THE UNIVERSITY OF AUCKLAND

SECOND SEMESTER, 2012
Campus: City

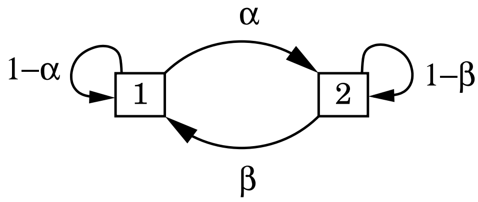
STATISTICS

Stochastic Processes

(Time allowed: THREE hours)

NOTE: Attempt **ALL** questions. Marks for each question are shown in brackets.
There are 100 marks in total.
An **Attachment** containing useful information is found on page 8.

1. Let $\{X_1, X_2, X_3, \ldots\}$ be a Markov chain on the state space $S = \{1, 2\}$, with the following transition diagram:



Assume that $0 < \alpha < 1$ and $0 < \beta \leq 1$. (Note that β can be 1, but α can not be 1.)

(a) Write down the transition matrix for the Markov chain represented by this diagram. Be careful to insert α and β in the matrix correctly. (2E)

(b) Suppose the chain is equally likely to be in either of the two states at time 1. Find $\mathbb{P}(X_1 = 2, X_2 = 2, X_3 = 1)$. (2E)

(c) Find an equilibrium distribution, $\boldsymbol{\pi}$, for this Markov chain. (3E)

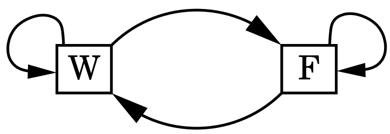
(d) Does the Markov chain converge to the equilibrium distribution in (c) as $t \to \infty$, for all possible α and β with $0 < \alpha < 1$ and $0 < \beta \leq 1$? Explain why or why not. (3E)

A factory relies on a machine that is often faulty. Every day, the machine is switched on. If it was working on the previous day, it fails with probability 0.1 when switched on for the new day's business. When the machine fails, the day's business is lost and an engineer is hired to fix it.

Every day that the machine is not working, the manager spends amount \$$100p$ on repairing it. With this amount of money spent on repairs, the machine is fixed on that day with probability p, independently of outcomes on other days. Once fixed, the machine works again on the following day.

Overall, on each day that the machine is working, it earns \$100. For each day that it is being fixed and therefore not working, it costs \$$100p$ in repair bills.

(e) The factory problem can be formulated as a two-state Markov chain. The states are W for days that the machine is working, and F for days when the machine has failed and is being fixed. Copy the diagram below and add the correct probabilities to each arrow, taken from the description in the previous paragraph. (2M)



(f) Let $f(p)$ be the **long-term expected daily earnings** from the machine for a chosen value of p. Show that

$$f(p) = \frac{90p}{0.1 + p}.$$

You must fully justify all parts of your answer. (3M)

(g) Recall that the machine is fixed with probability p if \$$100p$ is spent. Use the expression in part (f) to say what value of p the factory manager should choose.
[Hint: what is the gradient of $f(p)$?] (3M)

[18 marks]

2. A test for a particular disease operates as follows. The patient provides a blood sample, which goes through a first test. If the test shows that the patient is infected with the disease, the blood sample is sent for a second test to diagnose cause and severity.

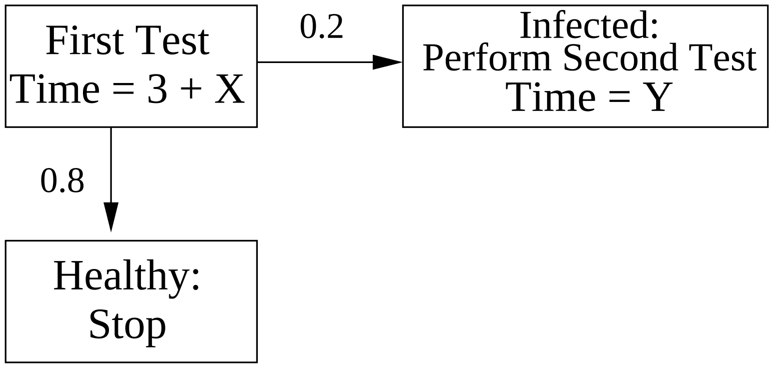
   The first test takes $3 + X$ days to complete, where $X \sim \text{Poisson}(1)$. The second test takes a further Y days to complete, where $Y \sim \text{Poisson}(2)$. Assume that X and Y are independent.

   Define the indicator random variable D, such that

$$D = \begin{cases} 1 & \text{if patient has the disease,} \\ 0 & \text{if patient is healthy.} \end{cases}$$

   Suppose that $\mathbb{P}(D = 1) = 0.2$ in the population being tested.

   A patient's progress through the testing process is shown in the diagram below.



   Define the random variable T to be the total time taken before the sample testing is complete. Thus T is the total time taken for the first test and (if needed) the second test.

   (a) Show that the total expected testing time is $\mathbb{E}(T) = 4.4$ days. You may use any results from the Attachment that you need. (2E)

   (b) The probability generating functions (PGFs) of X and Y are $G_X(s) = \mathbb{E}(s^X) = e^{(s-1)}$ for $X \sim \text{Poisson}(1)$, and $G_Y(s) = \mathbb{E}(s^Y) = e^{2(s-1)}$ for $Y \sim \text{Poisson}(2)$. Find $G_{X+Y}(s)$, the PGF of $X + Y$, and hence name the distribution of $X + Y$, giving the values of any parameters. (3E)

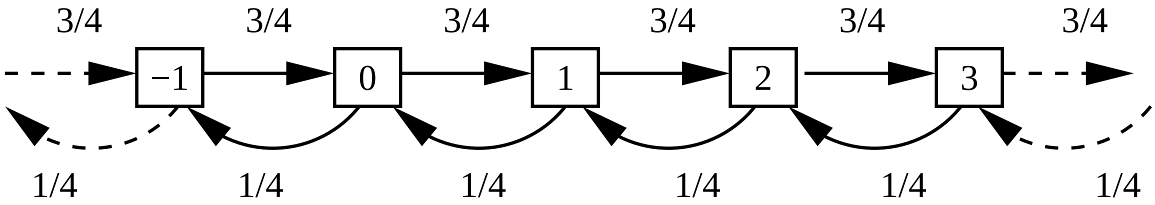
   (c) Find $\mathbb{P}(T \geq 5 \,|\, D = 0)$, the probability that a healthy sample will take at least 5 days to process; and $\mathbb{P}(T \geq 5 \,|\, D = 1)$, the probability that a diseased sample will take at least 5 days to process. (3M)

   (d) An anxious patient has already waited 5 days for her test result. She wants to know the probability that she has the disease, given that her sample has taken at least 5 days to process. Find $\mathbb{P}(D = 1 \,|\, T \geq 5)$. (2M)

   (e) Using the law of total variance, or otherwise, find $\text{Var}(T)$. (5H)

[15 marks]

3. Let $\{X_0, X_1, X_2, \ldots\}$ be a random walk on the integers, with transition diagram below.



Let U be the number of steps taken to reach state 1, starting at state 0. Let $H_U(s) = \mathbb{E}(s^U)$ be the probability generating function of U.

(a) Show that $H_U(s)$ must be either the (+) root or the (−) root of the following expression:

$$H_U(s) = \frac{4 \pm \sqrt{16 - 12s^2}}{2s}.$$ (4M)

(b) By considering $\lim_{s \to 0} H_U(s)$, prove that $H_U(s)$ can **not** be the (+) root in the expression above. (2M)

(c) Using the expression $H_U(s) = \dfrac{4 - \sqrt{16 - 12s^2}}{2s}$, find whether U is a defective random variable. Hence state the probability that the process *never* reaches state 1, starting from state 0. (3M)

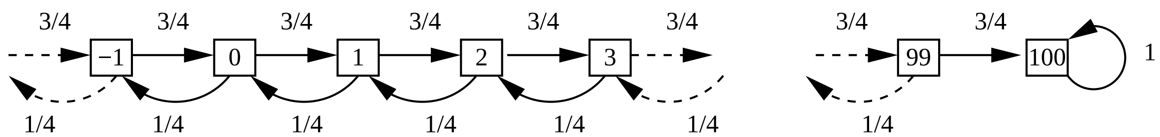
(d) Let V be the number of steps taken to reach state −1, starting at state 0. Let $H_V(s) = \mathbb{E}(s^V)$ be the probability generating function of V. You may assume that

$$H_V(s) = \frac{4 - \sqrt{16 - 12s^2}}{6s}.$$

What is the probability that the process *never* reaches state −1, starting from state 0? (2M)

(e) Let T be the number of steps taken to **first return to** state 0, starting at state 0. For example, if $X_0 = 0$, $X_1 = 1$, and $X_2 = 0$, then $T = 2$ steps are taken to return from 0 to 0 again. Let $G(s) = \mathbb{E}(s^T)$ be the probability generating function of T. Using $G(s)$, find the probability that the process *never* returns to state 0, starting from state 0. (3M)

(f) Now suppose that the random walk has a boundary at state 100, as shown in the transition diagram below. Are your calculations for the probabilities in parts (c), (d), and (e) still valid? Briefly say why or why not. (2M)



[16 marks]

4. A man owns N umbrellas. Every day, he walks between his home and his workplace and back again. On each of his trips, from home to work or from work to home, he carries an umbrella only if he has an umbrella available and it is raining. It rains on each trip with probability r, independently of all other trips, where $0 < r < 1$.

   Let $\{X_1, X_2, X_3, \ldots\}$ be a Markov chain on the state space $S = \{0, 1, 2, \ldots, N\}$, such that X_t is the **number of umbrellas available** on trip t. For example, if trip t is a trip from home to work, then $X_t = n$ if there are n umbrellas at home, with the other $N - n$ umbrellas being at work.

   (a) Draw the transition diagram when the man owns $N = 1$ umbrella. Write down the equilibrium equations, and solve them to find the equilibrium distribution $\boldsymbol{\pi}^T = (\pi_0, \pi_1)$ when $N = 1$. (4E)

   (b) Draw the transition diagram when the man owns $N = 2$ umbrellas. By direct substitution into the appropriate equilibrium equations, verify that the Markov chain has the following equilibrium distribution when $N = 2$:

   $$\boldsymbol{\pi}^T = (\pi_0, \pi_1, \pi_2) = \left(\frac{1-r}{3-r}, \quad \frac{1}{3-r}, \quad \frac{1}{3-r} \right).$$

   (4M)

   (c) Suppose the man has N umbrellas. We wish to prove that the Markov chain has the following equilibrium distribution:

   $$\boldsymbol{\pi}_\star^T = (\pi_0,\ \pi_1,\ \pi_2, \ldots,\ \pi_N) = \frac{1}{N+1-r} \Big(1-r,\ \ 1,\ \ 1, \ldots,\ \ 1 \Big).$$

   Find $\mathbb{E}\left(s^{X_{t+1}} \,|\, X_t\right)$, and hence show that $X_{t+1} \sim X_t$ if $X_t \sim \boldsymbol{\pi}_\star$. Deduce that $\boldsymbol{\pi}_\star$ is the equilibrium distribution for this Markov chain, and (with justification) state the long-term probability that the man gets wet on a trip. You may assume that the chain is aperiodic for any value of N. (6H)

[14 marks]

5. Let $\{X_1, X_2, X_3, \ldots\}$ be a Markov chain on the state space $S = \{1, 2, \ldots, N\}$, with transition matrix $P = (p_{ij})$, such that:

- for $i = 2, 3, \ldots, N$,
$$p_{ij} = \begin{cases} \frac{1}{i} & \text{for } j = 1, \ldots, i\,, \\ 0 & \text{for } j = i+1, \ldots, N\,; \end{cases}$$
- for $i = 1$,
$$p_{1j} = \begin{cases} 0 & \text{for } j = 1, \ldots, N-1\,, \\ 1 & \text{for } j = N\,. \end{cases}$$

(a) Draw the transition diagram for $N = 4$. (2M)

(b) Does the Markov chain $\{X_t\}_{t \geq 1}$ converge to an equilibrium distribution as $t \to \infty$, for any $N \geq 2$? Explain why or why not. (2M)

(c) For $x = 1, 2, \ldots, N$, define m_x to be the expected reaching time for state 1, starting from state x: that is,
$$m_x = \mathbb{E}(\text{ number of steps taken to hit state 1 } \mid \text{ start at state } x\text{ }).$$
We define $m_1 = 0$. Using first-step analysis, show that
$$m_{x+1} = m_x + \frac{1}{x} \quad \text{for } x = 2, \ldots, N-1,$$
and find m_2. (4H)

(d) A theorem that we have not studied states that, for an irreducible Markov chain on a finite state space, the equilibrium distribution $\boldsymbol{\pi}$ satisfies
$$\pi_k = \frac{1}{R_{kk}}$$
where R_{kk} is the **expected return time** for state k. That is, R_{kk} is the expected number of steps needed to return to state k for the first time, starting at state k.
(Note that $R_{kk} \geq 1$, because at least one step is needed to *return* to state k, starting from state k. The expected return times differ from the expected reaching times, which are defined as 0 when starting at the target state.)
Show that the long-run proportion of time that the Markov chain spends in state 1 is
$$\pi_1 = \left(2 + \sum_{r=1}^{N-1} \frac{1}{r}\right)^{-1}.$$
You must fully justify all parts of your answer. (4H)

[12 marks]

6. Let $\{Z_0, Z_1, Z_2, \ldots\}$ be a branching process, where Z_n denotes the population size at time n, and $Z_0 = 1$. Let Y be the family size distribution. Suppose that $Y \sim \text{Geometric}(p = 0.5)$, so that

$$\mathbb{P}(Y = y) = \left(\frac{1}{2}\right)^{y+1} \qquad \text{for } y = 0, 1, 2, \ldots$$

(a) Let $G(s) = \mathbb{E}(s^Y)$ be the probability generating function of Y. Show that

$$G(s) = \frac{1}{2-s},$$

and state the range of values of s for which this expression is valid. (3E)

(b) Find the probability of eventual extinction, γ. (3E)

(c) Let $G_n(s) = \mathbb{E}\left(s^{Z_n}\right)$ be the probability generating function of the population size at time n. State (without calculation) the correct expression for $G_{n+1}(s)$ in terms of $G_n(s)$. That is, rewrite the expression below, with $\star$ replaced by the correct quantity:

$$G_{n+1}(s) = G_n\left(\ \star\ \right).$$

(1E)

(d) For $Y \sim \text{Geometric}(p = 0.5)$ as above, prove by mathematical induction that

$$G_n(s) = \frac{n - (n-1)s}{(n+1) - ns} \qquad \text{for } n = 1, 2, 3, \ldots$$

(6M)

(e) Using the expression shown in (d), state expressions for the following probabilities:
(i) γ_n, that the process is extinct by generation n: that is, $\gamma_n = \mathbb{P}(Z_n = 0)$;
(ii) ρ_n, that there are surviving individuals in generation n: that is, $\rho_n = \mathbb{P}(Z_n > 0)$. (2M)

(f) An idea of interest in branching processes is the **Most Recent Common Ancestor**. If there are z individuals alive in generation n, we can trace back the parents, grandparents, and so on, for each of the z individuals until we get to a single individual in an earlier generation who is a common ancestor for all z individuals. The **most recent** common ancestor is the most recent of all common ancestors: that is, the common ancestor that lived closest in time to generation n.

Let T_n be the generation number of the most recent common ancestor for individuals surviving at generation n. For example, if all individuals at time n had the same parent at time $n-1$, then $T_n = n-1$. Alternatively, if there is no common ancestor until the single individual at generation 0, then $T_n = 0$. (Note that the individual who started the branching process at time 0 is always a common ancestor to all individuals alive at generation n.)

For the branching process above, with $Y \sim \text{Geometric}(p = 0.5)$, show that

$$\mathbb{P}(Z_n > 0 \text{ and } T_n = 0) = \sum_{y=2}^{\infty} \left(\frac{1}{2}\right)^{y+1} \left\{1 - \left(\frac{n-1}{n}\right)^{y} - y\left(\frac{n-1}{n}\right)^{y-1}\left(\frac{1}{n}\right)\right\}.$$

Hence find an expression for $\alpha_{n,0}$: the conditional probability, given that there are individuals still alive in generation n, that their most recent common ancestor lived in generation 0.

Lastly, find a similar expression for $\alpha_{n,1}$ in terms of $\alpha_{n-1,0}$, where $\alpha_{n,1} = \mathbb{P}(T_n = 1 \mid Z_n > 0)$. (10H)

[25 marks]

ATTACHMENT

1. **Discrete Probability Distributions**

Distribution	$\mathbb{P}(X = x)$	$\mathbb{E}(X)$	$\text{Var}(X)$	PGF, $\mathbb{E}(s^X)$
Geometric(p)	pq^x (where $q = 1 - p$), for $x = 0, 1, 2, \ldots$	$\dfrac{q}{p}$	$\dfrac{q}{p^2}$	$\dfrac{p}{1 - qs}$
	Number of failures before the first success in a sequence of independent trials, each with $\mathbb{P}(\text{success}) = p$.			
Binomial(n, p)	$\binom{n}{x} p^x q^{n-x}$ (where $q = 1 - p$), for $x = 0, 1, 2, \ldots, n$.	np	npq	$(ps + q)^n$
	Number of successes in n independent trials, each with $\mathbb{P}(\text{success}) = p$.			
Poisson(λ)	$\dfrac{\lambda^x}{x!} e^{-\lambda}$ for $x = 0, 1, 2, \ldots$	λ	λ	$e^{\lambda(s-1)}$

2. **Uniform Distribution:** $X \sim \text{Uniform}(a, b)$.
Probability density function, $f_X(x) = \frac{1}{b-a}$ for $a < x < b$. Mean, $\mathbb{E}(X) = \frac{a+b}{2}$.

3. **Properties of Probability Generating Functions**

Definition: $G_X(s) = \mathbb{E}(s^X)$

Moments: $\mathbb{E}(X) = G'_X(1)$ $\quad \mathbb{E}\left\{X(X-1)\ldots(X-k+1)\right\} = G_X^{(k)}(1)$

Probabilities: $\mathbb{P}(X = n) = \dfrac{1}{n!} G_X^{(n)}(0)$

4. **Geometric Series:** $1 + r + r^2 + r^3 + \ldots = \sum_{x=0}^{\infty} r^x = \dfrac{1}{1-r}$ for $|r| < 1$.

Finite sum: $\sum_{x=0}^{n} r^x = \dfrac{1 - r^{n+1}}{1 - r}$ for $r \neq 1$.

5. **Binomial Theorem:** For any $p, q \in \mathbb{R}$, and integer $n > 0$, $(p+q)^n = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x}$.

6. **Exponential Power Series:** For any $\lambda \in \mathbb{R}$, $\sum_{x=0}^{\infty} \dfrac{\lambda^x}{x!} = e^{\lambda}$.