ulm university universität uulm

Stochastic Simulations
SoSe 2017
22. Mai 2017

Universität Ulm
Dr. Kirsten Schorning
Dr. Vitalii Makogin

Exercise sheet 3 till May 26, 2017

Theory (total – 16 points)

Exercise 4-1 (3+3+2 points)

Let $\{X_n\}_{n\in\mathbb{N}}$ be a Markov chain, and $\tau = \inf\{n \geq 0 | X_n \in A\}$ be a time of the first visit of a set A by the chain. Denote $r_i = \mathbf{P}(\tau < \infty | X_0 = i)$, $\rho_i = \mathbf{E}(\tau | X_0 = i) \in (0, \infty]$. Prove that:
(a) Probabilities $\{r_i\}$ satisfy the system of linear equations

$$\begin{cases} r_i = \sum_j p_{ij} r_j, & i \notin A, \\ r_i = 1, & i \in A, \\ r_i = 0, & i \not\to A \text{ and } i \notin A. \end{cases}$$

(b) Expectations $\{\rho_i\}$ satisfy the following system of equations

$$\begin{cases} \rho_k = 1 + \sum_j p_{kj} \rho_j, & k \notin A, \\ \rho_k = 0, & k \in A. \end{cases}$$

(c) If the phase space is finite then the expectation ρ_i is finite if and only if for any state j for which $i \to j$ there exists a state $k \in A$ accessible from $j (j \to k)$.

Exercise 4-2 (4 points)

Let $\{X_n\}_{n\in\mathbb{N}}$ be Markov chain with state space $E = \{1, 2\}$, and transition matrix

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Classify the states. Suppose that $\alpha\beta > 0$ and $\alpha\beta \neq 1$. Find the n-step transition probabilities and show directly that they converge to the unique stationary distribution as $n \to \infty$. For what values of α and β is the chain reversible?

Exercise 4-3 (2+2 points)

Consider a Markov chain $\{X_n\}_{n\in\mathbb{N}}$ with transition matrix P and stationary distribution $\boldsymbol{\pi}$. Show that

(a) $\frac{P+\widehat{P}}{2}$ and

(b) $\widehat{P}P$

are stochastic matrices and invariant for $\boldsymbol{\pi}$, where $\widehat{P}$ is the transition matrix of the time-reversal of $\{X_n\}_{n\in\mathbb{N}}$.

Programming (total – 7 points)

Exercise 4-4 (3 points)

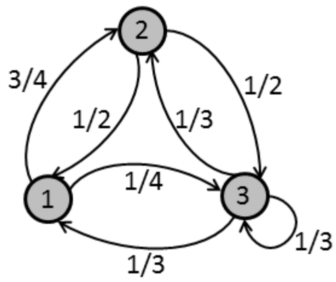
Consider the transition matrix:

$$P = \begin{pmatrix} 0.1500 & 0.3500 & 0.3500 & 0.1500 \\ 0.1660 & 0.3340 & 0.3340 & 0.1660 \\ 0.1875 & 0.3125 & 0.3125 & 0.1875 \\ 0.2000 & 0.3000 & 0.3000 & 0.2000 \end{pmatrix}.$$

Generate DNA sequences of nucleotides "A", "C", "G", "T" according to a first order Markov chain using the transition matrix given above. The initial distribution is $\{p_A = 0.2, p_C = 0.1, p_G = 0.1, p_T = 0.6\}$.
The GC content of a DNA sequence is defined as the percentage of C's and G's on the total number of bases of the sequence. Calculate the GC content for an infinitely long DNA sequence generated in accordance with the sampling model above. Confirm this by simulation: generate a DNA sequence of (say) length 100000 and calculate its GC content. Calculate the stationary distribution of the transition matrix P analytically and through matrix multiplication.

Exercise 4-5 (4 points)

Consider a Markov chain $\{Y_n\}_{n\in\mathbb{N}}$ with the following transition graph.



Is this Markov chain reversible? Provide a proof of your answer.
Consider this Markov chain, started from its invariant distribution $\boldsymbol{\pi}$.

(a) Write a R program to simulate this Markov chain *backwards* from $n = 50$ to $n = 0$ and plot a realization of it.

(b) Repeat the simulation in a) from (at least) $n = 10^5$ to $n = 0$ and estimate the transition probabilities P of the forward Markov chain $\{Y_n\}_{n\in\mathbb{N}}$.