INSTITUT FÜR MATHEMATISCHE STOCHASTIK SS 2004
UNIVERSITÄT KARLSRUHE
Priv.-Doz. Dr. D. Kadelka

Klausur

Wahrscheinlichkeitstheorie und Statistik
vom 15.9.2004

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	1	1.8	3.1	3.9	4.7	6.4	7	8.1	8.9	9.7
y_j	6.8	6.9	5.8	4.8	3.9	3.9	1.2	-2	-1	-3.5

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

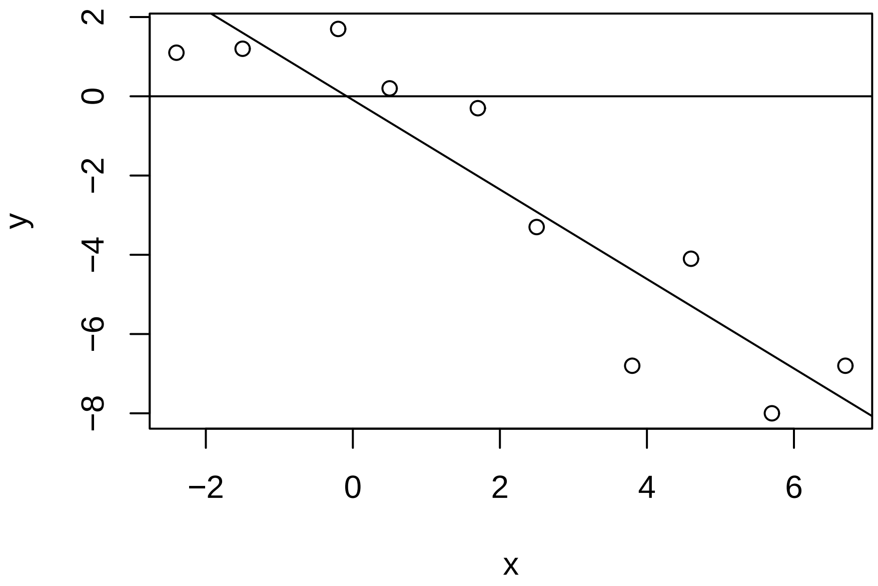
$$\sum_{j=1}^{n} (x_j - \bar{x}) \cdot (y_j - \bar{y}) = \sum_{j=1}^{n} x_j \cdot y_j - n \cdot \bar{x} \cdot \bar{y}$$

$$\bar{x} = 5.46 \qquad s_x = 3.017$$
$$\bar{y} = 2.68 \qquad s_y = 3.769$$
$$r_{xy} = -0.9555$$

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

$$\begin{aligned} b^* &= -1.194 \\ a^* &= 9.2 \end{aligned}$$

und die Regressionsgerade $y = 9.2 - 1.194 \cdot x$.



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

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

$$y_{()} = (-3.5,\ -2,\ -1,\ 1.2,\ 3.9,\ 3.9,\ 4.8,\ 5.8,\ 6.8,\ 6.9)$$

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

$$\bar{y}_{0.2} = \frac{1}{10 - 2 \cdot 2} \cdot (y_{(3)} + \ldots + y_{(8)}) = 3.1$$

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 $k_2 = [7.5] = 7$

$$\begin{aligned} \tilde{y}_{0.25} &= y_{(k_1+1)} = y_{(3)} = -1 \\ \tilde{y}_{0.75} &= y_{(k_2+1)} = y_{(8)} = 5.8 \end{aligned}$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 6.8$.

Aufgabe A2

In einem Rohr soll Wasser nur in einer Richtung fließen. Dazu baut man hintereinander in das Rohr zwei Ventile ein, die das Wasser jeweils nur in ein und derselben Richtung durchlassen sollen. Ventil 1 bzw. 2 ist zu einem bestimmten Zeitpunkt im Zustand $\left\{\begin{matrix} 0 \\ 1 \\ 2 \end{matrix}\right\}$, falls es das Wasser nur in $\left\{\begin{matrix} \text{der gewünschten Richtung} \\ \text{beide Richtungen} \\ \text{keiner Richtung} \end{matrix}\right\}$ durchlässt.

Sei X der zufällige Zustand des ersten, Y der des zweiten Ventils. Es ist also

$$\begin{aligned} A &:= \text{„Die Ventilkombination funktioniert richtig“} \\ &= [X = 0, Y = 0] + [X = 0, Y = 1] + [X = 1, Y = 0] = [X + Y \leq 1]. \end{aligned}$$

Es seien X und Y unabhängig mit

$$\mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = 0.9 \text{ und } \mathbb{P}(X = 1) = \mathbb{P}(Y = 1) = 0.09.$$

a) Bestimmen Sie $\mathbb{P}(Y = 2)$ und $\mathbb{P}(Y = 2 \mid X = 0)$.

b) Berechnen Sie $\mathbb{P}(X = 0, Y = 0)$.

c) Berechnen Sie $\mathbb{P}(X + Y = 0)$, $\mathbb{P}(X + Y = 1)$ und $\mathbb{P}(X + Y \leq 1)$.

d) Bestimmen Sie die Wahrscheinlichkeit, dass die Ventilkombination nicht richtig funktioniert.

Lösung:

a) Da Y nur die Werte $0, 1$ und 2 annimmt, gilt

$$\mathbb{P}(Y = 2) = 1 - \mathbb{P}(Y \neq 2) = 1 - \mathbb{P}(Y = 0) - \mathbb{P}(Y = 1) = 0.01$$

Da ferner X und Y stochastisch unabhängig sind, gilt

$$\mathbb{P}(Y = 2 \mid X = 0) = \mathbb{P}(Y = 2) = 0.01.$$

b) Wegen der Unabhängigkeit von X und Y gilt

$$\mathbb{P}(X = 0, Y = 0) = \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 0) = 0.9 \cdot 0.9 = 0.81.$$

c) $X + Y = 0$ gilt genau dann, wenn $X = 0$ und $Y = 0$, daher $\mathbb{P}(X + Y = 0) = \mathbb{P}(X = 0, Y = 0) \overset{b)}{=} 0.81$. Fallunterscheidung und Unabhängigkeit von X und Y ergibt

$$\begin{aligned} \mathbb{P}(X + Y = 1) &= \mathbb{P}(X = 0, Y = 1) + \mathbb{P}(X = 1, Y = 0) \\ &= \mathbb{P}(X = 0) \cdot \mathbb{P}(Y = 1) + \mathbb{P}(X = 1) \cdot \mathbb{P}(Y = 0) = 2 \cdot 0.9 \cdot 0.09 = 0.162 \end{aligned}$$

und damit

$$\mathbb{P}(X + Y \leq 1) = \mathbb{P}(X + Y = 0) + \mathbb{P}(X + Y = 1) = 0.972$$

d) Gesucht ist $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 1 - \mathbb{P}(X \leq 1) = 0.028$.

Aufgabe A3

Ein Transformatorkern habe die zufällige Dicke X (in mm) und eine Spule den zufällige Innendurchmesser Y. X und Y seien stochastisch unabhängige Zufallsvariable mit den Normalverteilungen $\mathcal{N}(22, 0.18)$ bzw. $\mathcal{N}(23, 0.07)$.

a) Welche Verteilung besitzt $Z := Y - X$?

b) Berechnen Sie die Kovarianz $C(Z, Y)$ und den Korrelationskoeffizienten $\rho(Z, Y)$ von Z und Y.

c) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(Z > 0)$, dass ein zufällig ausgewählter Kern in eine zufällig ausgewählte Spule passt.

d) Bestimmen Sie $t \in \mathbb{R}$ so, dass $\mathbb{P}(X > t) = 0.025$ gilt.

Lösung:

a) Es gilt $Z = Y + (-X)$. Wegen Satz 9.7 gilt $-X \sim \mathcal{N}(-22, 0.18)$, also nach der Faltungsformel

$$Z \sim \mathcal{N}(23 - 22, 0.07 + 0.18) = \mathcal{N}(1, 0.25)$$

(Die Varianzen dürfen hier **nicht** subtrahiert werden!)

b) Anwenden von Satz 12.23 ergibt

$$C(Z, Y) = C(Y + (-X), Y) = C(Y, Y) + C(-X, Y) = V(Y) + 0 = 0.07,$$

da $-X$ und Y stochastisch unabhängig sind. Damit ergibt sich der Korrelationskoeffizient zu

$$\rho(Z, Y) = \frac{C(Z, Y)}{\sqrt{V(Z) \cdot V(Y)}} = \frac{0.07}{\sqrt{0.25 \cdot 0.07}} = 0.52915$$

c) Wegen Satz 9.6 und (9.5) gilt

$$\begin{aligned}\mathbb{P}(Z > 0) = 1 - \mathbb{P}(Z \leq 0) = 1 - \Phi_{1,0.25}(0) = 1 - \Phi\left(\frac{0 - 1}{0.5}\right)& \\ = 1 - \Phi(-2) = \Phi(2) = 0.9772&\end{aligned}$$

d) Die Gleichung

$$0.025 = \mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t) = 1 - \Phi_{22,0.18}(t) = 1 - \Phi\left(\frac{t - 22}{\sqrt{0.18}}\right)$$

ist nach t aufzulösen, also

$$\Phi\left(\frac{t - 22}{\sqrt{0.18}}\right) = 1 - 0.025 = 0.975$$

Da $\Phi(1.96) = 0.975$ gilt, muss also

$$\frac{t - 22}{\sqrt{0.18}} = 1.96$$

gelten, also $t = 1.96 \cdot \sqrt{0.18} + 22 = 22.8316$

Aufgabe A4

Eine Zufallsvariable X habe die Verteilung $Exp(\lambda)$ mit dem Parameter $\lambda > 0$ und es sei $Y = \dfrac{X}{1+X}$ mit festem Parameter $\lambda > 0$.

a) Bestimmen Sie den Median von X für den Spezialfall $\lambda = \ln(2)$.

b) Berechnen Sie den Erwartungswert $\mathbb{E}\dfrac{Y^2}{(1-Y)^2}$.

Hinweis: $\dfrac{Y^2}{(1-Y)^2}$ ist eine einfache Funktion der Zufallsvariablen X.

c) Y besitzt die stetige Verteilungsfunktion

$$t \to F_Y(t) := \begin{cases} 0 & , t \leq 0 \\ 1 - e^{-\frac{\lambda t}{1-t}} & , 0 < t < 1 \\ 1 & , t \geq 1. \end{cases}$$

Bestimmen Sie die Dichte f_Y von Y.

Lösung:

a) Nach Beispiel 12.21 besitzt die Zufallsvariable X allgemein das p-Quantil

$$t_p = -\frac{1}{\lambda} \cdot \ln(1-p).$$

Für den Median ist $p = 1/2$. Speziell für $\lambda = \ln(2)$ ergibt sich

$$t_{1/2} = -\frac{1}{\ln(2)} \cdot \ln(1-1/2) = -\frac{1}{\ln(2)} \cdot \ln(1/2) = 1.$$

b) Es gilt $\dfrac{Y^2}{(1-Y)^2} = \dfrac{(\frac{X}{1+X})^2}{(1-\frac{X}{1+X})^2} = \dfrac{\frac{X^2}{(1+X)^2}}{\frac{1}{(1+X)^2}} = X^2$. Da X die Verteilung $Exp(\lambda)$ besitzt, folgt aus Beispiel 12.16 a) (mit $\alpha = 1$ und $\beta = \lambda$)

$$\mathbb{E}\frac{Y^2}{(1-Y)^2} = \mathbb{E}X^2 = \frac{2}{\lambda^2}.$$

Alternativ erhielte man das Resultat auch mit $\mathbb{E}X^2 = V(X) + (\mathbb{E}X)^2 = \dfrac{1}{\lambda^2} + \left(\dfrac{1}{\lambda}\right)^2$.

c) Nach der Aufgabenstellung ist die Verteilungsfunktion F_Y stetig. Aus der expliziten Darstellung folgt auch, dass F_Y bis auf eventuell die Stellen 0 und 1 auch stetig differenzierbar ist, so dass Satz 8.12 anwendbar ist. Da $F_Y(t)$ für $t \leq 0$ und auch für $t \geq 1$ konstant ist, folgt $f_Y(y) = 0$ für $y \leq 0$ und für $y \geq 1$. Für $0 < y < 1$ folgt schließlich mit der Kettenregel

$$f_Y(y) = F_Y'(y) = \lambda \cdot \left(\frac{y}{1-y}\right)' \cdot e^{-\lambda \cdot \frac{y}{1-y}} = \frac{\lambda}{(1-y)^2} \cdot e^{-\lambda \cdot \frac{y}{1-y}}.$$

Aufgabe A5

Ein Merkmal besitze die Weibull-Verteilung $W_{\vartheta,\beta}$. Es besitzt also die Dichte

$$f_\vartheta(x) := \begin{cases} \vartheta \cdot \beta \cdot x^{\beta-1} \cdot e^{-\vartheta \cdot x^\beta} & , \ x > 0 \\ 0 & , \ \text{sonst} \end{cases}$$

mit einem bekannten Parameter $\beta > 0$ und einem unbekannten Parameter $\vartheta > 0$. Beobachtet sei die feste Stichprobe $x = (x_1, \ldots, x_n)$ mit $x_i > 0$ für $i = 1, \ldots, n$.

a) Bestimmen Sie die Loglikelihood-Funktion zu x.

b) Berechnen Sie die Ableitung $M'_x(\vartheta)$.

c) Bestimmen Sie den Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ.

d) Bekannt ist, dass für eine Zufallsvariable Y mit Weibull-Verteilung $W_{\vartheta,\beta}$ die Zufallsvariable Y^β die Verteilung $Exp(\vartheta)$ besitzt (kein Beweis erforderlich).
Seien $X_1, \ldots, X_n$ die stochastisch unabhängig Stichprobenvariablen. Welche Verteilung besitzt $Z := \sum_{i=1}^n X_i^\beta$?

Lösung:

a) Wegen $\ln(f_\vartheta(x)) = \ln(\vartheta) + \ln(\beta) + (\beta-1) \cdot \ln(x) - \vartheta \cdot x^\beta$ gilt für die Loglikelihood-Funktion

$$\begin{aligned} M_x(\vartheta) = \sum_{i=1}^n \ln(f_\vartheta(x_i)) &= (\ln(\vartheta) + \ln(\beta) + (\beta - 1) \cdot \ln(x_1) - \vartheta \cdot x_1^\beta) \\ &\quad + \ldots + (\ln(\vartheta) + \ln(\beta) + (\beta - 1) \cdot \ln(x_n) - \vartheta \cdot x_n^\beta) \\ &= n \cdot \ln(\vartheta) + n \cdot \ln(\beta) + (\beta - 1) \cdot \sum_{i=1}^n \ln(x_i) - \vartheta \cdot \sum_{i=1}^n x_i^\beta \end{aligned}$$

b) Direkt aus a) ergibt sich Ableitung $M'_x(\vartheta)$ zu

$$M'_x(\vartheta) = \frac{n}{\vartheta} - \sum_{i=1}^n x_i^\beta$$

c) Zu bestimmen ist eine Maximumstelle von $\vartheta \to M_x(\vartheta)$. Wir gehen wie auf S. 164 im Skriptum vor und setzen die Ableitung $M'_x(\vartheta)$ gleich 0. Auflösen von $M'_x(\vartheta) = 0$ nach ϑ liefert die Lösung

$$\hat{\vartheta}(x) = n / \sum_{i=1}^n x_i^\beta$$

Wegen b) gilt

$$M''_x(\vartheta) = -\frac{n}{\vartheta^2} < 0$$

Da die zweite Ableitung von $M_x(\vartheta)$ also negativ ist, ist $\hat{\vartheta}(x)$ tatsächlich die gesuchte Maximumstelle und damit der Maximum-Likelihood-Schätzer.

d) Nach Voraussetzung besitzen die Stichprobenvariablen die Verteilung $W_{\vartheta,\beta}$, daher gilt $X_i^\beta \sim Exp(\vartheta) = \Gamma(1, \vartheta)$. Da die Zufallsvariablen X_i^β stochastisch unabhängig sind, gilt nach der Faltungsformel 11.16

$$Z = X_1^\beta + \ldots + X_n^\beta \sim \Gamma(\vartheta, n).$$

Aufgabe A6

Zwei Firmen erhalten Lieferungen gleichartiger Messinstrumente, von denen viele wegen Messfehlern noch nachjustiert werden müssen. Angenommen sei, dass sich die Messinstrumente nicht gegenseitig beeinflussen.

a) Von den $n = 18$ an die erste Firma gelieferten Messinstrumenten mussten 12 nachjustiert werden. Geben Sie ein zweiseitiges Konfidenzintervall zum Konfidenzniveau 0.95 für die Wahrscheinlichkeit p an, dass ein Messinstrument nachjustiert werden muss.

b) Die zweite Firma erhält eine Lieferung von $N = 1000$ Messinstrumenten eines anderen Typs. Von den gelieferten Instrumenten werden 160 überprüft; dabei stellt sich heraus, dass 120 der überprüften Instrumente nachjustiert werden müssen. Bestimmen Sie ein zweiseitiges Konfidenzintervall zum Konfidenzniveau 0.95 für den Anteil der Instrumenten in der gesamten Lieferung, die nachjustiert werden müssen.

Lösung:

a) Nach Voraussetzung kann wie in Beispiel 18.5 ein ideales Zufallsexperiment mit den zwei möglichen Ergebnissen „Es muss nachjustiert werden“ (Treffer) und „Es muss nicht nachjustiert werden“ (Niete) und der Trefferwahrscheinlichkeit $\vartheta := p$ angesehen werden. Nach diesem Beispiel und wegen 18.6 ist das gesuchte Konfidenzintervall $[l(x), L(x)]$, wobei die Konfidenzgrenzen $l(x)$ und $L(x)$ für $x = 12$ und $n - x = 6$ und $1 - \alpha = 0.95$ Tabelle A.4 entnommen werden. Dies ergibt das Konfidenzintervall

$$\mathcal{C}(x) = [l(x), L(x)] = [0.410\ ,\ 0.867]\ .$$

Es ist also $l(x) = 0.410$ und $L(x) = 0.867$.

b) Hier ist Beispiel 18.12 (Anteils-Schätzung bei endlicher Population) anwendbar. In der Lieferung sind $N = 1000$ Geräte, von denen r (r unbekannt) nachjustiert werden müssen. Es ist dann $T_n = 120/160 = 0.75$ der relative Anteil der Instrumente in der zufälligen Stichprobe vom Umfang $n = 160$, die nachjustiert werden müssen. Mit $\vartheta := r/N$ (= unbekannter Anteil der nachzujustierenden Instrumente), $h := u_{1-\alpha/2} = 1.9600$ und dem Endlichkeitskorrektur-Faktor $\gamma\ :=\ 1 - \frac{n-1}{N-1} = 0.8408$ bilden dann

$$\frac{T_n\ +\ \dfrac{h^2}{2n}\gamma\ \pm\ \dfrac{h}{\sqrt{n}} \cdot \sqrt{T_n(1 - T_n)\gamma + \dfrac{h^2}{4n}\gamma^2}}{1\ +\ \dfrac{h^2}{n}\gamma}$$

den oberen bzw. unteren Endpunkt eines approximativen $(1 - \alpha)$–Konfidenzintervalles für ϑ $(= \frac{r}{N})$.

Setzt man die Werte ein, so ergibt sich für den unbekannten Anteil r/N das Konfidenzintervall

$$\mathcal{C}'(x) = [l'(x), L'(x)] = [0.6839\ ,\ 0.8062]\ .$$

Es ist also $l'(x) = 0.6839$ und $L'(x) = 0.8062$.