Karlsruher Institut für Technologie

Institut für Stochastik
Dr. B. Ebner

Name:
Vorname:
Matr.-Nr.:

Klausur zum Fach

WAHRSCHEINLICHKEITSTHEORIE UND STATISTIK (STOCHASTIK)

für **Studierende der Informatik**

Datum: **12. Juli 2011**

Dauer: **90 Minuten**

Diese Klausur hat bestanden, wer mindestens 18 Punkte erreicht.

Aufgabe 1 (10 Punkte)

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	0.5	0.7	1.7	1.7	2.1	2.2	2.2	2.4	2.9	3.1	3.4	3.8
y_j	0.1	0.4	0.9	1.2	1.4	1.7	2.0	2.1	2.5	2.7	2.5	5.4

a) Berechnen Sie die Stichprobenmittel $\bar{x}$, $\bar{y}$, die Stichproben-Standardabweichungen s_x, s_y und den empirischen Korrelationskoeffizienten r_{xy}.
Hinweis:
$$\sum_{j=1}^{12} x_j = 26.7, \quad \sum_{j=1}^{12} x_j^2 = 70.39, \quad \sum_{j=1}^{12} y_j = 22.9, \quad \sum_{j=1}^{12} y_j^2 = 64.63, \quad \sum_{j=1}^{12} x_j \cdot y_j = 64.66.$$

b) Bestimmen Sie die zugehörige Regressionsgerade $y = a^* + b^* \cdot x$ von y auf x.

c) Berechnen Sie das 0.15-getrimmte Stichprobenmittel $\bar{y}_{0.15}$ von $(y_1, \ldots, y_{12})$.

d) Bestimmen Sie das Stichproben-0.3-Quantil $\tilde{y}_{0.3}$ von $(y_1, \ldots, y_{12})$.

Lösung:

a) Direkt aus den Daten ergibt sich gemäß Definition 1.8 und Paragraph 1.5 unter Ausnützung der Beziehung

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

$$\bar{x} = 2.225 \qquad s_x = 0.999$$
$$\bar{y} = 1.908 \qquad s_y = 1.379$$
$$r_{xy} = 0.904$$

b) Nach Paragraph 1.5 ist $b^* = r_{xy} \cdot \dfrac{s_y}{s_x}$ und $a^* = \bar{y} - b^* \cdot \bar{x}$, also

$$\begin{aligned} b^* &= 1.249 \\ a^* &= -0.870 \end{aligned}$$

und die Regressionsgerade $y = -0.870 + 1.249 \cdot x$.

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

$$y_{()} = (0.1, 0.4, 0.9, 1.2, 1.4, 1.7, 2.0, 2.1, 2.5, 2.5, 2.7, 5.4)$$

Mit $k = [12 \cdot 0.15] = 1$ ergibt sich

$$\bar{y}_{0.15} = \frac{1}{12 - 2 \cdot 1} \cdot (y_{(2)} + \ldots + y_{(11)}) = 1.74$$

d) Da $12 \cdot 0.3 = 3.6$ nicht ganzzahlig ist, ist mit $k = [3.6] = 3$

$$\tilde{y}_{0.3} = y_{(k+1)} = y_{(4)} = 1.2$$

Aufgabe 2 (10 Punkte)
Die Funktion f sei in Abhängigkeit des Parameters $c > -1$ gegeben durch

$$f(x) = \begin{cases} x^c, & \text{falls } 0 < x \leq 1, \\ 0, & \text{sonst.} \end{cases}$$

a) Zeigen Sie, dass die Funktion f nur für $c = 0$ eine Dichte einer Zufallsvariablen X ist.

b) Die Zufallsvariable X habe die Dichte f aus Teilaufgabe a). Bestimmen Sie die Verteilungsfunktion F_X der Zufallsvariablen X. Wie nennt man die Verteilung?

c) Betrachten Sie die Zufallsvariable $Y := 2 \cdot X - 1$. Geben Sie die Verteilung F_Y von Y an und bestimmen Sie den Erwartungswert $\mathbb{E}(Y)$.

d) Bestimmen Sie das zweite Moment $\mathbb{E}(Y^2)$ und die Varianz $V(Y)$ von Y.

e) Bestimmen Sie das erste Quartil der Verteilungsfunktion F_X.

Lösung:

a) Die Abbildung f ist für alle $c > -1$ und alle $x \in \mathbb{R}$ größer gleich 0. Weiter ist

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_0^1 x^c\, dx = \frac{1}{c+1} x^{c+1} \Big|_0^1 = \frac{1}{c+1} \stackrel{!}{=} 1$$

also muss $c = 0$ sein.

b) Die Verteilungsfunktion lautet für $t \in [0, 1]$

$$F_X(t) = \int_{-\infty}^{t} f(x)\, dx = \int_0^t 1\, dx = t$$

und $F_X(t) = 0$ für $t < 0$ und $F_x(t) = 1$ für $t > 1$. Es handelt sich also um die Gleichverteilung $\mathcal{U}(0, 1]$.

c) Die Zufallsvariable Y besitzt die Verteilung $\mathcal{U}(-1, 1)$. Es gilt

$$\mathbb{E}(Y) = 2\mathbb{E}(X) - 1 = 0.$$

d) Es gilt

$$\mathbb{E}(Y^2) = \int_{-1}^{1} x \cdot \frac{x+1}{2}\, dx = \frac{1}{6}x^3 + \frac{1}{4}x^2 \Big|_{-1}^{1} = \frac{1}{3}.$$

und damit

$$V(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \frac{1}{3}.$$

e) Das erste Quartil $u_{0.25}$ der Gleichverteilung $\mathcal{U}(0, 1]$ liegt bei $u_{0.25} = 0.25$.

Aufgabe 3 (10 Punkte)

Mit einer Maschine zur Herstellung von Plastikflaschen werden $n > 0$ grüne und $m > 0$ schwarze Flaschen produziert. Die Qualitätsprüfung erfolgt durch Laserabtastung. Eine Flasche wird (unabhängig von den anderen) aussortiert, wenn „zu große“Abweichungen von der eingestellten Norm festgestellt werden; dies geschehe mit Wahrscheinlichkeit $p = 0.05$. Sei X die zufällige Anzahl aussortierter grüner Flaschen und Y die zufällige Anzahl aussortierter schwarzer Flaschen.

a) Geben Sie die Wahrscheinlichkeitsverteilungen von X und Y an.

b) In einem Testlauf werden $n = 10$ grüne Flaschen hergestellt. Berechnen Sie den Erwartungswert $\mathbb{E}(X)$, die Varianz $V(X)$ und $\mathbb{P}(X > 1)$.

c) Die Zufallsvariablen X und Y seien stochastisch unabhängig. Welche Verteilung hat die Gesamtanzahl $Z := X+Y$ der aussortierten Flaschen, wenn $n = 10$ grüne und $m = 8$ schwarze Flaschen produziert werden? Berechnen Sie die Wahrscheinlichkeit, dass höchstens 2 Flaschen fehlerhaft sind.

d) Bestimmen Sie in der Situation von Teilaufgabe c) $\mathbb{P}(Z > 2|Y = 1)$.

e) Durch eine eingebaute Fehlerkorrektur gelte für die Fehlerwahrscheinlichkeit $p_n := \frac{\lambda}{n}$ mit einem $\lambda > 0$. Welche Verteilung hat die Zufallsvariable X für $n \to \infty$?

Lösung:

a) Es gilt (Treffer-Niete Experiment)

$$X \sim \mathrm{Bin}(n, p) \quad \text{und} \quad Y \sim \mathrm{Bin}(m, p).$$

b) Mit $n = 10$ gilt

$$\mathbb{E}(X) = n \cdot p = 0.5 \text{ und } V(X) = n \cdot p \cdot (1 - p) = 0.475.$$

Weiter ist

$$\mathbb{P}(X > 1) = 1 - \mathbb{P}(X \leq 1) = 1 - \left[\binom{10}{0} 0.05^0 (0.95)^{10} + \binom{10}{1} (0.05)(0.95)^9\right] \approx 0.086.$$

c) Mit der Faltungsformel für die Binomialverteilung hat Z die Verteilung $\mathrm{Bin}(18, p)$.

$$\mathbb{P}(Z \leq 2) = \sum_{k=0}^{2} \binom{18}{k} 0.05^k 0.95^{18-k} \approx 0.942.$$

d) Wegen der stochastischen Unabhängigkeit von X und Y gilt

$$\begin{aligned}\mathbb{P}(Z > 2|Y = 1) &= \frac{\mathbb{P}(Z > 2, Y = 1)}{\mathbb{P}(Y = 1)} \\ &= \frac{\mathbb{P}(X + Y > 2, Y = 1)}{\mathbb{P}(Y = 1)} \\ &= \frac{\mathbb{P}(X > 1, Y = 1)}{\mathbb{P}(Y = 1)} \\ &= \mathbb{P}(X > 1) \approx 0.086.\end{aligned}$$

e) Die Zufallsvariable X hat in diesem Fall eine Poisson-Verteilung mit Parameter λ.

Aufgabe 4 (10 Punkte)

Es seien X eine Zufallsvariable mit Werten in $\{0,1\}$ und Y eine Zufallsvariable mit Werten in $\{0,1,2\}$. Die folgende Tabelle gibt die gemeinsame Verteilung $\mathbb{P}(X=i, Y=j)$ des Zufallsvektors (X,Y) für die Werte $i=1,2$ und $j=0,1,2$ an.

	$j=0$	$j=1$	$j=2$
$i=0$	$\frac{1}{3}$	0	$\frac{1}{6}$
$i=1$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

a) Berechnen Sie die Randverteilung von Y, d.h. $\mathbb{P}(Y=0), \mathbb{P}(Y=1)$ und $\mathbb{P}(Y=c)$, und den Erwartungswert $\mathbb{E}Y$.

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

c) Berechnen Sie die bedingte Wahrscheinlichkeit $\mathbb{P}(X=1|Y>0)$.

d) Es sei $Z := X \cdot Y$. Berechnen Sie den Erwartungswert $\mathbb{E}Z$, das zweite Moment $\mathbb{E}Z^2$ und die Varianz $V(Z)$.

e) Sind X und Y unabhängig? Begründen Sie Ihre Antwort!

Lösung:

a) Die Randverteilung von Y berechnet sich durch

$$\begin{aligned}
\mathbb{P}(Y=0) &= \mathbb{P}(X=0, Y=0) + \mathbb{P}(X=1, Y=0) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}, \\
\mathbb{P}(Y=1) &= \mathbb{P}(X=0, Y=1) + \mathbb{P}(X=1, Y=1) = 0 + \frac{1}{6} = \frac{1}{6}, \\
\mathbb{P}(Y=2) &= \mathbb{P}(X=0, Y=2) + \mathbb{P}(X=1, Y=2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.
\end{aligned}$$

Der Erwartungswert ist gegeben durch

$$\mathbb{E}(Y) = 1 \cdot \mathbb{P}(Y=1) + 2 \cdot \mathbb{P}(Y=2) = \frac{1}{6} + \frac{2}{3} = \frac{5}{6}.$$

b) Es gilt

$$\mathbb{P}(X=0, Y>0) = \mathbb{P}(X=0, Y=1) + \mathbb{P}(X=0, Y=2) = \frac{1}{6}.$$

c) Weiter gilt

$$\mathbb{P}(X=1|Y>0) = \frac{\mathbb{P}(X=1, Y>0)}{\mathbb{P}(Y>0)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

d) Es gilt

$$\mathbb{E}(Z) = \mathbb{E}(X \cdot Y) = \sum_{(i,j)} i \cdot j P(X = i, Y = j) = \frac{1}{6} + \frac{2}{6} = \frac{1}{2}.$$

Weiter ist

$$\mathbb{E}(Z^2) = \mathbb{E}(X \cdot Y) = \sum_{(i,j)} (i \cdot j)^2 P(X = i, Y = j) = \frac{1}{6} + \frac{4}{6} = \frac{5}{6}.$$

und damit gilt für die Varianz von Z

$$V(Z) = \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2 = \frac{5}{6} - \frac{1}{4} = \frac{7}{12}.$$

e) Die Zufallsvariablen X und Y sind nicht stochastisch unabhängig, da

$$\mathbb{P}(X = 0, Y = 0) = \frac{1}{3} \neq \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathbb{P}(Y = 0) \cdot \mathbb{P}(X = 0).$$

Aufgabe 5 (10 Punkte)
Es soll der unbekannte Parameter $\vartheta > 0$ für die Verteilung mit der Dichte

$$f_\vartheta(x) = \begin{cases} \frac{1}{\vartheta^2} \cdot x \cdot \exp\left(-\frac{x}{\vartheta}\right), & 0 \leq x < \infty, \\ 0, & \text{sonst} \end{cases}$$

bestimmt werden.
Hinweis: Sie können ohne Beweis verwenden: Hat die Zufallsvariable X die Dichte f_ϑ, so gilt $\mathbb{E}_\vartheta(X) = 2\vartheta$.

a) Geben Sie die zur Stichprobe $x = (x_1, \ldots, x_n)$ gehörende Likelihood-Funktion $L_x(\vartheta)$ an und berechnen Sie die Loglikelihood-Funktion $M_x(\vartheta)$.

b) Geben Sie den Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ an.

c) Ist der Schätzer erwartungstreu für $\gamma(\vartheta) = \vartheta$? Ist der Schätzer asymptotisch erwartungstreu für $\gamma(\vartheta) = \vartheta$?

d) Geben Sie den Momenten-Schätzer $\tilde{\vartheta}(x)$ für ϑ an.

Lösung:

a) Die zur Stichprobe $x = (x_1, \ldots, x_n)$ gehörende Likelihood-Funktion $L_x(\vartheta)$ lautet

$$L_x(\vartheta) = \prod_{j=1}^n f_\vartheta(x_j) = \prod_{j=1}^n \frac{1}{\vartheta^2} \cdot x_j \cdot \exp\left(-\frac{x_j}{\vartheta}\right) = \frac{1}{\vartheta^{2n}} \cdot \exp\left(-\frac{1}{\vartheta}\sum_{j=1}^n x_j\right) \prod_{j=1}^n x_j.$$

Die Loglikelihood-Funktion $M_x(\vartheta)$ lautet entsprechend

$$M_x(\vartheta) = \log L_x(\vartheta) = -2n \log \vartheta - \frac{1}{\vartheta}\sum_{j=1}^n x_j + \sum_{j=1}^n \log(x_j).$$

b) Differenzieren von $M_x(\vartheta)$ nach ϑ liefert

$$M_x'(\vartheta) = -2n\frac{1}{\vartheta} + \frac{1}{\vartheta^2}\sum_{j=1}^n x_j = \frac{1}{\vartheta^2}\left(\sum_{j=1}^n x_j - 2n\vartheta\right) \stackrel{!}{=} 0.$$

also ist $\hat{\vartheta}(x) = \frac{1}{2n}\sum_{j=1}^n x_j$ ein stationärer Punkt von M_x und es ist wegen dem Vorzeichenwechsel von $+$ nach $-$ auch ein Maximum. Also ist $\hat{\vartheta}(x)$ der gesuchte Maximum-Likelihood-Schätzer.

c) Es gilt mit dem Hinweis

$$\mathbb{E}(\hat{\vartheta}(X_1, \ldots, X_n)) = \mathbb{E}\left(\frac{1}{2n}\sum_{j=1}^n X_j\right) = \frac{1}{2n}\sum_{j=1}^n \mathbb{E}\left(X_j\right) = \frac{1}{2}\mathbb{E}(X_1) = \frac{1}{2}2\vartheta = \vartheta.$$

Also ist $\hat{\vartheta}(x)$ ein erwartungstreuer Schätzer für ϑ und demzufolge auch asymptotisch erwartungstreu.

d) Der Momentenschätzer berechnet sich mit Hilfe des Hinweises durch

$$\frac{1}{n}\sum_{j=1}^{n} x_j = \mathbb{E}(X_1) = 2\vartheta.$$

Auflösen nach ϑ liefert

$$\tilde{\vartheta}(x) = \frac{1}{2n}\sum_{j=1}^{n} x_j = \hat{\vartheta}(x).$$