

Name:

Vorname:

Matr.-Nr.:

Klausur zum Fach
GRUNDLAGEN DER WAHRSCHEINLICHKEITSTHEORIE
UND STATISTIK
für Studierende der INFORMATIK

Datum: 08. Februar 2012

Dauer: 90 Minuten

Achtung:

Bei dieser Klausur werden **nur** diejenigen Ergebnisse gewertet, die in die vorgesehenen Kästchen auf dem extra ausgegebenen

Lösungsblatt

eingetragen sind! Die Herleitung wird **nicht** bewertet, es sei denn, dies wird auf dem Lösungsblatt explizit gefordert!

Die Aufgabenblätter werden nicht abgegeben und korrigiert!

Diese Klausur hat bestanden, wer mindestens 18 Punkte von 50 möglichen Punkten erreicht.

Aufgabe 1 (10 Punkte)

Gegeben sei eine Urliste mit den Paaren $(x_1, y_1), \dots, (x_8, y_8)$

j	1	2	3	4	5	6	7	8
x_j	1.5	1.7	2.7	2.7	3.1	3.2	3.2	3.4
y_j	2.6	2.9	3.4	3.7	3.9	4.2	4.5	4.6

- a) Berechnen Sie die Stichprobenmittel \bar{x} , \bar{y} , die Stichproben-Standardabweichungen s_x , s_y und den empirischen Korrelationskoeffizienten r_{xy} .

Hinweis:

$$\sum_{j=1}^8 x_j = 21.5, \quad \sum_{j=1}^8 x_j^2 = 61.37, \quad \sum_{j=1}^8 y_j = 29.8, \quad \sum_{j=1}^8 y_j^2 = 114.68, \quad \sum_{j=1}^8 x_j \cdot y_j = 83.57.$$

- b) Bestimmen Sie die zugehörige Regressionsgerade $y = a^* + b^* \cdot x$ von y auf x .
- c) Berechnen Sie das 0.2-getrimmte Stichprobenmittel $\bar{y}_{0.2}$ von (y_1, \dots, y_8) .
- d) Bestimmen Sie das Stichproben-0.3-Quantil $\tilde{y}_{0.3}$ von (y_1, \dots, y_8) .

Lösung:

- a) Direkt aus den Daten ergeben sich aufgrund der Abschnitte 1.4 und 1.5 im Skript mit Hilfe der Beziehung

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

die Ergebnisse

$$\begin{aligned} \bar{x} &= 2.6875 & s_x &= \left(\frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) \right)^{\frac{1}{2}} = 0.716 \\ \bar{y} &= 3.725 & s_y &= 0.7246 \\ r_{xy} &= 0.9589. \end{aligned}$$

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

$$\begin{aligned} b^* &= 0.9704 \\ a^* &= 1.117 \end{aligned}$$

und die Regressionsgerade $y = 1.117 + 0.9704 \cdot x$.

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

$$y_{(j)} = (2.6, 2.9, 3.4, 3.7, 3.9, 4.2, 4.5, 4.6).$$

Mit $k = [8 \cdot 0.2] = [1.6] = 1$ ergibt sich

$$\bar{y}_{0.2} = \frac{1}{8 - 2 \cdot 1} \cdot (y_{(2)} + \dots + y_{(7)}) = 3.77.$$

d) Da $8 \cdot 0.3 = 2.4$ nicht ganzzahlig ist, ist mit $k = [2.4] = 2$

$$\tilde{y}_{0.3} = y_{(k+1)} = y_{(3)} = 3.4.$$

e) Da $0.25 \cdot 8 = 2$ und $0.75 \cdot 8 = 6$ beide ganzzahlig sind, ergibt sich mit $k_1 = 2$ und $k_2 = 6$

$$\tilde{y}_{0.25} = \frac{y_{(k_1)} + y_{(k_1+1)}}{2} = \frac{y_{(2)} + y_{(3)}}{2} = 3.15$$

$$\tilde{y}_{0.75} = \frac{y_{(k_2)} + y_{(k_2+1)}}{2} = \frac{y_{(6)} + y_{(7)}}{2} = 4.35$$

und damit der Quartilsabstand zu $\tilde{y}_{0.75} - \tilde{y}_{0.25} = 1.2$.

Aufgabe 2 (10 Punkte)

Es seien X eine Zufallsvariable mit Werten in $\{-1, 1, 2\}$ und

$$\mathbb{P}(X = 1) = \frac{2}{3}, \quad \mathbb{P}(X = -1) = \mathbb{P}(X = 2) = \frac{1}{6}$$

sowie Y eine Zufallsvariable mit Werten in $\{0, 1\}$ und den bedingten Wahrscheinlichkeiten

$$\mathbb{P}(Y = 0 \mid X = -1) = \frac{1}{2}, \quad \mathbb{P}(Y = 0 \mid X = 1) = \frac{1}{4}, \quad \mathbb{P}(Y = 1 \mid X = 2) = \frac{3}{4}.$$

- Hinweis:** beachten Sie, dass $\mathbb{P}(Y = 0 \mid X = -1) + \mathbb{P}(Y = 1 \mid X = -1) = 1$ gilt. Berechnen Sie $\mathbb{P}(Y = 1 \mid X = -1)$, $\mathbb{P}(Y = 1 \mid X = 1)$ und $\mathbb{P}(Y = 0 \mid X = 2)$.
- Berechnen Sie die gemeinsame Verteilung von X und Y sowie die Verteilung von Y , d.h. $\mathbb{P}(X = x, Y = y)$ und $\mathbb{P}(Y = y)$ für alle $x \in \{-1, 1, 2\}$ und $y \in \{0, 1\}$.
- Bestimmen Sie die Erwartungswerte $\mathbb{E}(X)$, $\mathbb{E}(Y)$, $\mathbb{E}(X + Y)$.
- Berechnen Sie $\mathbb{E}(X \cdot Y)$ und die Varianz $V(X)$ von X .
- Sind X und Y stochastisch unabhängig? Begründen Sie Ihre Antwort.

Lösung:

- Nach Skript/Vorlesung ist $\mathbb{P}(\cdot \mid X = -1)$ eine Wahrscheinlichkeitsverteilung, also gilt

$$\mathbb{P}(Y = 0 \mid X = -1) + \mathbb{P}(Y = 1 \mid X = -1) = \mathbb{P}(Y \in \{0, 1\} \mid X = -1) = 1.$$

(Nachweis war nicht gefordert!) Wir erhalten hiermit

$$\begin{aligned} \mathbb{P}(Y = 1 \mid X = -1) &= 1 - \mathbb{P}(Y = 0 \mid X = -1) = 1 - \frac{1}{2} = \frac{1}{2}, \\ \mathbb{P}(Y = 1 \mid X = 1) &= 1 - \mathbb{P}(Y = 0 \mid X = 1) = 1 - \frac{1}{4} = \frac{3}{4}, \\ \mathbb{P}(Y = 0 \mid X = 2) &= 1 - \mathbb{P}(Y = 1 \mid X = 2) = 1 - \frac{3}{4} = \frac{1}{4}. \end{aligned}$$

- Wir erhalten

$$\mathbb{P}(Y = 0, X = -1) = \mathbb{P}(Y = 0 \mid X = -1)\mathbb{P}(X = -1) = \frac{1}{2}\mathbb{P}(X = -1) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}$$

und mit analoger Rechnung

$$\begin{aligned} \mathbb{P}(Y = 0, X = 1) &= \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{6}, & \mathbb{P}(Y = 0, X = 2) &= \frac{1}{4} \cdot \frac{1}{6} = \frac{1}{24}, \\ \mathbb{P}(Y = 1, X = -1) &= \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12}, & \mathbb{P}(Y = 1, X = 1) &= \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2}, \\ \mathbb{P}(Y = 1, X = 2) &= \frac{3}{4} \cdot \frac{1}{6} = \frac{1}{8}. \end{aligned}$$

Hieraus folgt weiter

$$\mathbb{P}(Y = 0) = \frac{1}{12} + \frac{1}{6} + \frac{1}{24} = \frac{7}{24},$$

$$\mathbb{P}(Y = 1) = \frac{1}{12} + \frac{1}{2} + \frac{1}{8} = \frac{17}{24}.$$

c) Es gilt

$$\mathbb{E}(X) = 1 \cdot \frac{2}{3} + (-1) \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} = \frac{5}{6},$$

$$\mathbb{E}(Y) = 1 \cdot \frac{17}{24} = \frac{17}{24},$$

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = \frac{37}{24}.$$

d) Zunächst ist

$$\mathbb{E}(X \cdot Y) = (-1) \frac{1}{12} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{8} = \frac{2}{3}.$$

Ferner gilt

$$\mathbb{E}(X^2) = \frac{2}{3} + \frac{1}{6} + 4 \cdot \frac{1}{6} = \frac{3}{2}$$

und daher

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{3}{2} - (5/6)^2 = \frac{29}{36}.$$

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

$$\mathbb{E}(X \cdot Y) = \frac{2}{3} \neq \frac{5}{6} \cdot \frac{17}{24} = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Aufgabe 3 (10 Punkte)

An einer E-Mail-Adresse kommen erwünschte E-Mails und Spam-Mails an. Jede ankommende E-Mail sei unabhängig von den vorangehenden mit der Wahrscheinlichkeit $p \in (0, 1)$ **erwünscht**. Sei S_n die zufällige Anzahl von **nicht erwünschten** Spam-Mails bei n eingegangenen E-Mails.

- a) Welche Verteilung hat die Zufallsvariable S_n ?
Berechnen Sie $\mathbb{P}(S_n \geq 2)$ für $p = 0.7$ und $n = 10$.
- b) Berechnen Sie den Erwartungswert $\mathbb{E}(S_n)$ und die Varianz $V(S_n)$ für $p = 0.7$ und $n = 10$.
- c) Schätzen Sie für $p = 0.7 = 1 - q$, $n = 500$ und $\varepsilon = 0.05$ die Wahrscheinlichkeit

$$\mathbb{P}\left(\left|\frac{1}{n} \cdot S_n - q\right| \geq \varepsilon\right)$$

nach oben mit Hilfe der Tschebyschev-Ungleichung ab.

- d) Sei nun $p \in (0, 1)$ wieder beliebig. Ferner bezeichne T_k die zufällige Anzahl an Spam-Mails bis zur k -ten erwünschten E-Mail, $k = 1, 2$.
Welche Verteilungen haben die Zufallsvariablen T_1 beziehungsweise T_2 ?
Geben Sie $\mathbb{E}(T_1)$ und $\mathbb{E}(T_2)$ an.
- e) Berechnen Sie die Wahrscheinlichkeit $\mathbb{P}(T_1 \geq 2)$ für allgemeines p sowie für $p = 0.7$.

Lösung:

- a) Wegen der vorausgesetzten Unabhängigkeit und der festen Wahrscheinlichkeit $q = 1 - p$ einer nicht erwünschten Spam-Mail (“Treffer”) ist S_n binomialverteilt mit $S_n \sim \text{Bin}(n, q)$, also

$$\mathbb{P}(S_n = k) = \binom{n}{k} q^k p^{n-k}, \quad k = 0, \dots, n.$$

Ferner ist

$$\begin{aligned} \mathbb{P}(S_n \geq 2) &= 1 - \mathbb{P}(S_n = 0) - \mathbb{P}(S_n = 1) = 1 - q^n - \binom{n}{1} q^1 p^{n-1} \\ &= 1 - 0.7^{10} - 10 \cdot 0.3 \cdot 0.7^9 \approx 0.8507. \end{aligned}$$

- b) Es gilt $\mathbb{E}(S_n) = n \cdot q = 10 \cdot 0.3 = 3$ und $V(S_n) = n \cdot q \cdot p = 10 \cdot 0.3 \cdot 0.7 = 2.1$.
- c) Die Tschebyschev-Ungleichung (Skript!) ergibt

$$\mathbb{P}\left(\left|\frac{1}{n} \cdot S_n - q\right| \geq \varepsilon\right) \leq \frac{q \cdot p}{n \cdot \varepsilon^2} = \frac{0.3 \cdot 0.7}{500 \cdot 0.05^2} = \frac{21}{125} \approx 0.168.$$

- d) Es gilt

$$\begin{aligned} \mathbb{P}(T_1 = j) &= \mathbb{P}(\text{zunächst } j \text{ Spam-Mails, dann eine erwünschte E-Mail}) \\ &= q^j \cdot p, \end{aligned}$$

also folgt T_1 einer $Nb(1, p)$ -Verteilung (negative Binomialverteilung), welche auch als geometrische Verteilung $G(p)$ bezeichnet wird. Ferner ist

$$\begin{aligned}\mathbb{P}(T_2 = j) &= \mathbb{P}(\text{die ersten } j + 1 \text{ E-Mails enthalten genau } j \text{ Spam-Mails und eine} \\ &\quad \text{erwünschte E-Mail, dann folgt die zweite erwünschte E-Mail}) \\ &= \binom{j+1}{1} q^j \cdot p \cdot p = (j+1)p^2 q^j,\end{aligned}$$

also folgt T_2 einer $Nb(2, p)$ -Verteilung (negative Binomialverteilung). Nach Skript gilt daher

$$\mathbb{E}(T_1) = \frac{1-p}{p} \quad \text{und} \quad \mathbb{E}(T_2) = 2 \frac{1-p}{p}.$$

e) Es gilt

$$\begin{aligned}\mathbb{P}(T_1 \geq 2) &= 1 - \mathbb{P}(T_1 = 0) - \mathbb{P}(T_1 = 1) = 1 - p - qp = q - qp \\ &= q(1-p) = q^2 = 0.3^2 = 0.09.\end{aligned}$$

Aufgabe 4 (10 Punkte)

Die Zufallsvariable X besitze die Verteilung $\mathcal{N}(2, 1)$. Weiter sei $Y := 1 - 2X$.

- Berechnen Sie den Erwartungswert $\mathbb{E}(Y)$ und die Varianz $V(Y)$.
- Welche Verteilung hat Y ?
- Drücken Sie die Wahrscheinlichkeit $\mathbb{P}(-1 \leq Y \leq 3)$ mit Hilfe der Verteilungsfunktion Φ der Standardnormalverteilung $\mathcal{N}(0, 1)$ aus.
- Bestimmen Sie die Kovarianz $C(X, Y)$ und den Korrelationskoeffizienten $\rho(X, Y)$.
- Sei U eine auf $[0, 1]$ gleichverteilte Zufallsvariable, das heißt $U \sim \mathcal{U}(0, 1)$. Welche Verteilung hat die Zufallsvariable $2 \cdot \Phi^{-1}(U) + 1$? Begründen Sie Ihre Antwort.

Lösung: Wir verwenden die aus der Vorlesung bekannten Eigenschaften der Normalverteilung aus Abschnitt 9.2.

- $\mathbb{E}(Y) = \mathbb{E}(1 - 2X) = 1 - 2\mathbb{E}(X) = 1 - 2 \cdot 2 = 1 - 4 = -3$ und $V(Y) = V(1 - 2X) = V(-2X) = 4V(X) = 4 \cdot 1 = 4$.
- Aus Satz 9.7 des Skriptums folgt $Y \sim \mathcal{N}(-3, 4)$. Bemerkung: Man kann a) und b) auch zugleich mit Satz 9.7 beantworten, wenn man die Maßzahlen (Erwartungswert und Varianz) der Normalverteilung aus dem Skriptum verwendet.
- Für eine Zufallsvariable $N \sim \mathcal{N}(0, 1)$ gilt $-3 + 2N \sim \mathcal{N}(-3, 4)$ und daher folgt zusammen mit b)

$$\begin{aligned}\mathbb{P}(-1 \leq Y \leq 3) &= \mathbb{P}(-1 \leq -3 + 2N \leq 3) = \mathbb{P}(1 \leq N \leq 3) \\ &= \Phi(3) - \Phi(1) \quad [= 0.9987 - 0.8413 = 0.1574].\end{aligned}$$

d)

$$\begin{aligned}C(X, Y) &= C(X, 1 - 2X) = C(X, -2X) = -2C(X, X) = -2V(X) = -2, \\ \rho(X, Y) &= C(X, Y) / \sqrt{V(X)V(Y)} = (-2) / \sqrt{1 \cdot 4} = -1.\end{aligned}$$

- Nach Skript (Kapitel 15) gilt zunächst $\Phi^{-1}(U) \sim \mathcal{N}(0, 1)$. Wiederum nach Skript (Satz 9.7) folgt daher $2\Phi^{-1}(U) + 1 \sim \mathcal{N}(1, 4)$.

Aufgabe 5 (10 Punkte)

Es soll der unbekannte Parameter $\vartheta > 0$ für die Verteilung mit der Dichte

$$f_{\vartheta}(t) = \begin{cases} \frac{4t^3}{\vartheta} \cdot \exp\left(-\frac{1}{\vartheta} \cdot t^4\right), & t > 0, \\ 0, & \text{sonst} \end{cases}$$

bestimmt werden.

- Geben Sie die zur Stichprobe $x = (x_1, \dots, x_n)$ gehörende Likelihood-Funktion $L_x(\vartheta)$ und die Loglikelihood-Funktion $M_x(\vartheta)$ an.
- Berechnen Sie die Ableitung $M'_x(\vartheta)$.
- Bestimmen Sie einen Maximum-Likelihood-Schätzer $\hat{\vartheta}(x)$ für ϑ zur Stichprobe x .
- Bestimmen Sie zunächst die Verteilungsfunktion von X_1 , wenn X_1 die Dichte f_{ϑ} hat. Zeigen Sie dann, dass X_1^4 einer $\text{Exp}(\frac{1}{\vartheta})$ -Verteilung folgt.
- Ist $\hat{\vartheta}$ ein erwartungstreuer Schätzer für ϑ ?

Lösung: Für die Lösung kann man annehmen, dass $x_j > 0$ für $j = 1, \dots, n$. Andernfalls wäre $L_x(\vartheta) = 0$ und $M_x(\vartheta) = -\infty$ für alle $\vartheta > 0$.

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

$$L_x(\vartheta) = \prod_{j=1}^n f_{\vartheta}(x_j) = \prod_{j=1}^n \left(\frac{1}{\vartheta} \cdot (4x_j^3) \cdot \exp\left(-\frac{x_j^4}{\vartheta}\right) \right) = \frac{1}{\vartheta^n} \cdot \exp\left(-\frac{1}{\vartheta} \sum_{j=1}^n x_j^4\right) \prod_{j=1}^n (4x_j^3).$$

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

$$M_x(\vartheta) = \log L_x(\vartheta) = -n \log \vartheta - \frac{1}{\vartheta} \sum_{j=1}^n x_j^4 + \sum_{j=1}^n \log(4x_j^3).$$

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

$$M'_x(\vartheta) = -\frac{n}{\vartheta} + \frac{1}{\vartheta^2} \sum_{j=1}^n x_j^4 = -\frac{n}{\vartheta^2} \left(\vartheta - \frac{1}{n} \sum_{j=1}^n x_j^4 \right).$$

- Die Bedingung $M'_x(\vartheta) = 0$ führt auf $\vartheta = \frac{1}{n} \sum_{j=1}^n x_j^4$. Dann ist $\hat{\vartheta}(x) = \frac{1}{n} \sum_{j=1}^n x_j^4$ ein Maximum-Likelihood-Schätzer für ϑ , da $M'_x(\cdot)$ in der einzigen Nullstelle $\hat{\vartheta}(x)$ einen Vorzeichenwechsel von $+$ nach $-$ hat.

d) Für $z \geq 0$ ist

$$F_{X_1}(z) = \int_0^z \frac{4}{\vartheta} t^3 e^{-t^4/\vartheta} dt = [-e^{-t^4/\vartheta}]_0^z = 1 - e^{-z^4/\vartheta^4}$$

und folglich

$$F_{X_1^4}(z) = \mathbb{P}(X_1^4 \leq z) = \mathbb{P}(X_1 \leq z^{1/4}) = 1 - e^{-z/\vartheta}.$$

Beide Verteilungsfunktionen sind offensichtlich Null für $z < 0$. Insbesondere gilt daher, dass $X_1^4 \sim \text{Exp}(1/\vartheta)$.

e) Der Schätzer ist erwartungstreu, da aufgrund des vorangehenden Aufgabenteils gilt:

$$\mathbb{E}_\vartheta(\hat{\vartheta}(x)) = \frac{1}{n} \sum_{j=1}^n \mathbb{E}(X_j^4) = \mathbb{E}(X_1^4) = (1/\vartheta)^{-1} = \vartheta.$$