THE UNIVERSITY OF AUCKLAND

SECOND SEMESTER, 2011
Campus: City

STATISTICS

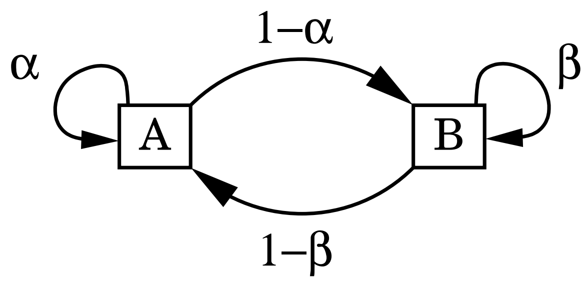
Special Topic in Applied Probability

(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. Consider a **two-armed bandit process** for trialling two treatments A and B on a succession of patients. All patients are independent. For any patient, the probability that treatment A is successful is α, and the probability that treatment B is successful is β. If the treatment for patient t was **successful**, the **same** treatment is used for patient $t + 1$. If the treatment for patient t was **unsuccessful**, the treatment is **changed** for patient $t+1$. The state of the process at time t is the treatment (A or B) that is given to patient t.

   The transition diagram for the two-armed bandit process is below. Assume that $0 < \alpha < 1$ and $0 < \beta < 1$.



   (a) Write down the transition matrix for the Markov chain represented by this diagram. (1E)

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

   (c) Does the Markov chain converge to the equilibrium distribution in (b) as $t \to \infty$? Explain why or why not. (2E)

   (d) Show that the long-run probability of *success* for each patient in this process is

$$\frac{\alpha + \beta - 2\alpha\beta}{2 - \alpha - \beta}.$$

   (2M)

   (e) Suppose that treatment A is better than treatment B, so $\alpha > \beta$. Show that the two-armed bandit strategy has a **lower** long-run probability of success for each patient than an alternative strategy that applies treatment A to every patient. In view of this, explain why we might ever wish to use the two-armed bandit strategy. (4M)

[12 marks]

2. Consider the model for gene spread studied in class. A population consists of N individuals, where N is constant. Each individual possesses one of two alleles: the harmful allele A, or the safe allele B. Let $X_0, X_1, X_2, \dots$ be a Markov chain such that X_t is the number of individuals in generation t that possess the harmful allele A.

The chain evolves according to the following relationship:

$$[\, X_{t+1} \,|\, X_t \,] \sim \text{Binomial}\left(N,\ \frac{X_t}{N} \right) .$$

The chain is said to reach **fixation** when either $X_t = 0$ or $X_t = N$ for some t. We are interested in the probability that the chain reaches fixation at state N, such that all individuals eventually possess the harmful allele A.

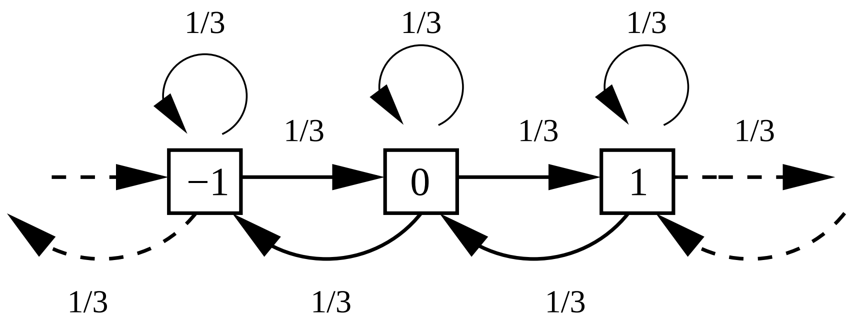
(a) Show that the chain $X_0, X_1, X_2, \dots$ is a martingale. (3M)

(b) Define the random variable T to be the generation at which fixation is reached. Explain why T is a stopping time with respect to $\{X_t\}$. (2M)

(c) Suppose that the chain begins with $X_0 = x$ individuals possessing allele A. Use the Optional Stopping Theorem for bounded martingales to find the probability that the population eventually becomes fixed for the harmful allele A. Your answer should include clear notation, and justification that the Optional Stopping Theorem can be applied to this process. You may assume that $\mathbb{E}(T) < \infty$. (5H)

[10 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:

$$H_U(s) = \frac{3 - s \pm \sqrt{9 - 6s - 3s^2}}{2s}\,.$$ (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 $H_U(s) = \dfrac{3 - s - \sqrt{9 - 6s - 3s^2}}{2s}$, find whether U is a defective random variable. (2M)

(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. Explain why $H_V(\cdot) = H_U(\cdot)$. (1M)

(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. Show that

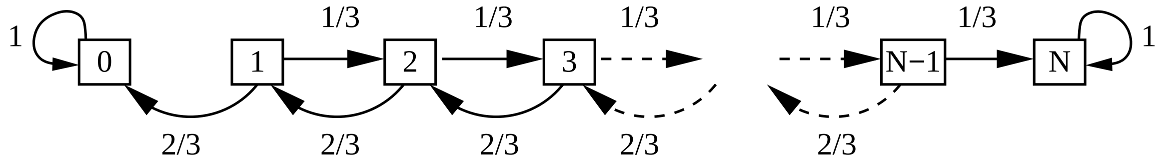
$$G(s) = 1 - \frac{1}{3}\sqrt{9 - 6s - 3s^2}.$$

(4M)

(f) If the process is currently in state 5, what is the probability it will **never** be in state 5 again? Explain your answer. (2M)

[15 marks]

4. Consider the Gambler's Ruin process represented by the transition diagram below.



Define

$$h_x = \mathbb{P}(\text{process reaches state } N \mid \text{start from state } x), \quad \text{for } x = 0, 1, 2, \ldots, N.$$

(a) Using the theory of second-order difference equations, show that

$$h_x = \frac{2^x - 1}{2^N - 1} \quad \text{for } x = 0, 1, \ldots, N.$$

Marks are awarded for setting out your answer clearly, including giving appropriate headings. (7M)

(b) Let m_x be the expected number of steps taken before absorption, starting from state x, for $x = 0, 1, \ldots, N$. Note that m_x corresponds to the expected number of arrows traversed from state x until the process first reaches state 0 or state N, and $m_0 = m_N = 0$. Again using the theory of second-order difference equations, find an expression for m_x for $x = 0, 1, \ldots, N$. You may use results from part (a) without rewriting them, as long as you label them in your working for part (a) and clearly refer to them here. **Do not re-use notation that you have already used in part (a).** (7M)

(c) Let Y be a random variable such that Y is the **maximum** value attained in the Gambler's Ruin process, starting from state x. For example, in the process that starts at $x = 2$ and has trajectory 2, 3, 2, 1, 0, the maximum value attained is $Y = 3$.
Find $\mathbb{P}(Y = y \mid \text{start from state } x)$ in terms of x and N, for any $y = 0, 1, \ldots, N$. (5H)

[19 marks]

5.(a) Let $Y \sim \text{Poisson}(\lambda)$. Working directly from the probability function of Y, show that the probability generating function (PGF) of Y is

$$G_Y(s) = \mathbb{E}(s^Y) = e^{\lambda(s-1)}\,.$$

(3E)

(b) Let $Y_1, Y_2, \ldots, Y_n$ be independent, such that each $Y_i \sim \text{Poisson}(\lambda)$. Let $T = Y_1 + Y_2 + \ldots + Y_n$. Show that $T \sim \text{Poisson}(n\lambda)$. (3M)

(c) Now let $X_0, X_1, X_2, \ldots$ be a Markov chain with state space $S = \{0, 1, 2, \ldots\}$, such that for any t the conditional distribution of X_{t+1}, given X_t, is

$$[\,X_{t+1} \,|\, X_t\,] \sim \text{Poisson}(X_t)\,.$$

(i) Suppose that $X_0 = 10$. What is the probability of the trajectory $(X_0, X_1, X_2) = (10, 8, 12)$? (2M)

(ii) Suppose that $X_0 = 1$. Find $\mathbb{P}(X_t \leq 1 \text{ for all } t = 0, 1, 2, \ldots)$, the probability that the chain never exceeds 1. (3M)

(iii) The Markov chain $X_0, X_1, X_2, \ldots$ is an example of a named process that we have studied in class. Using previous parts of the question to help you, give the name of the process, and identify any parameters of the process. Use your knowledge of this process to find the probability that the Markov chain ever reaches the state 0, starting from any state $x \in S$. Fully explain all your reasoning. (5H)

[16 marks]

6. Let $\{Z_0, Z_1, Z_2, \ldots\}$ be a branching process, where Z_n denotes the number of individuals born at time n, and $Z_0 = 1$. Let Y be the family size distribution, and suppose the probability generating function of Y is $G(s) = \mathbb{E}(s^Y)$.

Let γ be the probability of eventual extinction in the process, starting from $Z_0 = 1$.

(a) Suppose that there are k individuals alive in a particular generation. Give an expression for the probability of eventual extinction, starting from k individuals, for any $k = 0, 1, 2, \ldots$. Write your answer in terms of γ. (2M)

(b) A theorem we have studied states that if $\{Z_0, Z_1, \ldots\}$ is a Markov chain, and A is any state, then the vector of hitting probabilities for state A is the minimal non-negative solution to the appropriate first-step analysis equations. In this question, the Markov chain is the branching process $\{Z_0, Z_1, \ldots\}$, and we are interested in the probability of eventual extinction starting from $Z_0 = 1$ individual. Using the theorem stated, together with part (a), prove that the probability of eventual extinction γ is the minimal non-negative solution to the equation $G(s) = s$. (6H)

[8 marks]

7. **Disease model.**

A school contains n children. On day t, let X_t of the children be absent with sickness. The model is described as follows:

- If a child is absent on day t, he or she will be absent again on day $t+1$ with probability p.
- If a child is **present** on day t, he or she will be **absent** on day $t+1$ with probability a.
- All children are independent.

According to this model, $X_1, X_2, \ldots$ is a Markov chain, such that:

$$\begin{aligned} X_{t+1} &= U_{t+1} + W_{t+1}, \\ \text{where} \quad [\,U_{t+1} \,|\, X_t\,] &\sim \text{Binomial}(X_t,\ p); \\ [\,W_{t+1} \,|\, X_t\,] &\sim \text{Binomial}(n - X_t,\ a); \\ U_{t+1},\ W_{t+1} &\ \text{are independent, given } X_t. \end{aligned}$$

Here, U_{t+1} is the number of children who have remained ill from day t to day $t+1$, and W_{t+1} is the number of children who have a new bout of illness on day $t+1$.

Assume that $0 < p < 1$ and $0 < a < 1$.

(a) In terms of n, p, and a, find $\mathbb{P}(X_{t+1} = 0 \,|\, X_t = 5)$ and $\mathbb{P}(X_{t+1} = 1 \,|\, X_t = 5)$. (4M)

(b) Does the Markov chain $\{X_t\}$ converge to an equilibrium distribution as $t \to \infty$? Explain why or why not. (3M)

(c) For any random variable $Y \sim \text{Binomial}(m, \beta)$, show that the probability generating function of Y is

$$G_Y(s) = \mathbb{E}(s^Y) = (\beta s + 1 - \beta)^m \ .$$

Work directly from the probability function of Y, and show your working. (3E)

(d) In the disease model above, show that

$$\mathbb{E}\left(s^{X_{t+1}} \,|\, X_t\right) = (as + 1 - a)^n \left\{ \frac{ps + 1 - p}{as + 1 - a} \right\}^{X_t} .$$

(4M)

(e) Assume that $X_t \sim \text{Binomial}(n, \pi)$ for some $0 < \pi < 1$. Find the distribution of X_{t+1}, specifying the distribution name and all parameters. Hence find an equilibrium distribution for the Markov chain $\{X_t\}$. (6H)

[20 marks]

ATTACHMENT

1. **Discrete Probability Distributions**

Distribution	$\mathbb{P}(X = x)$	$\mathbb{E}(X)$	PGF, $\mathbb{E}(s^X)$
Geometric(p)	pq^x (where $q = 1 - p$), for $x = 0, 1, 2, \ldots$	$\frac{q}{p}$	$\frac{p}{1 - qs}$
	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	$(ps + q)^n$
	Number of successes in n independent trials, each with $\mathbb{P}(\text{success}) = p$.		
Poisson(λ)	$\frac{\lambda^x}{x!} e^{-\lambda}$ for $x = 0, 1, 2, \ldots$	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:	$\mathbb{E}(X) = G'_X(1)$	$\mathbb{E}\left\{X(X-1)\ldots(X-k+1)\right\} = G_X^{(k)}(1)$
Probabilities:	$\mathbb{P}(X = n) = \frac{1}{n!} G_X^{(n)}(0)$	

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

Finite sum: $\sum_{x=0}^{n} r^x = \frac{1 - r^{n+1}}{1 - r}$ 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 $\lambda \in \mathbb{R}$, $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$.