Diskrete Strukturen II

Zentralübung

Hanjo Täubig Sebastian Wernicke

16.07.2004

Raumänderung

Die Zentralübung findet an folgendem Tag nicht statt, sondern wie folgt:

- Freitag, 23.07.2004: ZÜ entfällt,
  stattdessen:
  Sprechstunde am Mittwoch, 21.07.2004
  ab 9:00 Uhr in Raum 03.09.011

Aufgabe 1

Das Computeralgebra-Programm Maple ist auf den TU-Rechnern installiert und eine große Hilfe im Leben allgemein sowie bei vielerlei Rechnungen.
Zum Thema "Wahrscheinlichkeitstheorie und Statistik mit Maple" finden Sie zwei sehr gute Einführungen unter
`www.engineering.usu.edu/cee/faculty/gurro/MyBooks/MapleStatIntro.pdf`
sowie unter
`www.mas.ncl.ac.uk/~ndjw1/teaching/maple/`.

Ein Tutorial zur Matrizenrechnung mit Maple finden Sie z.B. unter
`oregonstate.edu/~peterseb/maple/docs/341mp.pdf`.

Bitte machen Sie sich mit den Grundfunktionen vertraut, und lösen Sie dann folgende Aufgabe mit Hilfe von Maple:

Aufgabe 1

Gegeben eine Indikatorvariable I und eine Variable $X \in \{A, C, G, T\}$[1]. Wir beobachten nun I und X und erhalten dabei folgende Tabelle mit den Häufigkeiten gemeinsamen Auftretens:

	X="A"	X="C"	X="G"	X="T"
I=1	1240	1558	210	1250
I=0	1301	1695	165	1200

Machen Sie einen χ^2-Test, der mit 99%-iger Sicherheit die Hypothese "H_0: I und X sind abhängig" verwirft, wenn Sie falsch ist.

[1] Das Problem stammt aus der Anwendung von Markovketten in der Bioinformatik beim Finden von Genen in Genomsequenzen.

Aufgabe 1

Sei $f_i(x)$ die Anzahl der gleichzeitigen Auftreten von $X = x$ und $I = i$. Ferner sei $n = \sum_{i,x} f_i(x)$.
Für den χ^2-Test gehen wir wie folgt vor:
Angenommen, H_0 gilt nicht. Dann würden wir erwarten, dass

$$\begin{aligned}\mathbb{E}_i(X) := f_i(x) &= \frac{f(i) \cdot f(x)}{n} \\ &= \frac{(f_i(A) + f_i(C) + f_i(G) + f_i(T)) \cdot (f_0(x) + f_1(x))}{n}.\end{aligned}$$

Unter der Bedinung, dass H_0 gilt, hat

$$c^2 := \sum_{X,I} \frac{(f_I(X) - \mathbb{E}_i(X))^2}{\mathbb{E}_i(X)}$$

eine χ^2-Verteilung mit $(2-1) \cdot (4-1) = 3$ Freiheitsgraden.

Aufgabe 1

Die Hypothese wird genau dann akzeptiert, wenn

$$c^2 < \chi^2_{3,0.99}$$

In Maple gehen wir also wie folgt vor:

...

Also können wir die Hypothese H_0 auf einem Signifikanzniveau von 99% annehmen.

Aufgabe 2

Gegeben eine Markovkette mit Zustandsmenge $\{\mathcal{S} = \{0, 1, 2, \dots\}\}$, $p_{n(n+1)} = 2/3$ und $p_{n0} = 1/3$ für alle $n \in \mathcal{S}$.

Wie groß ist die Wahrscheinlichkeit, sich nach langer Zeit im Zustand i zu befinden?

Aufgabe 2

Sei $\pi(n)$ die Wahrscheinlichkeit, sich in Zustand n zu befinden. Da wir ja quasi nach der stationären Verteilung der Markovkette suchen, verändert sich π nicht für verschiedene Zeitpunkte. Daher gilt

$$\pi(0) = \frac{1}{3} \cdot \underbrace{(\pi(0) + \pi(1) + \pi(2) + \dots)}_{=1}$$

und

$$\pi(n) = \frac{2}{3}\pi(x-1), \quad n \geq 1$$

woraus unmittelbar

$$\pi(i) = \frac{1}{3} \cdot \left(\frac{2}{3}\right)^i$$

folgt.

Aufgabe 3

Zwei Zustände A und B einer Markovkette gehören zu einer Kommunikationsklasse genau dann, wenn A von B aus erreichbar ist und umgekehrt. Gegeben eine Markovkette mit Zustandsmenge $\mathcal{S} = \{0, 1, 2, 3, 4, 5\}$ und Übergangsmatrix

$$M = \begin{pmatrix} .5 & .5 & 0 & 0 & 0 & 0 \\ .3 & .7 & 0 & 0 & 0 & 0 \\ 0 & 0 & .1 & 0 & .9 & 0 \\ .25 & .25 & 0 & 0 & .25 & .25 \\ 0 & 0 & .7 & 0 & .3 & 0 \\ 0 & .2 & 0 & .2 & .2 & .4 \end{pmatrix}$$

Welche Zustände bilden eine Kommunikationsklasse?
Welche davon sind rekurrent, welche transient?
Wir starten im Zustand 0. Wie groß ist die Wahrscheinlichkeit, sich nach einer längeren Zeit im Zustand 0 zu befinden?

Aufgabe 3

Die Klassen $\{0,1\}$ und $\{2,4\}$ sind rekurrent.
Die Klasse $\{3,5\}$ ist transient.
Da die Klasse $\{0,1\}$ rekurrent ist und wir in ihr anfangen, ist die Wahrscheinlichkeit, sich nach längerer Zeit im Zustand 0 zu befinden, durch die stationäre Verteilung von

$$\begin{pmatrix} .5 & .5 \\ .3 & .7 \end{pmatrix}$$

gegeben:

$$\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \cdot \begin{pmatrix} .5 & .5 \\ .3 & .7 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \Rightarrow \begin{array}{l} .5\pi_1 + .3\pi_2 = \pi_1 \\ .5\pi_1 + .7\pi_2 = \pi_2 \end{array} \Rightarrow \begin{array}{l} \pi_1 = 3/8 \\ \pi_2 = 5/8 \end{array}$$

Also ist $\Pr[\text{Zustand 0 nach längerer Zeit}] = \pi_1 = 3/8$.

Aufgabe 4

Betrachten Sie nochmals die Markovkette aus der letzten Aufgabe.

Wie groß ist die Wahrscheinlichkeit, sich nach einer längeren Zeit im Zustand 0 zu befinden, wenn wir im Zustand 5 starten?

Hinweis: Betrachten Sie zwei Teilmatrizen von M. Die eine, Q, stellt die Übergänge der transienten Klasse in sich selbst dar; die andere, S, die Übergänge von der transienten Klasse in die rekurrente. Die Inverse einer Matrix M kann mit Maples `LinearAlgebra[MatrixInverse](M)` Befehl berechnet werden.

Aufgabe 4

Unsere Strategie ist die folgende: Wir beginnen in der transienten Klasse. Das heißt, wir müssen zunächst die Wahrscheinlichkeit berechnen, aus dieser die rekurrente Klasse zu erreichen, welche den Zustand 0 enthält. Wir suchen also nach einer Matrix K, die uns die Wahrscheinlichkeit angibt, dass wir –ausgehend vom Zustand 5– die Klasse $\{0, 1\}$ erreichen, ohne vorher in der Klasse $\{2, 4\}$ gewesen zu sein.

Setze nun

$$Q := \begin{pmatrix} p_{33} & p_{35} \\ p_{53} & p_{55} \end{pmatrix} = \begin{pmatrix} 0 & .25 \\ .2 & .4 \end{pmatrix}$$

und

$$S := \begin{pmatrix} p_{30} & p_{31} & p_{32} & p_{34} \\ p_{50} & p_{51} & p_{52} & p_{54} \end{pmatrix} = \begin{pmatrix} .25 & .25 & 0 & .25 \\ 0 & .2 & 0 & .2 \end{pmatrix}$$

Aufgabe 4

Sei I die Einheitsmatrix. Dann gilt

$$S = I \cdot K - Q \cdot K$$

Interpretation: Die Wahrscheinlichkeit, die Klasse $\{3, 5\}$ zu verlassen, ist 1 minus die Wahrscheinlichkeit, in der Klasse zu bleiben.

Hieraus folgt

$$K = (I - Q)^{-1} S \Rightarrow K = \begin{pmatrix} 3/11 & 4/11 & 0 & 4/11 \\ 1/11 & 5/11 & 0 & 5/11 \end{pmatrix}$$

Wir haben also

$$p = \Pr[\text{Wir erreichen die Klasse } \{0, 1\}] = 1/11 + 5/11 = 6/11.$$

Aufgabe 4

Der gleiche Wert hätte sich übrigens auch aus der Überlegung

$$p = p_{50} + p_{51} + p_{55} \cdot p + p_{53} \cdot (p_{30} + p_{31}) + p_{53} \cdot (p_{35} + \sum_{k=0}^{\infty} p_{33}) \cdot p$$

heraus ergeben. Dieser Ansatz ist aber, wie man sieht, sehr kompliziert und für größere Markovketten zu aufwendig zu berechnen.
Sind wir erst einmal in dieser Klasse, ist die Wahrscheinlichkeit, sich im Zustand 0 zu befinden, 3/8 laut der vorhergehenden Aufgabe. Wir haben also

$$\Pr[\text{Zustand 0 nach längerer Zeit}] = 6/11 \cdot 3/8 = 9/44 \approx 20.5\%.$$

Raumänderung

Die Zentralübung findet an folgendem Tag nicht statt, sondern wie folgt:

- Freitag, 23.07.2004: ZÜ entfällt,
  stattdessen:
  Sprechstunde am Mittwoch, 21.07.2004
  ab 9:00 Uhr in Raum 03.09.011