Networks and Random Processes

Problem sheet 1

Sheet counts $45/100$ homework marks, [x] indicates weight of the question.
Please put solutions in my pigeon hole or give them to me by **Tuesday, 16.10.2018, 2pm**.

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) [15]

(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 500 time steps and plot the fraction of time spent in each state as a histogram, comparing to the stationary distribution.

1.2 Geometric random walk [15]

Let $X_1, X_2, \ldots$ be a sequence of iidrv's with $X_i \sim \mathcal{N}(\mu, \sigma^2)$ where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Consider the discrete-time random walk (DTRW) on state space $\mathbb{R}$

$$(Y_n : n \geq 0) \quad \text{with} \quad Y_{n+1} = Y_n + X_{n+1} \quad \text{and} \quad Y_0 = 0\ .$$

(a) State the weak law of large numbers and the central limit theorem for Y_n.

(b) What is the distribution of Y_n for any arbitrary $n \geq 0$?

Now consider the discrete-time process $(Z_n : n \geq 0)$ on the state space $[0, \infty)$ with $Z_n = \exp(Y_n)$, which is called a **geometric random walk**.

(c) Give a recursive definition of $(Z_n : n \geq 0)$ analogous to the above.
Show that Z_n has a log-normal distribution for all $n \geq 1$ by deriving the PDF.
Give the mean, variance and median of Z_n (you can look this up on the web).

(d) Simulate $M = 500$ realizations of Z_n for $n = 0, \ldots, 100$ with $\mu = 0$ and $\sigma = 0.2$.
Plot the **empirical average** $\hat{\mu}_n^M := \frac{1}{M} \sum_{i=1}^{M} Z_n^i$ as a function of time n, with error bars indicating the standard deviation.
At times $n = 10$ and 100 produce boxplots, plot the empirical PDF using a kernel density estimation, and compare it to the theoretical prediction.
For a single realization, plot the **ergodic average** $\bar{\mu}_N := \frac{1}{N} \sum_{n=1}^{N} Z_n$ as a function of N up to $N = 100$.

(e) For fixed $\sigma = 0.2$ pick μ such that $\mathbb{E}[Z_n] \equiv 1$ for all $n \geq 0$, and repeat the previous simulation. This time, plot the empirical tail of the data and compare to the theoretical prediction (given by $1-$CDF) at times $n = 10$ and 100.

(f) In the following consider the choice of parameters from (e).
What is the long-time behaviour of Z_n as $n \to \infty$? Does $(Z_n : n \geq 0)$ have a stationary distribution on $[0, \infty)$? Is the process ergodic?
As the number of realizations $M \to \infty$, does the empirical average $\hat{\mu}_n^M$ converge for fixed $n \geq 0$, and what is the limit?
For fixed finite M, do you think that $\hat{\mu}_n^M$ converges as $n \to \infty$, and what is the limit?
Do the limits commute, i.e. $\lim_{n\to\infty} \lim_{M\to\infty} \hat{\mu}_n^M = \lim_{M\to\infty} \lim_{n\to\infty} \hat{\mu}_n^M$?
What is the limit of $\bar{\mu}_N$ as $N \to \infty$? Why does this not contradict the ergodic theorem?

1.3 Wright-Fisher model of population genetics [15]

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.