TU Berlin, Institut für Mathematik
SoSe 16

Prof. Dr. Noemi Kurt
Dr. Mathias Rafler

Stochastik für Informatiker (alt), (6LP)

Klausur

30. Juli 2016

Name ______________________ Matrikelnummer ______________

Vorname ______________________ Studiengang ______________

Informationen

Füllen Sie bitte zuerst das Deckblatt vollständig und leserlich aus. Vergewissern Sie sich, dass das Aufgabenblatt vollständig ist.

Schreiben Sie auf *jedes* von Ihnen benutzte Papier *sofort* Ihren Namen und Ihr Matrikelnummer.

Bei der Klausur sind 50 Punkte erreichbar. Mit 25 Punkten ist die Klausur bestanden. Als Hilfsmittel darf, wie angekündigt, ein beidseitig handbeschriebenes DIN-A4-Blatt sowie ein nicht programmierbarer Taschenrechner benutzt werden. Weitere Hilfsmittel sind nicht zugelassen.

Geben Sie stets den *vollständigen* Rechenweg an. Begründen Sie Ihre Schritte falls notwendig.

Die Bearbeitungszeit beträgt 90 Minuten.

1	2	3	4	5	Σ

Aufgabe 1 *10 Punkte*

Mithilfe eines neuen Bluttests wird auf eine Pollenallergie getestet. Von dem Test ist bekannt, dass eine Person mit Allergie mit einer Wahrscheinlichkeit von 95% als solche erkannt wird, und eine nicht betroffene Person mit einer Wahrscheinlichkeit von 90% als nicht betroffen erkannt wird. Aus Erfahrung ist bekannt, dass 10% der Bevölkerung die Pollenallergie haben. Bezeichne mit A das Ereignis, dass eine zufällig ausgewählte Person diese Allergie hat und B das Ereignis, dass der Bluttest diese Allergie anzeigt.

(a) Berechnen Sie

(i) $\mathbb{P}(B|A^c)$ (ii) $\mathbb{P}(B^c \cap A^c)$ (iii) $\mathbb{P}(B^c)$ (iv) $\mathbb{P}(A|B)$

(b) Die folgenden Frage können ohne Begründung gelöst werden:

(i) Welche Paare von Ereignissen sind unabhängig?
1. B und A, 2. B^c und A^c, 3. B und A^c, 4. keines.

(ii) Falls der Test so verbessert werden kann, dass nun $\mathbb{P}(B|A) = 1$ gilt, bedeutet das: $\mathbb{P}(A|B)$ ist dann
1. fast 1, 2. leicht größer als der in (a) berechnete Wert, 3. kleiner als der in (a) berechnete Wert, 4. komplett was anderes.

(iii) Ein Labor führt den Test an den Blutproben verschiedener Personen nacheinander durch. Die Anzahl der Tests bis das erste Mal positiv angezeigt wird, hat die folgende Verteilung:
1. binomial, 2. geometrisch, 3. Poisson, 4. uniform (diskret).

(iv) Ein Labor führt täglich den Test an den Blutproben von 600 verschiedenen Personen nacheinander durch. Die Anzahl der Tests, die positiv angezeigen, hat die folgende Verteilung:
1. binomial, 2. geometrisch, 3. Poisson, 4. uniform (diskret).

Aufgabe 2 *10 Punkte*

Es sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum $S = \{1, \dots, 4\}$ und Übergangsmatrix

$$P = \begin{bmatrix} 1/4 & 0 & 0 & 3/4 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/8 & 7/8 & 0 \\ 3/4 & 0 & 0 & 1/4 \end{bmatrix}$$

(a) Skizzieren Sie den Übergangsgraphen zu dieser Matrix. Ist die Kette irreduzibel? Begründen Sie kurz.

(b) Bestimmen Sie alle ggf. vorhandenen invarianten Verteilungen dieser Kette.

(c) Geben Sie $\lim_{n\to\infty} \mathbb{P}(X_n = 3 \,|\, X_0 = 1)$ und $\lim_{n\to\infty} \mathbb{P}(X_n = 4 \,|\, X_0 = 1)$ an (mit kurzer Begründung).

Aufgabe 3 *10 Punkte*

Gegeben sei eine Zufallsvariable T mit *Verteilungsfunktion*

$$F_T(x) = \begin{cases} 1 - \left(\frac{5}{x}\right)^2, & \text{falls } x \geq a \\ 0 & \text{falls } x < a, \end{cases}$$

wobei $a > 0$ ein Parameter ist.

(a) Bestimmen Sie a so, dass F_T eine stetige Verteilungsfunktion ist und berechnen Sie die Dichte von T.

(b) Berechnen Sie die folgenden Wahrscheinlichkeiten

(i) $\mathbb{P}(T > 11)$ (ii) $\mathbb{P}(3 < T \leq 7)$

(c) Berechnen Sie $\mathbb{E}[T]$.

Aufgabe 4 *10 Punkte*

Sei $\theta > 0$ ein unbekannter Parameter, und seien $X_1, X_2, \ldots, X_n$ unabhängige Zufallsvariablen mit Dichte

$$f_\theta(x) = \begin{cases} \sqrt{\frac{2}{\theta\pi}}\, \mathrm{e}^{-\frac{x^2}{2\theta}} & \text{falls } x > 0 \\ 0 & \text{sonst.} \end{cases}$$

(a) Berechnen Sie die Likelihoodfunktion für Beobachtungen $x_1, \ldots, x_n$ und vereinfachen Sie soweit wie möglich.

(b) Bestimmen Sie den Maximum-Likelihood-Schätzer $\bar{\theta}_n$ für θ. Begründen Sie dabei auch, dass Sie tatsächlich ein Maximum gefunden haben.

(c) Wählen Sie jeweils die richtige Antwort. Eine Begründung ist nicht notwendig.

(i) Um die Erwartungstreue zu nachzuweisen, ist zu prüfen:
1. $\lim_{n\to\infty} \bar{\theta}_n = \theta$, 2. $\lim_{n\to\infty} \mathbb{E}_\theta[\bar{\theta}_n] = \theta$, 3. $\mathbb{E}[\bar{\theta}_n] = \theta$, 4. $\lim_{n\to\infty} \mathbb{V}_\theta[\bar{\theta}_n] = 0$.

(ii) Für die Effizienz von $\bar{\theta}_n$ ist zu prüfen:
1. $\lim_{n\to\infty} \bar{\theta}_n = \theta$, 2. $\lim_{n\to\infty} \mathbb{E}_\theta[\bar{\theta}_n] = \theta$, 3. $\mathbb{E}[\bar{\theta}_n] = \theta$, 4. $\lim_{n\to\infty} \mathbb{V}_\theta[\bar{\theta}_n] = 0$.

(iii) Für die Konsistenz von $\bar{\theta}_n$ ist zu prüfen:
1. $\lim_{n\to\infty} \bar{\theta}_n = \theta$, 2. $\lim_{n\to\infty} \mathbb{E}_\theta[\bar{\theta}_n] = \theta$, 3. $\mathbb{E}[\bar{\theta}_n] = \theta$, 4. $\lim_{n\to\infty} \mathbb{V}_\theta[\bar{\theta}_n] = 0$.

Aufgabe 5 *10 Punkte*

Es seien $a, b \in [0, 1]$ und X und Y Zufallsvariablen mit folgender gemeinsamer Verteilung:

Y \ X	-2	-1	0	1
0	0	0.1	0.2	a
1	0.2	b	0	0.1

(a) Bestimmen Sie a und b so, dass $\mathbb{E}[X] = 0$ ist.

(b) Berechnen Sie $\mathbb{E}[Y], \mathbb{V}(Y)$ und $\mathbb{E}[XY]$. (Wenn Sie in Teilaufgabe (a) kein Ergebnis gefunden haben, rechnen Sie hier mit $a = b = 0.2$ weiter.)

(c) Geben Sie bei folgenden Aussagen an, ob sie wahr oder falsch sind. Eine Begründung ist hier nicht notwendig.

(i) Wenn Z und W unabhängige Zufallsvariablen sind, so gilt $p_Z(a) \cdot p_W(b) = 0$ für alle $a \in Z(\Omega)$, $b \in W(\Omega)$.

(ii) Wenn Z und W unabhängig sind, so gilt $\operatorname{cov}(Z, W) = 0$.

(iii) Wenn Z und W unabhängig sind, so gilt $\mathbb{P}(Z = a \,|\, W = b) = \mathbb{P}(Z = a) \cdot \mathbb{P}(W = b)$ für alle $a \in Z(\Omega)$, $b \in W(\Omega)$.

(iv) Wenn Z und W unabhängig sind, so kann aus den Randverteilungen die gemeinsame Verteilung berechnet werden.

Lösungen

Lösungsskizze zu Aufgabe 1. Es sind $\mathbb{P}(A) = 0.1$, $\mathbb{P}(B|A) = 0.95$ sowie $\mathbb{P}(B^c|A^c) = 0.9$

(a) (i) $\mathbb{P}(B|A^c) = 1 - \mathbb{P}(B^c|A^c) = 0.1$;

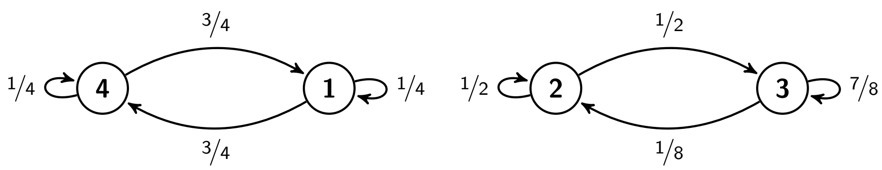
(ii) $\mathbb{P}(B^c \cap A^c) = \mathbb{P}(B^c|A^c)\mathbb{P}(A^c) = 0.9 \cdot 0.9 = 0.81$

(iii) $\mathbb{P}(B^c) = \mathbb{P}(B^c|A)\mathbb{P}(A) + \mathbb{P}(B^c|A^c)\mathbb{P}(A^c) = 0.05 \cdot 0.1 + 0.9 \cdot 0.9 = 0.815$ mittels totaler Wahrscheinlichkeit

(iv) $\mathbb{P}(A|B) = \mathbb{P}(B|A) \cdot \frac{\mathbb{P}(A)}{\mathbb{P}(B)} = 0.95 \cdot \frac{0.1}{1-0.815} = 0.514$ mittels Bayes

(b) (i) (b)((i))4., (ii) (b)((ii))2., (iii) (b)((iii))2., (iv) (b)((iv))1..

Lösungsskizze zu Aufgabe 2.



(a) Wegen der zwei Zusammenhangskomponenten des Übergangsgraphen ist die Kette nicht irreduzibel.

(b) Löse $(P - I)^T \pi = 0$ mit der Nebenbedingung $\pi_1 + \pi_2 + \pi_3 + \pi_4 = 1$, d.h.

$$\begin{array}{rrrrl} -{}^3\!/\!_4\pi_1 & & & +{}^3\!/\!_4\pi_4 & = 0 \\ & -{}^1\!/\!_2\pi_2 & +{}^1\!/\!_8\pi_3 & & = 0 \\ & +{}^1\!/\!_2\pi_2 & -{}^1\!/\!_8\pi_3 & & = 0 \\ -{}^3\!/\!_4\pi_1 & & & +{}^3\!/\!_4\pi_4 & = 0 \end{array}$$

Aus der ersten und vierten kommen jeweils $\pi_1 = \pi_4$, aus der zweten und dritten jeweils $4\pi_2 = \pi_3$. Setze $\pi_2 = t$, so folgt daraus sowie aus der Nebenbedingung

$$\pi = \left({}^{1-5t}\!/\!_2,\ t,\ 4t,\ {}^{1-5t}\!/\!_2\right), \qquad t \in [0, {}^1\!/\!_5].$$

(c) Startet die Kette in 1, so kann sie ne die 3 erreichen und damit

$$\lim_{n\to\infty} \mathbb{P}(X_n = 3 \,|\, X_0 = 1) = 0.$$

Die 4 ereicht sie problemlos und (X_n) sieht weder 2 noch 3. Die Kette auf $\{1, 4\}$ ist irreduzibel und aperiodisch, daher konvergiert das gegen den passenden Wert der invarianten Verteilung (für $t = 0$),

$$\lim_{n\to\infty} \mathbb{P}(X_n = 4 \,|\, X_0 = 1) = {}^1\!/\!_2.$$

Lösungsskizze zu Aufgabe 3.

(a) Es gilt $F_T(a) = 1 - ({}^5\!/\!_a)^2$, und das muss mit 0 übereinstimmen. Das geht nur mit $a = 5$ ($a = -5$ fällt wegen der Bedingung der Positivität weg). Leite nun ab, für $x < a$ ist $f_T(x) = 0$, sonst

$$f_T(x) = -2 \cdot {}^5\!/\!_x \cdot (-{}^5\!/\!_{x^2}) = {}^{50}\!/\!_{x^3}.$$

(b) Die Wahrscheinlichkeiten berechnen sich unmittelbar aus der Verteilungsfunktion:

(i) $\mathbb{P}(T > 11) = 1 - \mathbb{P}(X \leq 11) = 1 - F_T(11) = {}^{96}\!/_{121}$;

(ii) $\mathbb{P}(3 < T \leq 7) = F_T(7) - F_T(3) = {}^{14}\!/_{49} - 0$.

(c) Der Erwartungswert ist

$$\begin{aligned}\mathbb{E}[T] &= \int_{-\infty}^{\infty} t f_T(t)\,\mathrm{d}t = \int_5^{\infty} \frac{50}{t^2}\,\mathrm{d}t \\ &= -\frac{50}{t}\bigg|_{t=5}^{\infty} = 10.\end{aligned}$$

Lösungsskizze zu Aufgabe 4.

(a) Die Likelihoodfunktion ist das Produkt der Dichten,

$$\begin{aligned}L\big(\theta,(x_1,\dots,x_n)\big) &= f_\theta(x_1)\cdots f_\theta(x_n) \\ &= \begin{cases} \left(\frac{2}{\theta\pi}\right)^{n/2}\cdot \mathrm{e}^{-\frac{x_1^2}{2\theta}}\cdots \mathrm{e}^{-\frac{x_n^2}{2\theta}} = \left(\frac{2}{\theta\pi}\right)^{n/2}\cdot \mathrm{e}^{-\frac{1}{2\theta}\sum_{i=1}^n x_i^2}, & \text{falls } x_1,\dots,x_n > 0 \\ 0 & \text{sonst.}\end{cases}\end{aligned}$$

(b) Logarithmiere und leite dann für positive Beobachtung ab

$$\begin{aligned}l\big(\theta,(x_1,\dots,x_n)\big) &= \frac{n}{2}\ln 2 - \frac{n}{2}\ln\theta - \frac{n}{2}\ln\pi - \frac{1}{2\theta}\sum_{i=1}^n x_i^2 \\ l'\big(\theta,(x_1,\dots,x_n)\big) &= -\frac{n}{2}\cdot\frac{1}{\theta} + \frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 \\ &= \frac{n}{2\theta^2}\left(-\theta + \frac{1}{n}\sum_{i=1}^n x_i^2\right) \stackrel{!}{=} 0 \\ l''\big(\theta,(x_1,\dots,x_n)\big) &= \frac{n}{2}\cdot\frac{1}{\theta^2} - \frac{1}{\theta^3}\sum_{i=1}^n x_i^2 \\ &= \frac{n}{2\theta^3}\left(\theta - \frac{2}{n}\sum_{i=1}^n x_i^2\right)\end{aligned}$$

Nullsetzen liefert $\bar{\theta}_n = \frac{1}{n}\sum_{i=1}^n x_i^2$, und das unten eingesetzt liefert

$$l''\big(\bar{\theta}_n,(x_1,\dots,x_n)\big) = \frac{n}{2\bar{\theta}_n^3}\big(\bar{\theta}_n - 2\bar{\theta}_n\big) < 0$$

Damit ist $\bar{\theta}_n$ ein Maximum-Likelihood-Schätzer.

(c) (i) (c)((i))3.; (ii) (c)((ii))4.; (iii) (c)((iii))1..

Lösungsskizze zu Aufgabe 5.

(a) Einerseits muss die Summe über alle Felder 1 ergeben, also $0.6+a+b = 1$, d.h. $a+b = 0.4$, andererseits muss X zentriert sein, d.h.

$$\begin{aligned}0 \stackrel{!}{=} \mathbb{E}[X] &= -2\cdot 0.2 - 1\cdot(0.1+b) + 0\cdot 0.2 + 1\cdot(0.1+a) \\ &= -0.4 + a - b\end{aligned}$$

bzw. $a - b = 0.4$. Addition beider Gleichungen liefert $a = 0.4$ und Subtraktion $b = 0$.

(b) Damit ist $\mathbb{P}(Y = 1) = 0.3$, und da Y bernouliverteilt,

$$\mathbb{E}[Y] = 0.3, \qquad \mathbb{V}(Y) = 0.3 \cdot 0.7 = 0.21.$$

Wegen der vielen Nullen ist

$$\mathbb{E}[XY] = \sum_{i,j} ij p_{X,Y}(i,j) = -2 \cdot 1 \cdot 0.2 + 1 \cdot 1 \cdot 0.1 = -0.3.$$

(b)' Falls mit Alternativwerten weitergerechnet, so $\mathbb{P}(Y = 1) = 0.5$ und damit

$$\mathbb{E}[Y] = 0.5, \qquad \mathbb{V}(Y) = 0.5 \cdot 0.5 = 0.25$$
$$\mathbb{E}[XY] = -0.5$$

(c) (i) falsch; (ii) wahr; (iii) falsch; (iv) wahr.