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 distributions for the candidates' information

Binomial: If a random variable X is Binomial with parameters n and p, $n \geq 1$, $0 < p < 1$,

$B(n,p)$ its probability function is

$$f(x / p) = P(X=x) = {}^{n}C_{x} p^{x} (1-p)^{n-x} \quad (x = 0, 1, ..., n),$$

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

Poisson: If a random variable X is Poisson with mean $m > 0$, its probability

$P(m)$ function is:

$$f(x / m) = P(X=x) = e^{-m} m^{x}/x! \quad (x = 0, 1, ...),$$

and $E(X) = m$, $V(X) = m$.

Normal: If a random variable X is Normally distributed with mean T and variance $ſ^{2}$, its

$N(T, ſ^2)$ probability density function is, *with* $-\infty < T < \infty$, $0 < ſ^2 < \infty$,

$$f(x / T, ſ^2) = (2\Xi ſ^2)^{-½} \exp\{-(2ſ^2)^{-1} (x-T)^2\}, \qquad -\infty < x < \infty.$$

SECTION A

1. Events $B_1, B_2, ..., B_n$ defined on a probability space, $\mathcal{A}$, are mutually exclusive and exhaustive, that is,

$$B_i \cap B_j = \varnothing, \qquad i \neq j,\ i, j = 1, ..., n;$$
$$\bigcup_{j=1}^{n} B_j = \Omega \quad 1.$$

Suppose that A is another event defined on $\mathcal{A}$. Explain why

$$P(A) = \sum_{j=1}^{n} P(A|B_j)\, P(B_j) \quad 2.$$

Hence prove Bayes' rule, that

$$P(B_s|A) = \frac{P(A|B_s)\, P(B_s)}{\sum_{j=1}^{n} P(A|B_j)\, P(B_j)} \quad .3$$

[4 marks]

A multiple choice test offers m alternative answers for each question (of these one is correct). A candidate is required to tick the answer he or she thinks is correct. The proportion of a large number of candidates who know the answer to a particular question is p. If a candidate does not know the answer, he or she guesses and the probability of guessing correctly is $1/m$; if the candidate knows the correct answer, that answer will be given.

For a particular question, evaluate in terms of m and p, the proportion of correct answers and show that the probability that a candidate knows the correct answer and did not guess *given* that the answer provided is correct is

$$\frac{mp}{mp + (1-p)}.$$

[6 marks]

4

A second but different question examines the same issues as the first. If a student knows the answer to the first question, he or she will also be able to answer the second correctly. A student who did not know the answer to the first question but guessed would again have to guess the answer to the second. However, a proportion, q, of those who did guess the first correctly, will recognise the connection between the questions and give the correct answer.

Show that the proportion of candidates who will answer both questions correctly is

$$\frac{m^2 p + mq(1-p) + (1-q)(1-p)}{m^2} \quad 5.$$

[10 marks]

[Hint: You may use without proof the result that for any three events, C, D and E,

$$P(C \cap D / E) = P(C / E)\, P(D / C \cap E)$$

2a) The proportion of faulty items in a batch is alleged to be at most 10%. Eight items chosen at random from the batch are tested, and three prove faulty. Stating any assumptions that you make, evaluate the probability of observing this event, that is, of finding 3 items out of the 8 tested faulty under the hypothesis that the proportion defective is not more than 10%. Hence comment on whether or not your finding casts serious doubt on the hypothesis.

[8 marks]

b) An hotel is considering the installation of an automatic fire protection system based on the use of a large number of water sprinklers and smoke detectors in the ceiling of each room, the detectors being wired independently to the master valve of the water supply. The firm manufacturing the systems claims that there is a probability p that a given detector will be alerted within five minutes of smoke in the room reaching a critical density, independently of other detectors doing so. However, the water flows only if k or more detectors are in an alert state simultaneously.

i) In the hotel ballroom, it is proposed that 150 smoke detectors be installed. If p is 0.6 and k is 80, determine the probability that the sprinklers will be turned on within 5 minutes of smoke reaching the specified density. [8 marks]

ii) Find also the largest value to which k can be set if it is desired to make this probability not less than 0.99. [4 marks]

3. The (coded) lifetimes, X_i $(i = 1, 2, 3, 4)$ of a batch of components, C_i $(i = 1, ..., 4)$ are independently distributed, each being Exponential with probability density function

$$f(x) = \begin{cases} \Sigma \exp(-\lambda x) & x \geq 0, \\ 0 & x < 0, \end{cases}$$

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

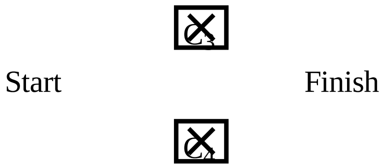
i) Suppose that a piece of equipment has been assembled by placing two of these components in a sequence and using the schematic design shown below:

The piece will work only if both the components C_1 and C_2 are working. Let Y_1 denote the lifetime of the piece. Show that the probability density function of Y_1 is given by

$$h(y) = \begin{cases} 2\Sigma \exp(-2\lambda y) & y \geq 0, \\ 0 & y < 0. \end{cases}$$

and find the meantime to failure, $E(Y)$, of this piece. [10 marks]

ii) 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.

Show that the probability density function of W is given by

$$k(w) = \begin{cases} 2\Sigma \exp(-\Sigma w)\{1-\exp(-\Sigma w)\} & w \geq 0, \\ 0 & w < 0. \end{cases}$$

and find the meantime to failure, $E(W)$, of this piece. [8 marks]

Q3 continued overleaf

Q3 continued

Show that

$$E(W) > E(Y)$$

and explain why this result is deducible directly from the diagrams above. [2 marks]

[N.B. You may use without proof what the standard result that for all $I \geq 1$

$$\int_0^\infty u^{\alpha-1} e^{-u} \, du = \Gamma(\alpha) = (\alpha-1)! \text{ . 6]}$$

SECTION B

4. In a learning experiment, subjects are given a sequence of three tests. At each test, the subject makes either a correct response or an incorrect response. Let S denote a correct response and F an incorrect response.

Make a list of the eight possible sequences, and find the probabilities associated with each of these response sequences in the following two cases:

a) There is no learning effect. The result of the tests are independent with a constant probability p of a correct response at each test. [4 marks]

b) There is a learning effect. The probability of a correct response on the first test is p, but for the second and third tests, it depends upon the result of the immediately preceding test. The probability of a correct response remains unaltered at p if the response to the immediately preceding test was incorrect, but increases to 1 if the response was correct. [6 marks]

Let the random variables X and Y denote the number of correct responses in three tests under hypotheses a) and b) above. Obtain

i) the probability distributions of X and Y, [3 marks]

ii) the expected values $E(X)$ and $E(Y)$ of X and Y and deduce that $E(Y) > E(X)$. [1 mark]

The following frequency distribution was obtained when 60 randomly selected subjects took part in the learning experiment described above:

No. of correct responses	0	1	2	3	Total
No. of subjects	2	8	18	32	60

If $p = ½$, fit the probability distributions obtained under the hypothesis a) and b) above to the data. Compare the observed and expected frequencies, but do not carry out any formal goodness of fit test, and comment on the sources of any perceived differences between these frequencies. [6 marks]

5. A non-negative continuous random variable, X, has probability density function $f(x)$, distribution function $F(x)$ and hazard function $h(x)$. Given the definition

$$h(x) = \frac{f(x)}{1 - F(x)}, 7$$

prove that

$$F(x) = 1 - \exp\{-\int_0^x h(t)dt\} 8. \qquad \text{[5 marks]}$$

In insurance industry, the risk that a large claim is paid out is known to decrease slowly with the claim size. As a probability model for this behaviour, an insurance analyst specifies that the hazard function, $h(x)$, of the claim-size distribution is given by

$$h(x) = (\Lambda/\Pi)\{1 + (x/\Pi)\}^{-1}, \qquad x > 0,$$

where $\Pi > 0, \Lambda > 0$ are parameters of the distribution.

Show that the claim-size, X, say, is postulated to follow a Pareto distribution with probability density function

$$f(x) = (\Lambda/\Pi)\{1 + (x/\Pi)\}^{-(\delta+1)}, \qquad x > 0. \qquad \text{[6 marks]}$$

An alternative probability model for the claim-size is a Weibull distribution, with probability density function

$$\mathrm{f}(x) = (\vartheta/I)(x/I)^{\beta-1} \exp\{-(x/I)^{\vartheta}\}, \qquad x > 0,$$

where $I > 0, \vartheta > 0$ are parameters of the distribution.

Show that the reliability function, $R(x) = P(X > x)$, for a Weibull distribution is given by

$$R(x) = exp\{-(x/I)^{\vartheta}\}, \qquad x > 0. \qquad \text{[5 marks]}$$

Compare the reliability functions of the Weibull and Pareto distributions and explain why Pareto distribution assigns a greater probability to occurrence of large values of X than a Weibull distribution, and that, in this sense, a Pareto distribution is 'more cautious' than a Weibull distribution. [4 marks]

6. The probability density function of a Cauchy random variable, X, with parameter Π, $-\infty < \Pi < \infty$, is given by

$$f(x) = \frac{1}{\pi\left\{1+(x-\theta)^2\right\}}, 9 \qquad -\infty < x < \infty.$$

Deduce that $f(x)$ is symmetric around Π. [1 mark]

Show that the distribution function, $F(x) = P(X \leq x)$, of X is given by

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x-\theta), 10 \qquad -\infty < x < \infty.$$ [6 marks]

Hence demonstrate that the quantiles of X corresponding to known values of $F(x)$ are given by

$$x = \theta + \tan[\pi\{F(x) - \tfrac{1}{2}\}]. \tag{1}$$

[2 marks]

If a set of n observations, $x_1, \ldots, x_n$, is available, give a numerical definition of the empirical cumulative distribution function of the data. Discuss how the result (1) above may be used for constructing a graphical probability plotting procedure for informally checking the hypothesis that the data are a random sample from a Cauchy distribution. If Π is unknown, discuss how an estimate of Π may be constructed from your graphical procedure.

[6 marks]

If you are given a random number, u, generated from a Uniform distribution on the interval (0,1) explain how the result (1) may be used to generate a random number from a Cauchy distribution with parameter Π. If $u = 0.9662$, generate the corresponding value of X *with* $\Pi = 0$. [5 marks]

[N.B. You may use without proof the standard results that

i) $\displaystyle\int \frac{dx}{1+x^2} = \tan^{-1}(x), 11$

ii) $\displaystyle\tan^{-1}(\infty) = \frac{\pi}{2}. 12$]

7. A sequence of events occurring randomly in time follows a non-homogenous Poisson process, $\{X(t)\}$, with mean function $T(t)$ and intensity function $\Sigma(t)$, that is, if $X(t)$ denotes the number of events occurring in $(0,t)$, where $t > 0$,

$$P\{X(t) = k\} \;=\; \exp\{-\mu(t)\} \left\{\frac{\mu(t)^k}{k!}\right\} \qquad (k \;=\; 0,\, 1,\, \ldots) \, 13$$

and $\lambda(t) \;=\; \dfrac{d}{dt}\, \mu(t)$ 14.

I) Let T denote the time taken from the start of the process until the occurrence of the first event and let $h(t)$ denote the hazard function of T. Explain why

$$h(t) = \Sigma(t).$$ [4 marks]

II) For a known integer $i > 0$ and a constant $t_i > 0$, suppose that i events are known to have occurred in the time interval $(0, t_i]$, and let T_{i+1} denote the total time required for $(i+1)$ events to occur. Explain why, for t > t_i,

i) $P\{X(t) = k+i \,/ X(t_i) = i\}$

$$= \exp\left[-\{\mu(t) - \mu(t_i)\}\right] \frac{\{\mu(t) - \mu(t_i)\}^k}{k!} \qquad (k \;=\; 0,1,\ldots) \, 15;$$ [7 marks]

ii) $F(t_{i+1} \,/ t_i) = P\{T_{i+1} \le t_{i+1} \,/ X(t_i) = i\} \;= 1 - exp[\, - \{\, T(t_{i+1}) - T(t_i)\}].$ [5 marks]

[N.B. You may use without proof but should state clearly a standard result concerning the distribution of the sum of two independent Poisson random variables.]

III) As a statistical model for the occurrence of software failures, an analyst has suggested that since frequently occurring errors tend to be experienced first and have the greatest impact in reducing the failure rate, the number of failures during the debugging phase be modelled as a non-homogenous Poisson process with an intensity function, $\Sigma(t)$, that decreases exponentially with respect to the mean function, T(t), of the number of failures already experienced. Thus, in his model,

$$\lambda(t) \;=\; \frac{d}{dt}\, \mu(t) \;=\; \lambda_0 \;\exp\{-\theta\mu(t)\}, 16$$

where t stands for the software execution time, Σ_0 denotes the failure intensity at the beginning of the debugging phase, Π denotes the rate at which $\Sigma(t)$ decreases as faults are isolated and fixed and $T(0) = 0$.

(Q7. continued overleaf)

(Q7. continued)

Demonstrate that for this model:

$$\frac{d}{dt}\left[\exp\{\theta\mu(t)\}\right] = \lambda_0\,\theta\,.17 \qquad \text{[2 marks]}$$

Comment on the usefulness of this model for the number of faults experienced during the debugging phase of a software. [2 marks]