THE UNIVERSITY OF AUCKLAND

SECOND SEMESTER, 2014
Campus: City

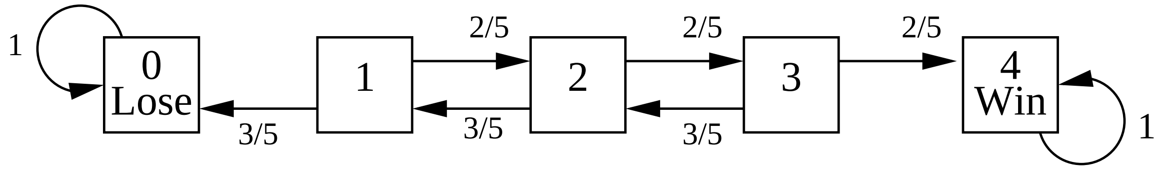
STATISTICS

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. A certain game corresponds to a Markov chain $\{X_0, X_1, X_2, \ldots\}$, following the transition diagram below.



Define the following events:

$$W = \{\text{process ever reaches state } 4\}\,;$$
$$D = \{\text{process ever reaches state } 1\}\,;$$
$$S_x = \{\text{process starts in state } x\} \quad \text{for } x = 0, 1, 2, 3, 4.$$

Define the following sets of hitting probabilities:

$$h_x = \mathbb{P}(W \mid S_x) \quad \text{for } x = 0, 1, 2, 3, 4\,;$$
$$d_x = \mathbb{P}(D \mid S_x) \quad \text{for } x = 2, 3, 4\,.$$

(a) Using first-step analysis, show that $d_2 = \frac{15}{19}$. You should show all working and equations. (3E)

(b) Write down the first-step analysis equations needed to find the vector $\boldsymbol{h} = (h_0, h_1, h_2, h_3, h_4)$. You do *not* need to solve the equations. (4E)

(c) The solution to the equations in part (b) is:

$$\boldsymbol{h} = \left(0\,,\ \tfrac{8}{65}\,,\ \tfrac{20}{65}\,,\ \tfrac{38}{65}\,,\ 1\right)\,.$$

Using this, and any other results you need from parts (a) and (b), find $\mathbb{P}(D \mid S_2 \cap W)$. Show all your working. Say which probability is larger, out of $\mathbb{P}(D \mid S_2 \cap W)$ and $\mathbb{P}(D \mid S_2)$, and briefly comment on why you would expect this to be the case. (5M)

(d) Write down all communicating classes of the Markov chain $\{X_0, X_1, X_2, \ldots\}$ denoted by the diagram above. For each class, say whether or not it is closed. (2E)

(e) Does the Markov chain $\{X_t\}$ converge to an equilibrium distribution that does not depend upon its start state as $t \to \infty$? Explain why or why not. (2E)

[16 marks]

2. 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 that the probability generating function (PGF) of Y is:

$$G(s) = \mathbb{E}(s^Y) = \frac{1}{1 + \mu - \mu s},$$

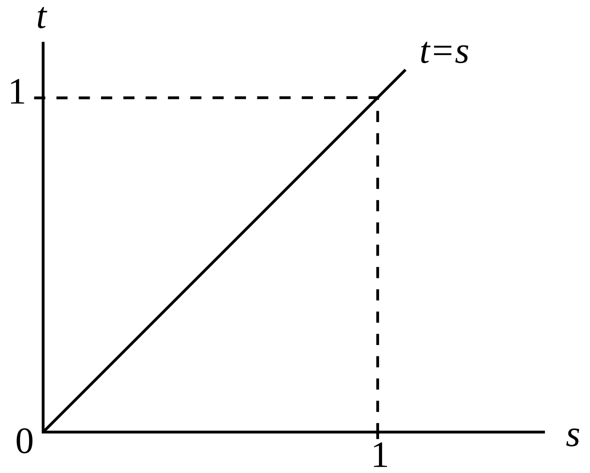
where $\mu > 0$ is a parameter.

(a) Using the fact that $\mathbb{E}(Y) = G'(1)$, show that $\mathbb{E}(Y) = \mu$. (2M)

(b) Let γ be the probability of eventual extinction. Working directly from the PGF of Y, find an expression for γ in terms of μ. Give your answer for the three cases (i) $\mu < 1$; (ii) $\mu = 1$; and (iii) $\mu > 1$. (5M)

(c) The diagram below shows a graph of $t = s$ for $0 \le s \le 1$. Make two copies of the diagram, one for the case $\mu < 1$, and the other for the case $\mu > 1$. For each copy, mark on it the following features:

(i) the curve $t = G(s)$;
(ii) the probability of eventual extinction, γ;
(iii) the mean, μ;
(iv) $\mathbb{P}(Y = 0)$. (4M)



(d) Now suppose that $\mu = 1$. Find $\mathbb{P}(Z_1 = 0)$ and $\mathbb{P}(Z_2 = 0)$. (2E)

(e) Continue to suppose that $\mu = 1$. Let $G_n(s)$ be the probability generating function of Z_n, for $n = 1, 2, \ldots$. Prove by mathematical induction that:

$$G_n(0) = \frac{n}{n+1} \quad \text{for } n = 1, 2, 3, \ldots. \qquad (\star)$$

You may assume that $G_{n+1}(s) = G\Big(G_n(s)\Big)$ for $n = 1, 2, \ldots$. Marks are awarded for setting out your answer clearly. (5M)

(f) Continue to suppose that $\mu = 1$. Using results from earlier parts of the question, find the probability that the population first goes extinct at generation $n = 10$. (2M)

[20 marks]

3. Let $Y_1, Y_2, \ldots$ be independent random variables with $\mathbb{P}(Y_t = 1) = \frac{2}{5}$ and $\mathbb{P}(Y_t = -1) = \frac{3}{5}$. Let (X_t) be a random walk process with $X_0 = 100$ and

$$X_t = 100 + \sum_{s=1}^{t} Y_s \quad \text{for } t = 1, 2, 3, \ldots$$

(a) Find the unique value of λ for which

$$M_t = X_t + \lambda t$$

is a martingale with respect to (X_t).

(b) Using the value of λ found in part (a), find $\mathbb{E}\left[M_{500}\right]$.

(c) Let T be the first time t with $X_t = 0$. Explain why T is a stopping time with respect to (X_t).

(d) Find $\mathbb{E}\left[T\right]$. You may assume that $\mathbb{E}\left[T\right] < \infty$.

[20 marks]

4. Three Markov chains $\{X_0, X_1, X_2, \ldots\}$ are defined below by their transition matrices $\mathcal{P}$. The (i, j) element of $\mathcal{P}$ is written as p_{ij}.

A. Voter process or **Gene spread model:** $p_{ij} = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$ for $i, j \in \{0, 1, \ldots, N\}$.

B. Two-armed bandit process: $\mathcal{P} = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix}$ where $\alpha, \beta \in (0, 1)$.

C. Random walk on the integers: $p_{ij} = \begin{cases} p & \text{if } j = i+1, \\ 1-p & \text{if } j = i-1, \\ 0 & \text{otherwise;} \end{cases}$ where $0 < p < 1$ and $i, j \in \mathbb{Z}$.

Three possible questions of interest about Markov chains are below:

Q1. Hitting probability for state 0. **Q2.** Expected time to absorption. **Q3.** Equilibrium distribution.

For each of processes A, B, and C, say whether each of questions Q1, Q2, and Q3 is of interest for the process. Briefly explain your answers.

[6 marks]

5. The Negative Binomial distribution with parameters k and p is defined to be the sum of k independent Geometric(p) random variables. That is, if $Y_i \sim \text{Geometric}(p)$ for all $i = 1, \ldots, k$, and $Y_1, \ldots, Y_k$ are independent, then $N = Y_1 + \ldots + Y_k$ has the Negative Binomial distribution with parameters k and p. We write $N \sim \text{NegBin}(k, p)$. Here, $k \in \mathbb{N}$, and $0 < p < 1$.

(a) Suppose $Y \sim \text{Geometric}(p)$, so $\mathbb{P}(Y = y) = p(1-p)^y$ for $y = 0, 1, 2, \ldots$ Working directly from the probability function of Y, show that the probability generating function (PGF) of Y is:

$$G_Y(t) = \mathbb{E}\left(t^Y\right) = \frac{p}{1-(1-p)t},$$

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

(b) Let $N \sim \text{NegBin}(k, p)$. Using the formulation $N = Y_1 + \ldots + Y_k$ where $Y_1, \ldots, Y_k$ are independent Geometric(p) random variables, show that the PGF of N is:

$$G_N(s) = \mathbb{E}\left(s^N\right) = \left\{\frac{p}{1-(1-p)s}\right\}^k,$$

and state the range of values of s for which this expression is valid. (2E)

(c) Now let X and Y be random variables defined as follows:

$$Y \sim \text{Geometric}\left(\tfrac{2}{5}\right), \qquad [\,X \mid Y\,] \sim \text{NegBin}\left(Y+1,\ \tfrac{2}{3}\right).$$

Show that the PGF of X is: $\quad G_X(s) = \mathbb{E}(s^X) = \dfrac{4}{9-5s},$

and hence name the distribution of X, with parameters. Show all your working. (6H)

(d) Let $Z = X + Y$. Show that $Z \sim \text{Geometric}(\frac{4}{15})$. Show all your working. (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 $Y \sim \text{Geometric}(p = 0.5)$ be the family size distribution.

   Define $T^{(n)} = 1 + Z_1 + Z_2 + \ldots + Z_n$ to be the **total progeny up to generation n**, starting from a single ancestor, for $n = 1, 2, \ldots$. You may assume that $T^{(n)}$ satisfies the following recursive relationship:

$$T^{(n)} = 1 + T_1^{(n-1)} + T_2^{(n-1)} + \ldots + T_Y^{(n-1)} \quad \text{for } n = 2, 3, \ldots ,$$

   where the random variables $T_1^{(n-1)}, T_2^{(n-1)}, \ldots, T_Y^{(n-1)}$ are independent of each other and of Y, and each have the same distribution as $T^{(n-1)}$.

   Define T to be the **total progeny over all generations**. That is: $T = \lim_{n\to\infty} T^{(n)}$.

   Define the following probability generating functions:

$$G(s) = \mathbb{E}\left(s^Y\right) ; \qquad H_n(s) = \mathbb{E}\left(s^{T^{(n)}}\right) \text{ for } n = 1, 2, \ldots ; \qquad H(s) = \mathbb{E}\left(s^T\right) .$$

   You may assume that, for $Y \sim \text{Geometric}(p = 0.5)$, the probability generating function is

$$G(s) = \mathbb{E}\left(s^Y\right) = \frac{1}{2 - s} ,$$

   and that all the PGFs listed above exist and are continuous for $-1 < s < 1$.

   (a) Using the expression $T^{(n)} = 1 + T_1^{(n-1)} + T_2^{(n-1)} + \ldots + T_Y^{(n-1)}$ for random Y, show that

$$H_n(s) = \frac{s}{2 - H_{n-1}(s)} \quad \text{for } n = 2, 3, \ldots \qquad (\star)$$

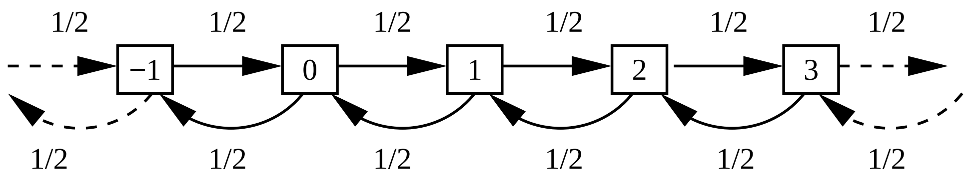
   Show all your working. (6H)

   (b) Using the expression $(\star)$, show that

$$H(s) = 1 - \sqrt{1 - s} .$$

   Justify all steps in your working. You may assume that $H(s) = \lim_{n\to\infty} H_n(s)$ for $-1 < s < 1$. (4M)

   Now suppose that $\{X_0, X_1, X_2, \ldots\}$ is a symmetric random walk on the integers, following the transition diagram below.



   (c) Let W 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 $W = 2$ steps are taken to return from 0 to 0 again. Let $J(s) = \mathbb{E}\left(s^W\right)$ be the probability generating function of W. Show that

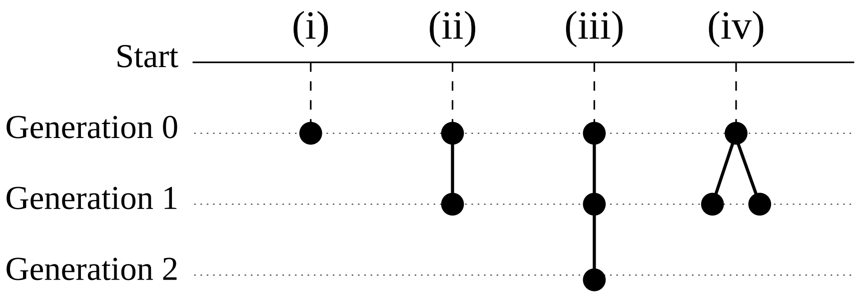
$$J(s) = 1 - \sqrt{1 - s^2} .$$

   (6H)

(d) Using the PGFs shown in parts (b) and (c), specify the exact relationship between the random variables T and W. Hence explain why and how you might expect every **completed graph of a branching process** to correspond to a **path in a random walk.** Using the symbol A to indicate a step **away** from 0 in the random walk, and the symbol B to indicate a step **back** towards 0 in the random walk, suggest paths in the random walk that correspond to the four branching process graphs shown in (i), (ii), (iii), and (iv).

Note: You are *not* expected to prove that your suggested paths are correct. Marks are awarded for suggesting a suitable way of matching paths to graphs.

[**Hint:** the answer to (i) is AB.] (6H)



[22 marks]

ATTACHMENT

1. **Discrete Probability Distributions**

Distribution	$\mathbb{P}(X = x)$	$\mathbb{E}(X)$	$\mathrm{Var}(X)$	PGF, $\mathbb{E}(s^X)$
Geometric(p)	pq^x (where $q = 1 - p$), for $x = 0, 1, 2, \ldots$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1 - qs}$
	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	npq	$(ps + q)^n$
	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	λ	$e^{\lambda(s-1)}$

2. **Uniform Distribution:** $X \sim \mathrm{Uniform}(a, b)$.
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 $\lambda \in \mathbb{R}$, $\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$.