THE UNIVERSITY OF AUCKLAND

SECOND SEMESTER, 2013
Campus: City

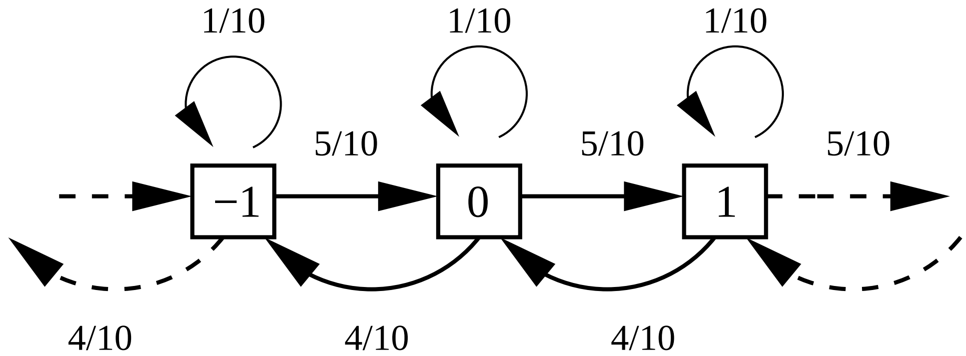
STATISTICS

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. 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:

$$H_U(s) = \frac{10 - s \pm \sqrt{100 - 20s - 79s^2}}{8s}.$$ (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 $H_U(s) = \dfrac{10 - s - \sqrt{100 - 20s - 79s^2}}{8s}$, find whether U is a defective random variable. Hence state the probability that the process ever 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, where

$$H_V(s) = \frac{10 - s - \sqrt{100 - 20s - 79s^2}}{10s}.$$

What is the probability that the process ever 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 $G(s) = \mathbb{E}\left(s^T\right)$ be the probability generating function of T. Using $G(s)$, or otherwise, find the probability that the process ever returns to state 0, starting from state 0. (3M)

[14 marks]

2. Let (X_t) be a simple random walk on $\{0, \ldots, N\}$, with each step being equally likely to be up or down, independently of previous steps. Let T be the number of steps until the walk first reaches either 0 or N, and let $t_k = \mathbb{E}\,(T \mid X_0 = k)$.

   (a) Explain why $t_k < \infty$ for $k = 0, \ldots, N$, i.e. why $\mathbb{E}(T) < \infty$ regardless of where the walk starts. (4)

   (b) Write down a difference equation satisfied by (t_k), including any boundary conditions. (4)

   (c) Solve your equation from (b) to obtain an expression for t_k. (12)

[20 marks]

3. A maker of chocolate bars puts a discount voucher inside the wrapper of every bar. The voucher is worth \$1, \$2, or \$3, with equal probability. Vouchers are coloured yellow, orange, or red, according to their value. The \$1 vouchers are yellow, the \$2 vouchers are orange, and the \$3 vouchers are red.

Several possible transition matrices describing different Markov chains are listed below. Where the pattern repeats indefinitely, this is indicated by dots, as in $[\cdots]$.

$$\textbf{A.} \quad P_A = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad \textbf{B.} \quad P_B = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix} \qquad \textbf{C.} \quad P_C = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix}$$

$$\textbf{D.} \quad P_D = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & \cdots \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 & 0 & \cdots \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 & \cdots \\ \frac{2}{3} & 0 & 0 & 0 & \frac{1}{3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix} \qquad \textbf{E.} \quad P_E = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 & 0 & \cdots \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & \cdots \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 & \cdots \\ 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{pmatrix}$$

Three different Markov chains $\{X_1, X_2, X_3, \ldots\}$ relating to the chocolate bar vouchers are defined in parts (a), (b), and (c) below. In each case, the chain begins at time 1 when the first chocolate bar is opened, and X_t describes the state of the chain when the t'th chocolate bar is opened. Assume outcomes from all chocolate bars are independent.

For each chain in parts (a), (b), and (c), you should answer the following questions:

(i) Specify the state space, S.

(ii) Select the correct transition matrix, A, B, C, D, or E from those listed above, giving a brief explanation.

(iii) Write down as a vector, $\boldsymbol{\alpha}$, the probability distribution of X_1, where element i of $\boldsymbol{\alpha}$ is $\alpha_i = \mathbb{P}(X_1 = i)$ for states $i \in S$.

(iv) State, with justification, whether or not the chain $\{X_t\}$ converges to an equilibrium distribution as $t \to \infty$.

(v) If the chain does converge to equilibrium as $t \to \infty$, state or calculate the equilibrium distribution, $\boldsymbol{\pi}$.

The chains for which you should answer questions (i) to (v) are described here.

(a) X_t is the **number of different colours** in the vouchers obtained from the first t bars. (4M)

(b) X_t is the **maximum voucher value** (in dollars) that has been attained in the first t bars. (4M)

(c) X_t is the **number of consecutive vouchers of the same colour** obtained when bar t is opened. For example, if bar t is the 3rd red voucher in a row, then $X_t = 3$. If the sequence of colours starting from time 1 is Red, Red, Yellow, Orange, Orange, then the trajectory of the chain for $t = 1, \ldots, 5$ is $X_1 = 1, X_2 = 2, X_3 = 1, X_4 = 1, X_5 = 2$. (7H)

[15 marks]

4. **Queue Model.** People join a queue in a bank according to a Poisson process with rate λ people per hour. This means that the time in hours measured from any instant until a new person **arrives** is $X \sim \text{Exponential}(\lambda)$.

If there is at least one person in the queue, then the time in hours measured from any instant until the next person **leaves** the queue is $Y \sim \text{Exponential}(\mu)$.

We begin measuring X and Y at the beginning of the process. Subsequently, we begin new measurements of X and Y whenever the system changes state. Assume that X and Y are independent, and that their distributions remain the same whenever we begin a new measurement (memoryless property).

(a) For $X \sim \text{Exponential}(\lambda)$, the probability density function of X is $f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$. Show by integration that

$$\mathbb{P}(X > x) = e^{-\lambda x} \text{ for any } x > 0.$$

(2M)

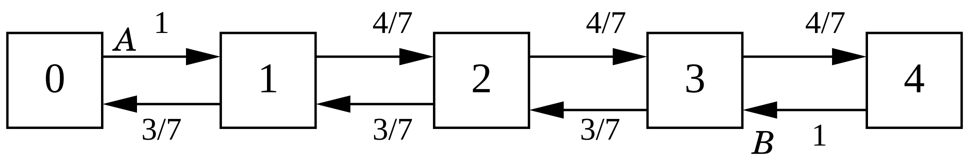
(b) For any $X \sim \text{Exponential}(\lambda)$ and $Y \sim \text{Exponential}(\mu)$, such that X and Y are independent, show that

$$\mathbb{P}(X > Y) = \frac{\mu}{\lambda + \mu}.$$

Show all working and notation. (5M)

In the bank, the arrival rate is $\lambda = 8$ people per hour. For security reasons, no more than 4 people are allowed in the queue at any time. Any additional arrivals will be turned away.

The transition diagram for the queue is given below, where the *state* corresponds to the number of people in the queue.



(c) Using the diagram above, and the fact that $\lambda = 8$, as well as your answer to previous parts of the question, find the value of μ. (3M)

(d) You are given the information that the expected time spent on arrow A is $1/8$ hours, the expected time spent on arrow B is $1/6$ hours, and the expected time spent on all other arrows is $1/14$ hours. Time is only spent on arrows, not in boxes. The queue is currently in state 4, and the bank staff want to know the expected time before it first reaches state 0, when they can have a break. Using the information given, define a suitable notation and write down a system of equations that can be solved to answer this question. Do *not* solve the equations. You should clearly indicate which quantity corresponds to the required answer. (4M)

(e) Write down the transition matrix, P, for the Markov chain shown in the diagram above. Also identify all communicating classes, and state whether or not each class is closed. (5E)

(f) Does the Markov chain $\{X_t\}$ represented by P converge to an equilibrium distribution that does not depend upon start state as $t \to \infty$? Explain why or why not. (3E)

[22 marks]

5. Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed discrete random variables, with common probability generating function $G_X(s)$. Let N be a discrete random variable with probability generating function $G_N(s)$, where N is independent of $X_1, X_2, \ldots$.

   Let T be the randomly stopped sum of the X_i's, such that

$$T = X_1 + X_2 + \ldots + X_N \,,$$

   and let $H(s)$ be the probability generating function of T.

   (a) Show that $H(s) = G_N\Big(G_X(s)\Big)$. (4M)

   (b) Let $Y \sim \text{Geometric}(p)$, with probability function $\mathbb{P}(Y = y) = pq^y$ for $y = 0, 1, 2, \ldots$, where $q = 1 - p$ and where $0 < p < 1$. Working directly from the probability function of Y, show that the PGF of Y is

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

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

   Now 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 $Y \sim \text{Geometric}(p)$ be the family size distribution, and let $G(s) = \mathbb{E}\left(s^Y\right)$ be the PGF of Y.

   Define $T^{(n)}$ to be the **total progeny up to generation *n***, starting from a single ancestor, for $n = 1, 2, \ldots$. That is,

$$T^{(n)} = 1 + Z_1 + Z_2 + \ldots + Z_n.$$

   Further, define the overall total progeny over all generations to be

$$T = \lim_{n\to\infty} T^{(n)} = 1 + Z_1 + Z_2 + \ldots$$

   Let $H_n(s) = \mathbb{E}\left(s^{T^{(n)}}\right)$ be the PGF of $T^{(n)}$ for $n = 1, 2, \ldots$, and let $H(s) = \mathbb{E}\left(s^T\right)$ be the PGF of T.

   (c) Show that

$$H_{n+1}(s) = s\,G\Big(H_n(s)\Big) \quad \text{for } n = 1, 2, \ldots \qquad (\star)$$

   Marks will be awarded for the clarity and completeness of your explanation. (5H)

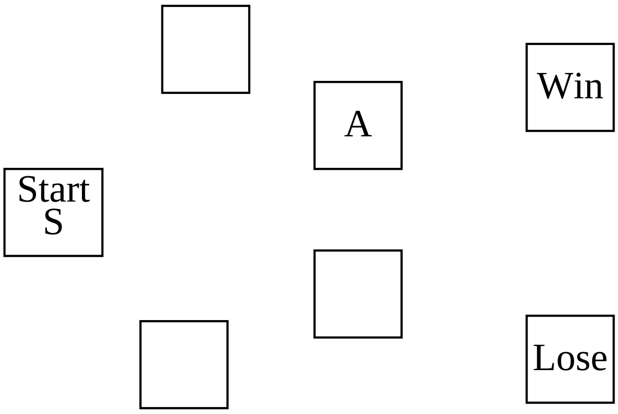
   (d) Because $G(\cdot)$ is continuous, we can take the limit as $n \to \infty$ of both sides of equation $(\star)$ to obtain

$$H(s) = s\,G\Big(H(s)\Big).$$

   Using this, find an expression for $H(s)$. Hence say whether the total progeny T is defective in the two cases (i) $p = 0.5$, and (ii) $p = 0.4$. Deduce the probability of ultimate extinction, γ, for the cases that $p = 0.5$ and $p = 0.4$. (8H)

[20 marks]

6. Consider a Markov chain that involves: one Start state, S; two finish states, *Win* and *Lose*; an intermediate state A; and otherwise *any* combination of states, arrows, and probabilities. An example of a suitable transition diagram is given below, with arrows and probabilities not shown.



The sample space is $\Omega = \{$ all paths starting in state S and finishing in state *Win* or *Lose* $\}$.
Define the event W to be:

$$W = \{\text{process finishes in state } \textit{Win}\}.$$

Define the random variable N to be the number of times the process enters state A, starting at state *Start* and finishing when the process reaches either state *Win* or state *Lose*. Assume that $\mathbb{P}(N = n) > 0$ for all $n = 0, 1, 2, \ldots$.

(a) Suppose that $\mathbb{P}(W \mid N \geq n)$ is constant for all integers $n \geq 1$. Specifically, suppose that $\mathbb{P}(W \mid N \geq n) = \alpha$ for all $n \geq 1$. Prove that this implies that $\mathbb{P}(W \mid N = n) = \alpha$ for all integers $n \geq 1$. (3H)

(b) Define the following notation for any state x:

$$\begin{aligned} p_x &= \mathbb{P}(W \mid \text{start at state } x) \\ m_x &= \mathbb{E}(\text{number of times the process enters state } A \mid \text{start at state } x). \end{aligned}$$

Using this notation, explain why it is true that $\mathbb{P}(W \mid N \geq n) = \alpha$ for all $n \geq 1$, and give the correct expression for α. Marks will be awarded for the clarity of your explanation. [Hint: consider the relevant sample space.] (3H)

(c) Using the notation defined in part (b), find an expression for $\mathbb{E}(N \mid W)$. Explain how you would calculate the quantities involved. (3H)

[9 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 $x=0,1,2,\dots$	$\dfrac{q}{p}$	$\dfrac{q}{p^2}$	$\dfrac{p}{1-qs}$
	Number of failures before the first success in a sequence of independent trials, each with $\mathbb{P}(\text{success})=p$.			
Binomial(n,p)	$\binom{n}{x}p^xq^{n-x}$ (where $q=1-p$), for $x=0,1,2,\dots,n$.	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 $x=0,1,2,\dots$	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:	$\mathbb{E}(X)=G'_X(1)$	$\mathbb{E}\left\{X(X-1)\dots(X-k+1)\right\}=G_X^{(k)}(1)$
Probabilities:	$\mathbb{P}(X=n)=\dfrac{1}{n!}G_X^{(n)}(0)$	

4. **Geometric Series:** $1+r+r^2+r^3+\dots \;=\; \sum_{x=0}^{\infty} r^x \;=\; \dfrac{1}{1-r}$ 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 $n>0$, $(p+q)^n \;=\; \sum_{x=0}^{n}\binom{n}{x}p^xq^{n-x}$.

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