M2S1 - ASSESSED COURSEWORK 1

To be handed in no later than Tuesday, 29th October
For this coursework, MAPLE may NOT be used.

The following series expansions are useful for many calculations relating to discrete probability distributions:

$$\text{GEOMETRIC : for } |z| < 1, \ \sum_{k=0}^{\infty} z^k = \frac{1}{1-z}$$

$$\text{EXPONENTIAL : for real } z, \ \sum_{k=0}^{\infty} \frac{z^k}{k!} = e^z$$

$$\text{BINOMIAL : for } n > 0, \ |z| < 1, \ \sum_{k=0}^{n} \binom{n}{k} z^k = (1+z)^k$$

$$\text{NEGATIVE BINOMIAL : for } n > 0, \ |z| < 1, \ \sum_{k=0}^{\infty} \binom{n+k}{k} z^k = \frac{1}{(1-z)^{n+1}}$$

$$\text{LOGARITHMIC : for } |z| < 1, \ \sum_{k=1}^{\infty} \frac{z^k}{k} = -\log(1-z)$$

(a) Consider a discrete random variable X with range $\mathbb{X} = \{1, 2, 3, ..., 9\}$ and probability mass function (pmf) f_X given by

$$f_X(x) = k_1 \log_{10}\left(1 + \frac{1}{x}\right) \qquad x = 1, 2, 3, ...9$$

for some value of k_1. Show by finding k_1 that this is a valid pmf.

[4 MARKS]

(b) Consider a discrete random variable Y with range $\mathbb{Y} = \{1, 2, 3, ...\}$ and probability mass function (pmf) f_Y given by

$$f_Y(y) = k_2 \frac{\theta^y}{y 3^y} \qquad y = 1, 2, 3, ...$$

for some value of k_2 where $0 < \theta < 3$.

(i) Find the value of k_2

[2 MARKS]

(ii) For the discrete random variable Y, the *probability generating function* (pgf) of Y is denoted G_Y, and is defined, for real value $t \in (1-h, 1+h)$ for some $h > 0$ (i.e. a neighbourhood of 1) by

$$G_Y(t) = \sum_{y=1}^{\infty} t^y f_Y(y).$$

Compute the pgf of Y.

[4 MARKS]

(c) A simple model for the time until an organism succumbs to a viral infection is specified as follows. Suppose that the organism is infected by N virus units. The organism succumbs to the virus when the first of these units successfully infiltrates and establishes itself in the tissue of the organism. The virus units can be considered to be acting independently, and a good model for the infiltration time is given by the following probabilistic specification: define the events E_{ix} and E_x as

$$\begin{aligned} E_{ix} &\equiv \text{“infiltration occurs later than time } x \text{ for virus unit } i\text{”} \\ E_x &\equiv \text{“organism succumbs to the virus later than time } x\text{”} \end{aligned}$$

and define, for $x > 0$,

$$P(E_{ix}) = e^{-\lambda x}$$

for some $\lambda > 0$ for all i. The value $x = 0$ corresponds to the time of infection; we wish to model the time from infection until infiltration.

(i) Suppose that the infection involves $N = n$ virus units. Show that, conditional on the event $F_n \equiv$ “$N = n$”

$$P(E_x|F_n) = P(E_x|N = n) = e^{-n\lambda x}$$

[Note that, for any n,

$$E_x = \bigcap_{i=1}^{n} E_{ix}$$

and that the events $E_{1x}, ..., E_{nx}$ are independent (and thus conditionally independent given $N = n$)]

[2 MARKS]

(ii) Suppose now the number of virus units is not observed, but that it can be modelled probabilistically. Specifically, suppose that N is a discrete random variable, with range $\{0, 1, 2, ...\}$ and pmf

$$f_N(n) = P[N = n] = \frac{\mu^n e^{-\mu}}{n!} \qquad n = 0, 1, 2, ...$$

for some $\mu > 0$.

Using the partition induced by the different possible values of N, the Theorem of Total Probability, and the result from part (i), find $P(E_x)$

[The partition via N uses events $F_n \equiv$ “$N = n$” for $n = 0, 1, 2, ...$]

. *[6 MARKS]*

(iii) Show that

$$P(E_x) \to \theta > 0$$

as $x \to \infty$, for some θ. Comment on the probability that the organism ultimately succombs to infection, and explain the result in terms of a component of the model.

[2 MARKS]

M2S1 - SUPPLEMENTARY 1
(not to be handed in)

1. For which values of the constant c do the following functions define valid probability mass functions for a discrete random variable X, taking values on range $\mathbb{X} = \{1, 2, 3, ...\}$:

$$\text{(a)} \quad f_X(x) = c/2^x \qquad \text{(b)} \quad f_X(x) = c/(x2^x)$$

$$\text{(c)} \quad f_X(x) = c/(x^2) \qquad \text{(d)} \quad f_X(x) = c2^x/x!$$

In each case, calculate (where possible) $\mathrm{P}[X > 1]$ and $\mathrm{P}[X \text{ is even}]$

2. n identical fair coins are tossed. Those that show Heads are tossed again, and the number of Heads obtained on the second set of tosses defines a discrete random variable X. Assuming that all tosses are independent, find the range and probability mass function of X.

3. A point is to be selected from an integer lattice restricted to the triangle with corners at $(1, 1)$, $(n, 1)$ and (n, n) for positive integer n. If all points are equally likely to be selected, find the probability mass functions for the two discrete random variables X and Y corresponding to the $x-$ and $y-$ coordinates of the selected points respectively.

4. A continuous random variable X has pdf given by

$$f_X(x) = c(1-x)x^2 \qquad 0 < x < 1$$

and zero otherwise. Find the value of c, the cdf of X, F_X, and $\mathrm{P}[X > 1/2]$.

5. A function f is defined by

$$f(x) = k/x^{k+1} \qquad x > 1$$

and zero otherwise. For what values of k is f a valid pdf ? Find the cdf of X.

6. A continuous random variable X has pdf given by

$$f_X(x) = \begin{cases} x & 0 < x < 1 \\ 2-x & 1 \le x < 2 \end{cases}$$

and zero otherwise. Sketch f_X, and find the cdf F_X.

7. A continuous random variable X has cdf given by

$$F_X(x) = c(\alpha x^\beta - \beta x^\alpha) \qquad 0 \le x \le 1$$

for constants $1 \le \beta < \alpha$, and zero otherwise.

Find the value of constant c, and evaluate the rth moment of X.

8. A continuous random variable X has cdf given by

$$F_X(x) = \frac{2\beta x}{\beta^2 + x^2} \qquad 0 \le x \le \beta$$

for constant $\beta > 0$. Find the pdf of X, and show that the expectation of X is

$$\beta(1 - \log 2)$$

9 Consider discrete random variables X and Y with ranges $\{0, 1, 2, ...\}$ and probability mass functions (pmfs) f_X and f_Y respectively. Define discrete random variable S by $S = X + Y$, and suppose that for all x, y

$$P[(X = x) \cap (Y = y)] = P[X = x, Y = y] = P[X = x] P[Y = y]$$

so that X and Y are *independent* random variables. It can be shown, using the Theorem of Total Probability, that the probability mass function of S is given by

$$f_S(s) = \sum_{x=0}^{s} f_X(x) f_Y(s - x) \qquad s = 0, 1, 2, ...$$

The *probability generating function (pgf)*, denoted G_Z, of a discrete random variable Z with probability mass function f_Z and range $\{0, 1, 2, ...\}$ is defined uniquely for f_Z (when the sum is convergent) by

$$G_Z(t) = \sum_{z=0}^{\infty} t^z f_Z(z)$$

Show for the variables defined above that $G_S(t) = G_X(t) G_Y(t)$.

10. A study of the preferences of A-Level Mathematics students is to be carried out. In an attempt to minimize bias, a list of schools with A-Level Mathematics classes comprising precisely n students is identified, and each school is required to ask each student in the class.

"Which topic do you prefer: Statistics or Mechanics ?"

It is to be assumed that each student answers this question "Statistics" with probability θ $(0 < \theta < 1)$, and that students' answers are mutually independent. For school i that returns a complete response (that is, x "Statistics" answers and $n - x$ "Mechanics" answers, for some x) , the total number of "Statistics" answers is a random variable, S_i say, where the pmf of S_i is given by

$$f_{S_i}(x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \qquad x = 0, 1, 2, ..., n$$

(i) Find the pgf of S_i.

The number of schools that provide a complete response to the survey is also a non-negative random variable, R say, where the probability distribution of R is well-modelled by the pmf

$$f_R(x) = \frac{e^{-\lambda} \lambda^x}{x!} \qquad x = 0, 1, 2, ...$$

for some constant parameter $\lambda > 0$. Incomplete responses are discarded.

(ii) Using the results in (i) and any results stated in lectures (or from M1S), and **given** that $R = r$, find the (conditional) pmf of the random variable S as defined as

$$S = \sum_{i=1}^{r} S_i$$

(iii) Find the (unconditional) pgf of S, G_S, if R is **not** observed.