Stefan Grosskinsky, Emma Southall 03.10.2019

Stochastic Modelling and Random Processes

Problem sheet 1

Sheet counts $50/150$ homework marks, [x] indicates weight of the question.
Please put solutions in my pigeon hole or give them to me by **Friday, 18.10.2019, 1pm**.

All plots must contain axis labels and a legend (can be added by hand if necessary). Use your own judgement to find reasonable and relevant plot ranges.

1.1 Simple random walk (SRW) [18]

(a) Consider a SRW on $\{1, \ldots, L\}$ with probabilities $p \in [0, 1]$ and $q = 1 - p$ to jump right and left, respectively, and consider periodic as well as closed boundary conditions.
For both cases, sketch the transition matrix P of the process (see lectures).
Decide whether the process is irreducible, and give all stationary distributions $\boldsymbol{\pi}$ and state whether they are reversible. (Hint: Use detailed balance.)
Where necessary, discuss the cases $p = 1$ and $p = q = 1/2$ separately from the general case $p \in (0, 1)$.

(b) Consider the same SRW with absorbing boundary conditions, sketch the transition matrix P, decide whether the process is irreducible, and give all stationary distributions $\boldsymbol{\pi}$ and state whether they are reversible.
Let $h_k^L = \mathbb{P}[X_n = L \text{ for some } n \geq 0 | X_0 = k]$ be the absorption probability in site L.
Give a recursion formula for h_k^L and solve it for $p \neq q$ and $p = q$.

(c) Simulate 500 realizations of a SRW with $L = 10$, closed boundary conditions and with a value for $p = 1 - q \in (0.6, 0.9)$ of your choice. For all simulations use $X_0 = 1$.
Plot the empirical distribution after 10 and 100 time steps in form of a histogram, and compare it with the theoretical stationary distribution from (a).
Repeat a single realization of the same simulation up to 50 and 500 time steps and plot the fraction of time spent in each state as a histogram, comparing to the stationary distribution.

1.2 Generators and eigenvalues [14]

Consider the continuous-time Markov chain $(X_t : t \geq 0)$ with generator $G = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -4 & 3 \\ 0 & 1 & -1 \end{pmatrix}$.

(a) Draw a graph representation for the chain (i.e. connect the three states by their jump rates), and give the transition matrix P^Y of the corresponding jump chain $(Y_n : n \in \mathbb{N}_0)$.

(b) Consider the Taylor series of the matrix P_t and convince yourself that $\frac{d}{dt} P_t|_{t=0} = G$, $\frac{d^2}{dt^2} P_t|_{t=0} = G^2$ etc..
Assume that $G = B^{-1} \Lambda B$, with diagonal matrix $\Lambda \in \mathbb{C}^{3\times 3}$ and eigenvalues λ_i of G on the diagonal, and with some matrix $Q \in \mathbb{C}^{3\times 3}$. Show that

$$P(t) = \exp(tG) = Q^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{pmatrix} Q \,.$$

(It is not necessary to compute entries of the matrix Q!)

(c) Compute λ_2 and λ_3. Use this to compute $p_{11}(t)$, i.e. determine the coefficients in

$$p_{11}(t) = a + b\, e^{\lambda_2 t} + c\, e^{\lambda_3 t} .$$

(Again, it is not necessary to compute the matrix Q, instead use what you know about $p_{11}(0)$ and $\frac{d}{dt} p_{11}(t)|_{t=0}$ etc.)

(d) What is the stationary distribution $\boldsymbol{\pi}$ of X?

1.3 Pólya urn models [18]

Consider the following experiment: Place k balls each of distinct color indexed by $i = 1, \dots, k$ in an urn. Draw one ball uniformly at random, then replace *two* balls of the colour just drawn in the urn. Iterate.

(a) We want to keep track of the contents $\underline{X}(n)$ of the urn as a function of discrete time n. Give the state space S, the initial condition $\underline{X}(0)$ and the transition probabilities $p(\underline{x}, \underline{y})$, $\underline{x}, \underline{y} \in S$ to define a stochastic process $\big(\underline{X}(n) : n = 0, 1, \dots\big)$ which does the job.

(b) For $k = 2$ sketch the state space and the transition probabilities between states.
Still for $k = 2$, show that for all $(x_1, x_2) \in S$ and all $n \geq 1$

$$\mathbb{P}\big[\underline{X}(n) = (x_1, x_2)\big] = \frac{1}{n+1}\, \delta_{n+2, x_1+x_2} ,$$

i.e. the distribution at time n is uniform.

Use this to show that $\quad \dfrac{1}{n+2} \underline{X}(n) \to (U, 1-U) ,$

where $U \sim U[0,1]$ is a uniform random variable on $[0,1]$.

Consider a **generalized Pólya urn model** with k types or colours on the same state space S as above, but now with transition probabilities

$$p(\underline{x}, \underline{x} + \underline{e}_i) = \frac{f_i x_i^\gamma}{\sum_{j=1}^k f_j x_j^\gamma} \qquad \big(\text{and } p(\underline{x}, \underline{y}) = 0 \text{ if } \underline{y} \neq \underline{x} + \underline{e}_i\big) ,$$

where the $f_i > 0$ denote the **fitness** of type i and $\gamma \geq 0$ is a **reinforcement parameter**.

(c) Simulate the model for $k = 500$ types with equal fitness $f_i \equiv 1$ for $\gamma = 0,\ 0.5,\ 1$ and 1.5. For each γ, show the **empirical tail distributions** of $\underline{X}(n)$ for $n = 5000$, 20000 and 80000 in one plot, and do the same for the normalized data $\frac{1}{n+k} \underline{X}(n)$ (8 plots in total). Choose the plot ranges reasonably and explain what you observe.

Background info: For $\gamma > 1$ it is known that the system exhibits **monopoly**, i.e. as $n \to \infty$ almost all balls in the urn will be of a single type.

(d) Play around with the fitness parameters f_i, find something 'interesting', and show one plot and write a few sentences to explain it.

1.4 Wright-Fisher model of population genetics **(Class only)**

Consider a fixed population of L individuals. At time $t = 0$ each individuum i has a different type $X_0(i)$, for simplicity we simply put $X_0(i) = i$. Time is counted in discrete generations $t = 0, 1, \ldots$. In generation $t+1$ each individuum i picks a parent $j \sim U(\{1, \ldots, L\})$ uniformly at random, and adopts its type, i.e. $X_{t+1}(i) = X_t(j)$. This leads to a discrete-time Markov chain $(X_t : t \in \mathbb{N})$.

(a) Give the state space of the Markov chain $(X_t : t \in \mathbb{N})$. Is it irreducible? What are the stationary distributions?
(Hint: if unclear do (c) first to get an idea.)

(b) Let N_t be the number of individuals of a given species at generation t, with $N_0 = 1$. Is $(N_t : t \in \mathbb{N})$ a Markov process? Give the state space and the transition probabilities.
Is the process irreducible? What are the stationary distributions? What is the limiting distribution as $t \to \infty$ for the initial condition $N_0 = 1$?

(c) Simulate the dynamics of the full process $(X_t : t \in \mathbb{N})$ up to generation T.
Store the trajectory $(X_t : t = 1, \ldots, T)$ in a $T \times L$ matrix, with ordered types such that $X_t(1) \leq \ldots \leq X_t(L)$ for all t, and visualise the matrix with a heat map.
You may use the suggested parameter value $L = 100$ and appropriate T, or any other that make sense (it is a good idea to vary them to get a feeling for the model).
Address the following points, supported by appropriate visualisations:
- Explain the emerging patterns in a couple of sentences, what will happen when you run the simulation long enough?
- How long will it roughly take on average to reach stationarity (depending on L)?
Test your answer using (at least) three values for L, e.g. 10, 50 and 100.