TU Berlin, Institut für Mathematik Prof. Dr. Noemi Kurt
SoSe 16 Dr. Mathias Rafler

Stochastik für Informatik, (9LP)
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 **??** berechnete Wert, 3. kleiner als der in **??** 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*

Die Übertragungsrate eines Webservers (in MBit/s) wird mehrfach gemessen. Die Daten ergeben

Messung i	1	2	3	4	5	6	7	8	9	10
Rate x_i	9.5	9.4	10.2	9.8	9.5	10.3	9.4	10.8	10.1	9.6

Hierbei ist $\bar{\mu}_n(x_1, \ldots, x_{10}) = 9.86$ und $\bar{\sigma}_n^2(x_1, \ldots, x_{10}) = 0.22$ (diese Werte dürfen ohne Überprüfung verwendet werden).

(a) Wir möchten annehmen, dass diese Daten von normalverteilten Zufallsvariablen stammen. Wie heißt das theoretische Ergebnis, welches diese Annahme ermöglicht?

(b) Es soll ein zweiseitiges Konfidenzintervall für den Erwartungswert der Übertragungsrate zum Fehlerniveau $\alpha = 0.05$ bestimmt werden. Wählen Sie jeweils die richtige Antwort (eine Begründung ist in diesem Aufgabenteil nicht nötig)

(i) Welche Verteilung ist hier relevant?
1. Normalverteilung 2. t-Verteilung 3. χ^2-Verteilung 4. Exponentialverteilung

(ii) Welches Quantil der obigen Verteilung tritt in dem gesuchten Konfidenzintervall auf?
1. q_α, 2. $q_{1-\alpha}$, 3. $q_{\alpha/2}$, 4. $q_{1-\alpha/2}$.

(iii) Welchen Ansatz machen wir für das Konfidenzintervall?
1. $[\bar{\mu}_n - h, \bar{\mu}_n + h]$, 2. $\left[\bar{\sigma}_n^2 - h, \bar{\sigma}_n^2 + h\right]$, 3. $\left[-\frac{h}{\bar{\sigma}_n^2}, \frac{h}{\bar{\sigma}_n^2}\right]$, 4. $]-\infty, \bar{\mu}_n + h]$.

(c) Der Hersteller behauptet, die Übertragungsrate betrage mindestens 10 MBit/s . Dies soll mittels eines geeigneten t-Tests zum Fehlerniveau $\alpha = 0.05$ überprüft werden.

(i) Stellen Sie eine geeignete Nullhypothese auf, und begründen Sie **kurz** ihre Wahl.

(ii) Bestimmen Sie den Freiheitsgrad und den Vergleichswert.

(iii) Berechnen Sie den Testwert, und geben Sie das Ergebnis des Tests an.

Tabelle: Quantile $t_{\beta,f}$ der t-Verteilung für die Freiheitsgrade $f = 8, 9, 10, 11$ und Signifikanzniveaus $\beta \in \{0.01, 0.05, 0.95, 0.99\}$.

Freiheitsgrad f	$t_{0.01,f}$	$t_{0.025,f}$	$t_{0.05,f}$	$t_{0.95,f}$	$t_{0.975,f}$	$t_{0.99,f}$
$f = 8$	-2.90	-2.31	-1.86	1.86	2.31	2.90
$f = 9$	-2.82	-2.26	-1.83	1.83	2.26	2.82
$f = 10$	-2.76	-2.23	-1.81	1.81	2.23	2.76
$f = 11$	-2.72	-2.20	-1.80	1.80	2.20	2.72