Full marks may be obtained for complete answers to **five** questions, of which no more than two may be from Section A. Credit will only be given for the best **five** answers.

Some useful Formulae

1) For any two events A and B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$
$$P(A \cap B) = P(A|B)\, P(B),$$
$$P(A \cap \overline{B}) = P(A) - P(A \cap B).$$

2) For three events A, B and C

$$P\{A \cap (B \cup C)\} = P\{(A \cap B) \cup (A \cap C)\},$$

3) If X has a Binomial distribution with parameters n and p

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \qquad (x = 0, 1, \ldots, n)$$

and $E(X) = np, V(X) = np(1-p)$.

4) If X has a Poisson distribution with mean λ

$$P(X = x) = \frac{\lambda^x}{x!} \exp(-\lambda) \qquad (x = 0, 1, \ldots),$$

and $E(X) = \lambda, V(X) = \lambda$.

Moreover, under suitable conditions,

$$P(a \le X \le b) = \Phi(\beta) - \Phi(\alpha),$$

where $\beta = \frac{(b + 0.5 - \lambda)}{\{\lambda\}^{1/2}}, \quad \alpha = \frac{a - 0.5 - \lambda}{\{\lambda\}^{1/2}}$

and $\Phi(z)$ denotes the area to the left of z for a standard Normal distribution.

5) For all $\alpha > 0$

$$\int_0^\infty t^{\alpha-1} e^{-t} dt = \Gamma(\alpha) = (a-1)! = (\alpha - 1)\, \Gamma(\alpha - 1).$$

SECTION A

1. A family with two pre-school children, called Sarah and Jane, will watch a particular afternoon TV programme, A, say, only if the mother (M) agrees to watching the programme and either or both Sarah (S) and Jane (J) agree as well. Let M denote the event that the mother agrees to watching the programme A and define S and J in the same way. It is known that

$P(S) = 0.23$, $P(J) = 0.25$, $P(M) = 0.5$,
$P(S \cap J) = 0.06$, $P(S \cap M) = 0.125$, $P(J \cap M) = 0.12$,
$P(S \cap J \cap M) = 0.029$.

a) Find the probability that

i) the family watches the TV programme; [4 marks]

ii) either Sarah or Jane or both agree to watching the programme but the mother does not. [7 marks]

b) Given that the family watches the TV programme, find the conditional probability that all three, namely the mother, Sarah and Jane agreed to watching the programme.

[4 marks]

c) There is a second TV programme, B, say, which immediately follows the programme A and the family will certainly watch the programme B if they have not previously watched the programme A. The family will however also watch the programme B even when they have already watched the programme A, provided all three, namely Sarah, Jane and the Mother, have agreed to watching the programme A. Given that the family has watched the programme B, find the conditional probability that they have also watched the programme A.

[5 marks]

2. In a certain area, two traffic wardens, *A* and *B*, say, issue tickets for parking violations. The number of tickets issued by *A* in an hour may be modelled by a Poisson distribution with a mean of 5 and that by *B* may also be modelled by a Poisson distribution, but with a mean of 8 tickets per hour.

Find the probability that

a) *A* issues at least one parking ticket in an hour; [2 marks]

b) the combined total number of parking tickets issued by *A* and *B* in one hour is at least 5. [5 marks]

[**N.B.** You may assume without proof the standard result that if the random variables X and Y are independent, X is Poisson with mean λ, and Y is Poisson with mean μ, then $X+Y$ is Poisson with mean $\lambda+\mu$.]

Suppose that *A* and *B* each issue tickets for six hours per day but the tickets are in fact issued for a total of 10 hours per day, and two part-time traffic wardens are employed for the remaining 4 hours. The numbers of tickets issued by each of these two part-time wardens also follow independent Poisson distributions, but each with a mean of 3 tickets per hour, and independent of the numbers of tickets issued by *A* and *B*. Find

c) the probability that more than 125 tickets are issued per day; [6 marks]

d) an integer, k, such that the probability of the number of tickets issued per day being less than or equal to k is no more than 0.01. [7 marks]

[**N.B**. If Z is Normally distributed with mean 0 and variance 1

$$P(Z < -2.3263) = 0.01]$$

3. The lifetime, T, of a component is known to be distributed as Exponential, with probability density function

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & (t > 0), \\ 0 & (t \le 0), \end{cases}$$

where $\lambda > 0$ is a parameter of the distribution.

Show that

a) $R(t) = P(T > t) = e^{-\lambda t}$, for all $t > 0$; [2 marks]

b) $E(T) = \dfrac{1}{\lambda}$. [2 marks]

A system is constructed from three such components, whose lifetime distributions are mutually independent, each being Exponential with probability density function, $f_T(t)$, given above. The system will function provided at least two of the three components have not failed. Let L denote the lifetime of system. Show that

c) $P(L > t) = 3e^{-2\lambda t} - 2e^{-3\lambda t}$ $(t > 0)$; [5 marks]

d) $E(L) = \dfrac{5}{6\lambda}$. [5 marks]

e) A combined system is made by assembling n of these systems, $n > 1$, and the combined system will function provided at least k of the individual systems are working, $1 \le k \le n$. Let C denote the lifetime of the combined system. Assuming that the n system lifetimes are independent and each has the same distribution as L, explain why

$$P(C > t) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j},$$

where $p = P(L > t)$ is as in c) above. [4 marks]

f) If $n = 2$, $k = 1$, $\lambda = 1$ and $t = 1$, verify that $P(C > t)$ equals 0.52. [2 marks]

SECTION B

4. In a study of the gender distribution of children in families with 3 children, the numbers of families with *k* male children were recorded for *k* = 0, 1, 2, 3.

For a randomly selected family with 3 children, let *X* denote the random variable that *k* children are male. Explain the conditions under which *X* follows a Binomial distribution with parameters 3 and *p*, i.e.

$$P(X = k) = \binom{3}{k} p^k (1-p)^{3-k} \qquad (k = 0, 1, 2, 3). \qquad (1)$$

[4 marks]

The following table was constructed from the data collected by K. Pakrasi and A. Halder (1971) in a study of the gender distribution of children in India and it gives the numbers of families with k male children, k = 0, 1, 2, 3, in a sample of 19,788 families with 3 children in urban areas of India.

Table
Number of male children in families with 3 children living in urban areas of India

No of male children (*k*)	**No of families (0_k)**
0	1990
1	7134
2	8034
3	2630
Total	**19788**

Find an estimate, $\hat{p}$, say, of the probability that a child at birth is male, correct to 5 decimal places. [4 marks]

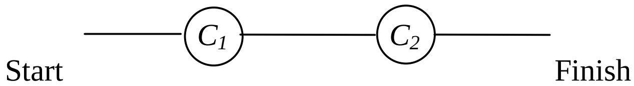
Fit the binomial distribution given in (1) above to this data and comment on the goodness of the fit. [8 marks]

Explain why a Binomial distribution may not be an appropriate probability model for the number of male children in a family, and suggest any special factors that may make the Binomial model particularly inappropriate for families living in the urban areas of India.
[4 marks]

5. The (coded) lifetimes of a batch of components are independently distributed each being Uniform on (0, 1), that is, with probability density function

$$f(x) = \begin{cases} 1 & 0 \le x \le 1, \\ 0 & \text{elsewhere.} \end{cases}$$

a) Suppose that a piece of equipment has been assembled by placing two of these components in a sequence and using the schematic design shown below:

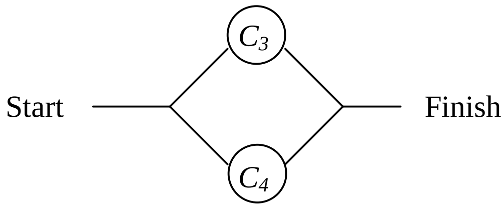


The piece will work only if both the components C_1 and C_2 are working. Let Y denote the lifetime of the piece. Derive the distribution function and the probability density function of Y and show that its hazard function is given by

$$h_Y(y) = \frac{2}{1-y} \qquad 0 < y < 1.$$

[6 marks]

b) Suppose instead that a second piece of equipment has been designed by placing two of these components in parallel and using the scheme shown below:



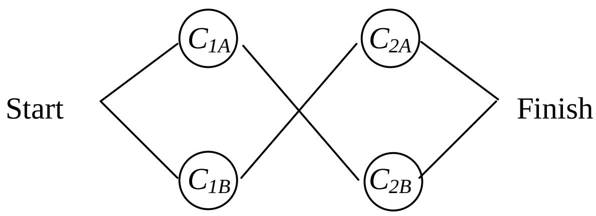
This second piece will fail only if both the components C_3 and C_4 have failed. Let W denote the lifetime of this second piece.

Derive the distribution function and the probability density function of W and show that its hazard function is given by

$$h_W(w) = \frac{2w}{1-w^2} \qquad 0 < w < 1.$$

[6 marks]

c) Suppose now that the design in a) above is modified by duplicating C_1 and C_2 and arranging these components in a schematic design below



Question 5 continued overleaf

Q 5 contd

The redesigned piece of equipment will work if C_{1A} or C_{1B} and C_{2A} or C_{2B} are working, and it may be assumed that the lifetimes of these four components are independently distributed, each being Uniform on (0, 1). Let V denote the lifetime of the redesigned equipment. Show that the distribution function of V is given by

$$F_V(v) = P(V \leq v) = \begin{cases} 0 & v < 0, \\ 1 - (1 - v^2)^2 & 0 < v < 1, \\ 1 & v \geq 1. \end{cases}$$ [4 marks]

Find the hazard function, $h_V(v)$ of V. [2 marks]

Show that

$$h_Y(v) - h_V(v) = \frac{2}{1+v} > 0, \quad 0 < v < 1,$$

and give an interpretation of this last result. [2 marks]

6. A random variable, X, has the probability density function

$$f(x) = \begin{cases} \frac{2x}{r^2} & 0 \leq x \leq r, \\ 0 & \text{elsewhere,} \end{cases}$$

where $r > 0$ is a parameter of the distribution.

a) Find the distribution function, $F(x) = P(X \leq x)$, for all $x \in (-\infty, \infty)$.
[4 marks]

b) Demonstrate that for known values of $F(x)$ such that $0 < F(x) < 1$, $F(x)$ may be inverted to give the quantiles of X by

$$x = r\{F(x)\}^{\frac{1}{2}}.$$ [3 marks]

c) If you are given a random number, u, generated from a Uniform distribution on the interval (0, 1), explain how the result in a) above may be used to generate a value of X. [3 marks]

If u = 0.1, find the corresponding value of X with $r = 1$. [2 marks]

d) If a set of n observations, $x_1, \ldots, x_n$, is available, describe how the result in b) above may be used for obtaining a probability plotting procedure for assessing graphically whether the observations form a random sample from the distribution with density function $f(x)$ and an unknown value of r. [6 marks]

e) Explain also how estimate of the parameter, r, may be obtained from your plot.
[2 marks]

7. A well-maintained piece of equipment can fail ‘at random’ and when it fails it may be repaired without loss of ‘real’ time. If $X = X(t)$ denotes the number of failures of the equipment in a given time interval of t units, $t > 0$, it may be assumed that X has a Poisson distribution,

$$P(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \qquad (k = 0, 1, \ldots,),$$

where $\lambda > 0$ is the average number (rate) of failures of the equipment in a unit time interval.

a) Let T denote the time to next failure, starting immediately after the last failure of the equipment. Show that T has an exponential distribution with probability density function,

$$f_T(t|\lambda) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0, \\ 0 & t < 0. \end{cases}$$

[5 marks]

b) Discuss why the Poisson distribution given above may not be suitable for modelling the number of failures experienced during the debugging process of a software. [4 marks]

Suppose now that the random variable T denotes the time to detect the next fault, starting immediately after the detection of the last fault, during the debugging phase of a software. It has been suggested that T could be treated as an Exponential random variable where the parameter, Λ say, has a Gamma distribution with parameters α and β; that is, conditional on $\Lambda = \lambda$, the probability density function T is given by $f_T(t|\lambda)$ as defined above in a) and the probability density function of Λ is given by $k_\Lambda(\lambda)$, where

$$k_\Lambda(\lambda) = \begin{cases} \{\Gamma(\alpha)\}^{-1} \beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta} & (\lambda \geq 0), \\ 0 & (\lambda < 0), \end{cases}$$

$\alpha > 0$ and $\beta > 0$ are the parameters of the distribution such that α is an integer and $\Gamma(\alpha) = (\alpha - 1)!$.

Let

$$f_T(t) = \int_0^\infty f_T(t|\lambda) k_\Lambda(\lambda) d\lambda$$

denote the unconditional probability density function of T.

c) Show that according to this model the unconditional distribution of T is the Pareto distribution with probability density function

$$f_T(t) = \alpha\beta^\alpha (\beta + t)^{-(\alpha+1)} \qquad (t > 0).$$

[8 marks]

d) By plotting $f_T(t)$, or otherwise, discuss whether a Pareto distribution would provide a suitable model for T, the time to detect the next fault during the debugging phase of a software, starting immediately after the detection of the last fault.

[3 marks]