b) Gegeben sind die folgenden beiden Dichtefunktionen:

$$f_{X_1}(x_1) = \begin{cases} \frac{1}{3} & \text{für } \; x_1 \in [1, 4] \\ 0 & \text{sonst} \end{cases}$$

$$f_{X_2}(x_2) = \begin{cases} 2 - 0.5x_2 & \text{für } \; x_2 \in [2, 4] \\ 0 & \text{sonst} \end{cases}$$

Da X_1 und X_2 unabhängig sind, gilt: $f_{(X_1,X_2)}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$, d.h.

$$f_{(X_1,X_2)}(x_1, x_2) = \begin{cases} \frac{1}{3}(2 - 0.5x_2) & \text{für } \; x_1 \in [1, 4] \text{ und } x_2 \in [2, 4] \\ 0 & \text{sonst} \end{cases}$$

Die Transformation lautet:

$$X = \binom{X_1}{X_2} \mapsto Y = \binom{Y_1}{Y_2} = \binom{X_1}{X_1 + X_2} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \binom{X_1}{X_2}$$

Es handelt sich hierbei um eine lineare Transformation. Die Transformation $g(x_1, x_2) = (g_1(x_1, x_2), g_2(x_1, x_2))$ ist daher gegeben durch

$$\begin{aligned} y_1 &= g_1(x_1, x_2) &= x_1 \\ y_2 &= g_2(x_1, x_2) &= x_1 + x_2 \end{aligned}$$

Die Ableitung von g lautet:

$$\begin{aligned} g'(a) &= \begin{pmatrix} \frac{\partial g_1(a)}{\partial x_1} & \frac{\partial g_1(a)}{\partial x_2} \\ \frac{\partial g_2(a)}{\partial x_1} & \frac{\partial g_2(a)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \text{ für alle } a \in (1, 4) \times (2, 4), \text{ d.h.} \\ \det g'(a) &= 1 \neq 0 \text{ für } a \in [1, 4] \times [2, 4] \end{aligned}$$

Die Umkehrtransformation lautet:

$$\begin{array}{rclcl} x_1 & = & g_1^{-1}(y_1, y_2) & = & y_1 \\ x_2 & = & g_2^{-1}(y_1, y_2) & = & y_2 - y_1 \end{array}$$

Hieraus ergibt sich nun die transformierte Dichte gemäß dem Transformationssatz:

$$\begin{aligned} f_{(Y_1,Y_2)}(y) &= \begin{cases} f_{X_1,X_2}(g_1^{-1}(y_1,y_2))\dfrac{1}{|\det(g'(g^{-1}(y)))|} & \text{für} \quad g_1^{-1}(y_1,y_2) = y_1 \in [1,4] \text{ und} \\ & \quad\;\; g_2^{-1}(y_1,y_2) = y_2 - y_1 \in [2,4] \\ 0 & \text{sonst} \end{cases} \\ &= \begin{cases} \frac{1}{3}(2 - 0.5(y_2 - y_1))\frac{1}{|1|} & \text{für} \quad y_1 \in [1,4] \text{ und } y_2 \in [2 + y_1, 4 + y_1] \\ 0 & \text{sonst} \end{cases} \\ &= \begin{cases} \frac{2}{3} - \frac{1}{6}(y_2 - y_1) & \text{für} \quad y_1 \in [1,4] \text{ und } y_2 \in [2 + y_1, 4 + y_1] \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

Bekanntlich gilt: $E(Y) = (E(Y_1), E(Y_2))$ d.h. der Erwartungswert von Y kann komponentenweise gebildet werden. Da $Y_1 = X_1$ ist, folgt:

$$E(Y_1) = E(X_1) = 2.5$$

Der Erwartungswert von Y_2 kann gemäß

$$E(Y_2) = E(X_1 + X_2) = E(X_1) + E(X_2)$$

aus dem Erwartungswert von X_1 und X_2 berechnet werden. Für den zweiten Summand erhält man:

$$\begin{aligned} E(X_2) &= \int_2^4 x_2(2 - 0.5x_2)\, dx_2 = x_2^2 - \frac{1}{6}x_2^3\Big|_2^4 \\ &= 16 - \frac{32}{3} - 4 + \frac{4}{3} = \frac{8}{3} \end{aligned}$$

und somit gilt:

$$E(Y_2) = \frac{5}{2} + \frac{8}{3} = \frac{31}{6}$$

Aufgabe 32

Sei $X = (X_1, X_2)$ eine bivariate Zufallsvariable mit folgender Dichtefunktion:

$$f_X(x_1, x_2) = \begin{cases} 0.1 \cdot x_1 x_2 & \text{für} \quad 1 \leq x_1 \leq 3,\ 2 \leq x_2 \leq 3 \\ \\ 0 & \text{sonst} \end{cases} .$$

Bestimmen Sie die Dichtefunktion der Zufallsvariablen $Y = (Y_1, Y_2)$ mit $Y_1 = X_1$ und $Y_2 = X_1 \cdot X_2$.

Lösung: (Wahrscheinlichkeitstheorie, S.199ff.)

Die Transformation h mit $y = h(x)$ besitzt die folgende Form:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} h_1(x_1, x_2) \\ h_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_1 \cdot x_2 \end{pmatrix}$$

Die partiellen Ableitungen existieren und es gilt:

$$\det(h'(x)) = \det \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \frac{\partial h_1}{\partial x_2} \\ \frac{\partial h_2}{\partial x_1} & \frac{\partial h_2}{\partial x_2} \end{pmatrix} = \det \begin{pmatrix} 1 & 0 \\ x_2 & x_1 \end{pmatrix} = x_1 \neq 0 \quad \text{für } x_1 \in [1, 3]$$

Die Umkehrtransformation lautet:

$$x_1 = h_1^{-1}(y_1, y_2) = y_1;\ x_2 = h_2^{-1}(y_1, y_2) = \frac{y_2}{y_1}$$

Damit ist $\det(h'(h^{-1}(y)) = h_1^{-1}(y_1, y_2) = y_1$ und für $y_1 = h_1^{-1}(y_1, y_2) \in [1, 3]$ und $\frac{y_2}{y_1} = h_2^{-1}(y_1, y_2) \in [2, 3]$ gilt:

$$\begin{aligned} f_Y(y_1, y_2) &= f_X(h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2)) \cdot \frac{1}{|\det(h'(h^{-1}(y_1, y_2)))|} \\ &= 0.1 \cdot y_1 \frac{y_2}{y_1} \cdot \frac{1}{y_1} = 0.1 \cdot \frac{y_2}{y_1} \end{aligned}$$

Die Dichtefunktion von (Y_1, Y_2) ist somit:

$$f_Y(y_1, y_2) = \begin{cases} 0.1 \cdot \frac{y_2}{y_1} & \text{für } y_1 \in [1, 3],\ y_2 \in [2y_1, 3y_1] \\ 0 & \text{sonst} \end{cases}$$

Aufgabe 33

Bestimmen Sie die Wahrscheinlichkeitsverteilung, den Erwartungswert und die Varianz der Summe einer $B(m,p)$-verteilten Zufallsvariablen X und einer $B(n,p)$-verteilten Zufallsvariablen Y. (X und Y seien unabhängig)

Lösung: (Wahrscheinlichkeitstheorie, S.182ff.)

Gesucht ist die Wahrscheinlichkeitsverteilung der Summe zweier unabhängiger binomialverteilter Zufallsvariablen, es handelt sich also um eine Faltung (diskreter Fall).

$$\begin{array}{ccc} X \sim Bin(n,p) & , & Y \sim Bin(m,p) \\ P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} & , & P(Y=y) = \binom{m}{y} p^y (1-p)^{m-y} \end{array}$$

Für $0 \leq z \leq n+m$ besitzt $Z = X+Y$ die Wahrscheinlichkeitsverteilung:

$$\begin{aligned} P(Z=z) &= P(X+Y=z) \\ &= \sum_{x,y\geq 0,\ x+y=z} P(X=x)P(Y=y) \\ &= \sum_{x=0}^{z} \binom{n}{x} p^x (1-p)^{n-x} \binom{m}{z-x} p^{z-x} (1-p)^{m-z+x} \\ &= \sum_{x=0}^{z} \binom{n}{x} \binom{m}{z-x} \cdot p^z (1-p)^{m+n-z} \\ &= \binom{n+m}{z} p^z (1-p)^{m+n-z}, \text{ da } \sum_{x=0}^{z} \binom{n}{x} \binom{m}{z-x} = \binom{n+m}{z} \text{ ist.} \end{aligned}$$

Z ist damit $Bin(n+m,p)$-verteilt. Daraus folgt:

$$E(Z) = (n+m)p, \quad Var(Z) = (n+m)p(1-p).$$

Aufgabe 34

Eine Maschine produziert Stahlstifte mit einer Soll-Länge von $\mu = 110$ mm. Ungenauigkeiten bei der Produktion können leider nicht ausgeschlossen werden, so dass die Längen der produzierten Stifte Realisationen einer Zufallsvariablen X darstellen. Die Streuung der realisierten Stiftlängen ist mit $\sigma = 0.1$ mm bekannt.

Um sich ein Bild von der Güte der Produktionseinheiten zu machen, wird die Wahrscheinlichkeit dafür, dass X höchstens um den Betrag c von μ abweicht, betrachtet. c wird dabei in der Regel in Vielfachen der Standardabweichung σ des Prozesses angegeben, d.h. gesucht ist $P(X \in [\mu - k\sigma, \mu + k\sigma])$.

Bestimmen und vergleichen Sie für $k = 1, 2, 3$ die Wahrscheinlichkeiten für die Fälle, dass

(a) über den Typ der Verteilung von X keine weiteren Kenntnisse verfügbar sind bzw.

(b) X normalverteilt ist mit Mittelwert μ und Varianz σ^2.

Wie erklären Sie sich Ihre Ergebnisse?

Lösung: (Wahrscheinlichkeitstheorie, S.208ff.)

(a) Da keine genaue Kenntnisse über den Verteilungstyp vorliegt, erfolgt die Abschätzung über die Tschebyscheffsche Ungleichung:

$$\begin{aligned} P(X \in [\mu - k\sigma; \mu + k\sigma]) &= P(|X - \mu| \leq k\sigma) \\ &= 1 - P(|X - \mu| > k\sigma) \\ &\overset{Tschebyscheff}{\geq} 1 - \frac{Var(X)}{k^2\sigma^2} = 1 - \frac{1}{k^2} \end{aligned}$$

Mit $k = 1, 2, 3$ ergibt sich:

$$\begin{aligned} P(X \in [\mu - \sigma; \mu + \sigma]) &\geq 1 - \frac{1}{1} = 0 \\ P(X \in [\mu - 2\sigma; \mu + 2\sigma]) &\geq 1 - \frac{1}{4} = 0.75 \\ P(X \in [\mu - 3\sigma; \mu + 3\sigma]) &\geq 1 - \frac{1}{9} = 0.\overline{8} \ (= \frac{8}{9}) \end{aligned}$$

Für $k = 1$ ist die Abschätzung über Tschebyscheff trivial und ohne Aussagekraft.

(b) Ist die Verteilung konkret bekannt, so lassen sich die Wahrscheinlichkeiten genau berechnen. Es gilt nun: $X \sim N(110, 0.1^2)$.

$$\begin{aligned} P(X \in [\mu - k\sigma; \mu + k\sigma]) &= \Phi\left(\frac{\mu + k\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - k\sigma - \mu}{\sigma}\right) \\ &= \Phi(k) - \Phi(-k) = 2\Phi(k) - 1 \end{aligned}$$

Für die konkreten Werte in der Aufgabe ergibt sich:

$$\begin{aligned} P(X \in [\mu - \sigma; \mu + \sigma]) &= 2\Phi(1) - 1 = 0.6826 \\ P(X \in [\mu - 2\sigma; \mu + 2\sigma]) &= 2\Phi(2) - 1 = 0.9544 \\ P(X \in [\mu - 3\sigma; \mu + 3\sigma]) &= 2\Phi(3) - 1 = 0.9974 \end{aligned}$$

Die Abschätzungen in (a) gelten für jede Verteilung von X. Daher ist es auch, wenn die Verteilung nicht bekannt ist, nur möglich eine untere Schranke für die Wahrscheinlichkeit anzugeben. Ist die Verteilung bekannt, so ist die Wahrscheinlichkeit (zumindest mit numerischer Genauigkeit) exakt anzugeben. Der Wert wird auf jeden Fall größer oder gleich der unteren Schranke sein. Bei der Normalverteilung erhalten wir erheblich größere Werte wegen der Symmetrie um μ und der hohen Dichtewerte im Bereich um μ.

Aufgabe 35

(a) Bei einem Großraumflugzeug ist die Auslastung pro Flug näherungsweise normalverteilt. Im Mittel fliegen 150 Passagiere mit dem Flugzeug, die Auslastung schwankt mit $\sigma = 25$ Passagiere. Mit welcher Anzahl von Fluggästen ist mit einer Wahrscheinlichkeit von 90% mindestens zu rechnen?

(b) Welche Abschätzung können Sie vornehmen, wenn der Typ der Verteilung nicht bekannt ist?

Lösung: (Wahrscheinlichkeitstheorie, S.86ff., S.102f., S.208f.)

(a) Die Anzahl X der Passagiere pro Flug ist näherungsweise $N(150; 25^2)$-verteilt. Gesucht ist: x_0 mit $P(X \geq x_0) \stackrel{!}{=} 0.9$. Sei Φ die Verteilungsfunktion der Standardnormalverteilung, dann gilt

$$\begin{aligned} P(X \geq x_0) &= 1 - P(X \leq x_0) = 1 - F_X(x_0) \\ &= 1 - \Phi\left(\frac{x_0 - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{x_0 - 150}{25}\right) \stackrel{!}{=} 0.9 \end{aligned}$$

Gesucht ist also x_0 mit $\Phi(\frac{x_0-150}{25}) \stackrel{!}{=} 0.1$. Man erhält:

$$\begin{aligned} \frac{x_0 - 150}{25} &= \Phi^{-1}(0.1) = -1.28 \ (= -\Phi^{-1}(1 - 0.1)) \\ \Rightarrow x_0 &= -1.28 \cdot 25 + 150 = 118 \end{aligned}$$

Es ist also mit einer Wahrscheinlichkeit von 90% mit mindestens 118 Passagieren zu rechnen.

(b) Ist die Verteilung nicht bekannt, so kann eine Abschätzung mit der Tschebyscheffschen Ungleichung durchgeführt werden. Gesucht ist also c mit

$$P(|X - 150| < c) \geq 1 - \frac{Var(X)}{c^2} \stackrel{!}{=} 0.9$$

Man erhält:

$$1 - \frac{Var(X)}{c^2} = 0.9 \text{ und somit } c = \sqrt{\frac{Var(X)}{0.1}} = \frac{25}{\sqrt{0.1}} = 79.06$$

Daraus folgt: Mit mindestens 90% Wahrscheinlichkeit gilt: $|X - 150| < 79.06$. Da X ganzzahlig ist, ist dies gleichwertig mit $X \in (70, 230)$, d.h. mit mindestens 90% Wahrscheinlichkeit gilt: $70 < X < 230$.

Bemerkung: Eine genauere Abschätzung ist mit Hilfe der Tschebyscheffschen Ungleichung nicht möglich, da über die Verteilung nichts bekannt ist! Insbesondere kann die Wahrscheinlichkeit für das gesuchte einseitig offene Intervall nur sehr ungenau bestimmt werden.

Aufgabe 36

(a) Erläutern Sie anhand eines idealen Würfels das "schwache Gesetz der großen Zahlen".

(b) Verdeutlichen Sie sich die wesentlichen Aussagen des zentralen Grenzwertsatzes.

(c) Eine Vertriebsgesellschaft besitzt in einer Großstadt 200 Zigarettenautomaten. Jeder Automat hat (unabhängig von den anderen) mit der Wahrscheinlichkeit $\frac{1}{20}$ pro Woche eine Störung. Für die Entscheidung über die Größe eines ständigen Reparaturtrupps sei die Wahrscheinlichkeit dafür von Interesse, dass in einer Woche die Anzahl X der defekten Automaten zwischen 5 und 15 liegt, von Interesse. Diese Wahrscheinlichkeit (der exakte Wert beträgt übrigens 0.9292) soll

(i) mittels der Poissonverteilung approximiert werden,

(ii) über die Ungleichung von Tschebyscheff nach unten abgeschätzt werden und

(iii) mit Hilfe des zentralen Grenzwertsatzes approximativ berechnet werden.

Lösung: (Wahrscheinlichkeitstheorie, S.212f., S.217ff., S.208f.)

(a) Das schwache Gesetz der großen Zahlen bezogen auf einen idealen Würfel besagt:

Bei wiederholtem unabhängigen Werfen eines Würfels konvergiert das arithmetische Mittel der gewürfelten Zahlen stochastisch gegen den Erwartungswert ($E(X) = 3.5$). D.h. für großes n ist die Wahrscheinlichkeit für eine größere Abweichung des arithmetischen Mittels vom Erwartungswert klein, oder formal:

Es bezeichne X_i, $i = 1, 2, \ldots$ eine Folge unabhängiger identisch verteilter Zufallsvariablen mit einem endlichen Erwartungswert. Dann gilt:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - EX\right| > \epsilon\right) = 0$$

(b) Die Aussage des Zentralen Grenzwertsatzes ist:

Seien $X_1, X_2, X_3, \ldots$ unabhängige Zufallsvariablen, deren Erwartungswerte $\mu_i = E(X_i)$ und Varianzen $\sigma_i^2 = Var(X_i)$ existieren ($i = 1, \ldots, n$). Falls $X_1, X_2, \ldots$ die Lindeberg-Bedingung erfüllen, gilt für die Verteilungsfunktion F_{Z_n} der Folge Z_n mit

$$Z_n = \frac{\sum\limits_{i=1}^{n} X_i - \sum\limits_{i=1}^{n} \mu_i}{\sqrt{\sum\limits_{i=1}^{n} \sigma_i^2}}$$

dass für alle $\alpha \in \mathbb{R}$

$$\lim_{n\to\infty} F_{Z_n}(\alpha) = \Phi(\alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx$$

gilt und die Konvergenz gleichmäßig in α ist. Die Lindeberg- Bedingung lautet:

Mit $D_n^2 = \sum\limits_{i=1}^{n} \sigma_i^2$ gelte für jedes $\varepsilon > 0$:

$$\lim_{n\to\infty} \frac{1}{D_n^2} \sum_{i=1}^{n} \int\limits_{|x_i-\mu_i|>\varepsilon D_n} (x_i - \mu_i)^2 \; dF_i(x) = 0.$$

Interpretation: Jeder der unabhängigen Summanden X_i trägt zur Summe nur einen kleinen Anteil bei, d.h. kein Summand dominiert. Dies gilt z.B. insbesondere falls die X_i identisch verteilt sind.

Die Summe n unabhängiger Zufallsvariablen kann für großes n als näherungsweise normalverteilt angenommen werden, wenn einzelne Summanden keinen "prägenden" Einfluß auf die Summe haben.

Anwendung: z.B. im Bereich der Qualitätssicherung, wo unsystematische Abweichungen einer überwachten Größe von ihrem Sollwert in der Regel als Ergebnis einer Überlagerung vieler zufälliger, für sich unbedeutender Abweichungen aufgefaßt werden (andere Anwendungen z.B. in der Kapitalmarkttheorie, ...)

(c) Die Zufallsvariable X gebe die Anzahl defekter Automaten in einer Woche an, d.h. gesucht ist $P(5 \le X \le 15)$. X ist binomialverteilt mit Parametern 200 und 0.05, d.h.

$$P(5 \le X \le 15) = \sum_{k=5}^{15} \binom{200}{k} (\frac{1}{20})^k (\frac{19}{20})^{200-k} = F_X(15) - F_X(4)$$

(i) Approximation über die Poissonverteilung (gute Näherung, falls $n \ge 50, p \le \frac{1}{10}$; Parameter $\lambda := n{\cdot}p =$ mit $\lambda =$ Erwartungswert von $Poi(\lambda)$ und $n{\cdot}p =$ Erwartungswert von $Bin(n,p)$), d.h. wir betrachten X als näherungsweise Poisson-verteilt mit $\lambda = n{\cdot}p$.

$$P(X = x) = \begin{cases} \frac{(np)^x}{x!} e^{-np} & \text{für } x = 0,1,2,\ldots \\ 0 & \text{sonst} \end{cases}$$

Damit ist

$$P(5 \le X \le 15) = \sum_{k=5}^{15} \frac{10^k}{k!} e^{-10} = 0.992$$

(ii) Schranke mit Hilfe der Tschebyscheffschen Ungleichung:

$$P(|X - E(X)| < c) \ge 1 - \frac{Var(X)}{c^2}$$

$$\begin{aligned} E(X) &= n \cdot p = 10 \\ Var(X) &= n \cdot p(1-p) = 9.5 \\ \Longrightarrow P(X \in [5;15]) &\overset{*}{=} P(|X-10| < 6) \ge 1 - \frac{9.5}{6^2} = 0.736 \; (* \text{ da } X \text{ diskret ist}) \\ \text{bzw. } P(|X-10| \le 5) &\ge P(|X-10| < 5) \ge 1 - \frac{9,5}{5^2} \\ &= 0.62 \end{aligned}$$

d.h. die Schranke sollte so groß wie möglich gewählt werden, um eine möglichst gute Abschätzung zu erhalten. Auch hier zeigt sich die ziemlich grobe Abschätzung bei der Tschebyscheffschen Ungleichung.

(iii) Approximation durch die Normalverteilung: X_i sei Bernoulli-verteilt für $i = 1, \dots, n$. Dann ist $\sum_{i=1}^n X_i$ näherungsweise $N(\mu, \sigma^2)-$verteilt mit $\mu = np$ und $\sigma^2 = np(1-p)$, falls n hinreichend groß ist. (Merkregel: relativ gute Abschätzung für $np \geq 5, n(1-p) \geq 5$). Mit $\mu = 10$ und $\sigma^2 = 9.5$ gilt für $X = \sum X_i$:

$$\begin{aligned} P(X \in [5;15]) &= P(X \leq 15) - P(X < 5) = P(X \leq 15.5) - P(X \leq 4.5) \\ &\approx^{1)} \Phi(\frac{15.5-10}{3.0822}) - \Phi(\frac{4.5-10}{3.0822}) = 0.9248. \end{aligned}$$

[1)] Stetigkeitskorrektur, d.h. beim Übergang von diskret nach stetig wird die Approximation verbessert, wenn als Schranken nicht 15 oder 5 sondern 15.5 und 4.5 gewählt werden, was bei der diskreten Zufallsvariable keine Auswirkung hat.

Aufgabe 37

Ein Unternehmen stellt Taschenlampenbatterien her. Für ein Marketingprojekt soll die Lebensdauer dieser Batterien mit einer Stichprobe untersucht werden, wobei eine Exponentialverteilung unterstellt wird. Geben Sie zu diesem Beispiel die drei Grundannahmen der schließenden Statistik an und erläutern Sie die drei Grundaufgaben. Geben Sie zu jeder der Grundaufgaben eine konkrete Fragestellung für das Unternehmen an, die zu dieser Aufgabe führt, und wie eine Entscheidungsfunktion dazu aussieht. Geben Sie jeweils eine sinnvoll erscheinende Entscheidungsfunktion an.

Lösung: (Induktive Statistik, S.33ff., S.40ff.)

Grundannahme 1: Der relevante Umweltzustand kann durch eine Zufallsvariable Y beschrieben werden.

Hier: Y sei die Lebensdauer einer zufällig ausgewählten Batterie.

Grundannahme 2: Die Verteilung von Y gehört einer bekannten Klasse von Wahrscheinlichkeitsverteilungen an.

Hier: Klasse der Exponentialverteilungen: $\{Exp(\lambda) | \lambda > 0\}$

Grundannahme 3: Zu Y existiert eine Stichprobe $(X_1, \ldots, X_n)$, deren Verteilung in bekannter Weise von der Verteilung von Y abhängt.

Hier: Verteilung einer einfachen Stichprobe $(X_1, \ldots, X_n)$ vom Umfang n zu Y:

$$P(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} P(X_i \leq x_i) = \prod_{i=1}^{n} P(Y \leq x_i)$$

Grundaufgabe 1: Parameterschätzung

Hier: Man interessiert sich für den unbekannten Parameter λ, da er die Verteilung von Y eindeutig festlegt, d.h. gesucht ist ein Schätzwert $\hat{\lambda}$ für λ anhand der Stichprobenrealisation $(x_1, \ldots, x_n)$.

Grundaufgabe 2: Intervallschätzung

Hier: Anstatt einen konkreten Schätzwert für λ anzugeben, sucht man einen Bereich, der den wahren Parameter mit vorgegebener Wahrscheinlichkeit enthält.

Grundaufgabe 3: Hypothesentests

Hier: Gegeben sind zwei Hypothesen über den unbekannten Parameter λ (z.B. $H_0 : \lambda \leq \lambda_0$, $H_1 : \lambda > \lambda_0$). Auf der Grundlage der Stichprobenrealisation ist eine Entscheidung über die Gültigkeit der Hypothesen zu treffen.

Entscheidungsfunktion zu 1):

$$\delta \; : \; \begin{array}{lcl} \mathcal{X} & \to & \Gamma \\ (x_1, \ldots, x_n) & \mapsto & \delta(x_1, \ldots, x_n) \end{array}$$

hierbei bezeichnet $\mathcal{X}$ den Stichprobenraum und Γ den Parameterraum.

Entscheidungsfunktion zu 2):

$$\delta \quad : \quad \begin{array}{lcl} \mathcal{X} & \rightarrow & \mathcal{P}(\Gamma) \\ (x_1, \ldots, x_n) & \mapsto & \delta(x_1, \ldots, x_n) = A \subseteq \Gamma \end{array}$$

wobei $P(\lambda \in A) \geq 1 - \alpha$ zu vorgegebenem α sei.

Entscheidungsfunktion zu 3)

$$\delta \quad : \quad \begin{array}{lcl} \mathcal{X} & \rightarrow & \{d_0, d_1\} \\ (x_1, \ldots, x_n) & \mapsto & \delta(x_1, \ldots, x_n) \end{array}$$

wobei d_0 und d_1 die Entscheidungen "Hypothese H_0 kann nicht abgelehnt werden bzw. H_0 wird angenommen" und "H_0 wird abgelehnt" symbolisieren.

Aufgabe 38

Ein Briefmarkensammler weiß, dass von einem bestimmten Ersttagsbrief eine gewisse Auflagenhöhe existiert. Von k dieser Ersttagsbriefe, die wie gewöhnlich mit 1 beginnend durchnumeriert wurden, kennt er per Zufall die Nummern $n_1, \ldots, n_k$. Aufgrund dieser Information möchte er einen Schätzwert für die Gesamtauflage N des Ersttagsbriefes erstellen. Begründen Sie, wie die im folgenden vorgeschlagenen Schätzer zustande gekommen sind:

(a) $N_1 = \max\limits_i \; n_i$

(b) $N_2 = \frac{k+1}{k} \; \max\limits_i \; n_i$

(c) $N_3 = \max\limits_i \; n_i + \min\limits_i \; n_i - 1$

(d) $N_4 = 2 \, \frac{1}{k} \sum\limits_{i=1}^{k} n_i$

(e) $N_5 = 2 \, \frac{1}{k} \sum\limits_{i=1}^{k} n_i - 1$

Lösung: (Induktive Statistik, S.40, S.61ff.)

Die Grundgesamtheit ist die Menge der Erstagsbriefe, $\Omega = \{\omega_1, \ldots, \omega_N\}$, N Auflagehöhe der Ersttagsbriefe, betrachtet wird der Laplacesche Wahrscheinlichkeitsraum mit Grundgesamtheit Ω sowie die Zufallsvariable Y, die zu jedem Erstagsbrief die zugehörige Nummer angibt: $Y(\omega_i) = i$ für $i = 1, \ldots, N$. Die bekannten Nummern $n_1, \ldots, n_k$ der Ersttagsbriefe werden als Stichprobenrealisation angesehen, mit deren Hilfe ein Schätzwert für N anzugeben sei.

(a) Schätzung von N mittels größtem Wert in der Stichprobe (ein kleinerer Wert wäre unsinnig).

(b) Hier liegt folgende Überlegung zugrunde: N_1 würde N nur treffen, wenn der größte Wert zufällig in der Stichprobe wäre, was sehr selten der Fall ist. Man wählt daher einen Korrekturfaktor "nach oben". Nachteil: i.a. keine ganze Zahl.

(c) Die Ersttagsbriefe sind folgendermaßen numeriert:
$1, \ldots, \; \min n_i, \ldots, \; \max n_i, \ldots, N$
Man nimmt an, dass der Abstand zwischen dem kleinsten Stichprobenwert und der 1 gleich groß ist, wie der Abstand zwischen der größten Stichprobennummer und dem Wert N, den man schätzen will. Hier wird sowohl der niedrigste als auch der höchste vorliegende Wert berücksichtigt, es wird also mehr Information in diesem Schätzer verarbeitet, als in (a) und (b).

(d) Ähnlich wie bei (c), nur dass jetzt die gesamte Stichprobeninformation verwendet wird. Es wird hier der Mittelwert der Stichprobe errechnet und angenommen, dass dieser auch das Mittel der Gesamtpopulation ist, d.h. dass rechts von ihm genausoviele Nummern liegen, wie links von ihm.

(e) Korrektutterm " -1 " da die Briefe von 1 und nicht von 0 an durchnummeriert sind. Ansonsten, wie (d).

Aufgabe 39

(a) Eine Zufallsvariable sei im Bereich $[0, a]$ gleichverteilt, a ist nicht bekannt und soll mit Hilfe einer Stichprobe mit Zurücklegen geschätzt werden. Überlegen Sie, was die für diese Aufgabe wesentliche Information der Stichprobenwerte $x_1, \dots, x_n$ ist. Überprüfen Sie, ob diese Stichprobenfunktion suffizient ist.

(b) Eine Zufallsvariable X habe die Dichtefunktion

$$f(x) \quad = \quad \begin{cases} 0 & \text{für} \quad x < a \\ \\ \lambda e^{-\lambda(x-a)} & \text{für} \quad x \geq a \end{cases}$$

mit $a > 0$ und bekanntem λ. Lösen Sie die zu (a) analoge Aufgabe.

Lösung: (Induktive Statistik, S.47ff.)

(a) Wesentliche Information ist der größte Stichprobenwert. Die Dichtefunktion des Stichprobenvektors ist

$$\begin{aligned} f(x_1, \dots, x_n) &= \prod_{i=1}^{n} \frac{1}{a} \mathbf{1}_{[0,a]}(x_i) = \frac{1}{a^n} \prod_{i=1}^{n} \mathbf{1}_{[0,a]}(x_i) \\ &= \begin{cases} \dfrac{1}{a^n} & \text{für} \quad \min x_i \geq 0, \max x_i \leq a \\ \\ 0 & \text{sonst} \end{cases} \\ &= \frac{1}{a^n} \mathbf{1}_{[0,a]}(\max_{i=1,\dots,n} x_i) \cdot \mathbf{1}_{[0,\infty)}(\min_{i=1,\dots,n} x_i) \\ &= g(\max_{i=1,\dots,n} x_i, a) \cdot h(x_1, \dots, x_n) \end{aligned}$$

mit

$$g(\max x_i, a) = \frac{1}{a^n} \mathbf{1}_{[0,a]}(\max_{i=1,\dots,n} x_i)$$

und

$$h(x_1, \dots, x_n) = \mathbf{1}_{[0,\infty)}(\min_{i=1,\dots,n} x_i)$$

oder

$$h(x_1, \dots, x_n) = \prod_{i=1}^{n} \mathbf{1}_{[0,\infty)}(x_i)$$

$\max x_i$ ist also eine suffiziente Stichprobenfunktion.

(b) Wesentliche Information über a ist der kleinste der Stichprobenwerte. (Man beachte dabei, dass λ als bekannt vorausgesetzt wurde. Betrachten Sie den Fall "λ und a unbekannt" als Übungsaufgabe!). Die Dichtefunktion des Stichprobenvektors ist:

$$\begin{aligned} f(x_1, \ldots, x_n) &= \prod_{i=1}^{n} \lambda e^{-\lambda(x_i - a)} \mathbf{1}_{[a,\infty)}(x_i) = \lambda^n e^{-\lambda \sum_{i=1}^{n} (x_i - a)} \mathbf{1}_{[a,\infty)}(\min_{i=1,\ldots n} x_i) \\ &= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} e^{n\lambda a} \mathbf{1}_{[a,\infty)}(\min x_i) = g(\min_{i=1,\ldots n} x_i, a) h(x_1, \ldots, x_n) \end{aligned}$$

mit

$$g(\min_{i=1,\ldots n} x_i, a) = e^{n\lambda a} 1_{[a,\infty)}(\min_{i=1,\ldots n} x_i)$$

und

$$h(x_1, \ldots, x_n) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.$$

$\min_{i=1,\ldots n} x_i$ ist also eine suffiziente Statistik.

Aufgabe 40

Bestimmen Sie mit Hilfe des Faktorisierungstheorems von Neyman suffiziente Statistiken für

(a) die Klasse der Exponentialverteilungen

(b) die Klasse der Gleichverteilungen auf dem Intervall $[-\vartheta, \vartheta]$ $(\vartheta > 0)$.

(c) die Klasse der Binomialverteilungen $B(m, p)$

Lösung: (Induktive Statistik, S.47ff.)

Gegeben eine einfache Stichprobe $X = (X_1, \ldots, X_n)$ zu einer Zufallsvariable Y.

(a) Y sei λ-exponentialverteilt.

Dichtefunktion zu X ist

$$\begin{aligned} f_X(x_1, \ldots, x_n) &= \prod_{i=1}^n \lambda e^{-\lambda x_i} \mathbf{1}_{[0,\infty)}(x_i) \\ &= \begin{cases} \lambda^n e^{-\lambda \sum\limits_{i=1}^n x_i} \cdot 1 & \text{für} \quad x_i > 0 \text{ für alle } i = 1, \ldots, n \\ 0 & \text{sonst} \end{cases} \\ &= g\left(\sum_{i=1}^n x_i, \lambda\right) \cdot h(x) \end{aligned}$$

mit

$$\begin{aligned} h(x_1, \ldots, x_n) &= \begin{cases} 1 & \text{für} \quad x_i > 0 \quad \text{für} \quad \text{alle } i = 1, \ldots, n \\ 0 & \text{für} \quad x_i \leq 0 \quad \text{für} \quad \text{mindestens ein } i \end{cases} \\ &= h(x_1, \ldots, x_n) = \mathbf{1}_{[0,\infty)}(\min x_i) = \prod_{i=1}^n \mathbf{1}_{[0,\infty)}(x_i). \end{aligned}$$

Also ist $T(X) = \sum\limits_{i=1}^n x_i$ suffiziente Statistik bzgl. λ.

(b) Y gleichverteilt auf $[-\vartheta, \vartheta], \vartheta > 0$.

Dichtefunktion von X ist:

$$\begin{aligned}
f_X(x_1,\dots,x_n) &= \prod_{i=1}^{n} \frac{1}{2\vartheta}\mathbf{1}_{[-\vartheta,\vartheta]}(x_i) \\
&= \frac{1}{(2\vartheta)^n}\prod_{i=1}^{n}\mathbf{1}_{[0,\vartheta]}(|x_i|) \\
&= \begin{cases} 0 & \max(|x_i|) > \vartheta \\ \frac{1}{(2\vartheta)^n} & \max(|x_i|) \le \vartheta \end{cases} \\
&= \frac{1}{(2\vartheta)^n}\mathbf{1}_{[0,\vartheta]}(\max|x_i|) \\
&= g(\max|x_i|,\vartheta)\cdot\ h(x_1,\dots,x_n) \text{ mit } h(x_1,\dots,x_n) = 1
\end{aligned}$$

$T(X) = \max\limits_i |x_i|$ ist suffiziente Statistik bzgl ϑ.

(c) Y binomialverteilt

Wahrscheinlichkeitsverteilung von X:

$$\begin{aligned}
P(X_1 = k_1, X_2 = k_2, \cdots, X_n = k_n) &= \prod_{i=1}^{n}\binom{m}{k_i}p^{k_i}(1-p)^{m-k_i} \\
&= p^{\sum\limits_{i=1}^{n} k_i}(1-p)^{n\cdot m - \sum\limits_{i=1}^{n} k_i}\cdot\prod_{i=1}^{n}\binom{m}{k_i} \\
&= g\left(\sum_{i=1}^{n} k_i, p\right)\cdot h(k_1,\cdots,k_n)
\end{aligned}$$

$T(k_1,\cdots,k_n) = \sum\limits_{i=1}^{n} k_i$ ist suffiziente Statistik bezüglich p.