EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)

GRADUATE DIPLOMA, 2001

Statistical Theory and Methods I

Time Allowed: Three Hours

Candidates should answer **FIVE** *questions.*

All questions carry equal marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use silent, cordless, non-programmable electronic calculators.

Where a calculator is used the ***method*** *of calculation should be stated in full.*

Note that $\binom{n}{r}$ *is the same as* ${}^{n}C_{r}$ *and that* ln *stands for* $\log_e$.

This examination paper consists of 10 printed pages. This front cover is page 1. The reverse of the front cover, which is intentionally left blank, is page 2. Question 1 starts on page 3.

There are 8 questions altogether in the paper.

1. Suppose that the discrete random variables X and Y independently follow Poisson distributions such that

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

and

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!}, \qquad y = 0, 1, \ldots .$$

(i) Show that the random variable $X + Y$ also follows a Poisson distribution.

(8)

(ii) Suppose now that $X + Y$ is known to equal z, where z is some non-negative integer. Determine $P(X = x \mid X + Y = z)$ for all possible values of x. What is the *conditional* distribution of X, *given* that $X + Y = z$?

(7)

(iii) When preparing student handouts, Lecturers A and B make typing errors at random, A at a rate of 1.5 errors per page and B at a rate of 0.5 errors per page. A course handout consists of 6 pages typed by Lecturer A and 12 pages typed by Lecturer B. It is found to contain a total of 14 typing errors. Show that the probability that Lecturer A made at least 10 of the mistakes on this handout is 0.279.

(5)

Turn over

2. (a) State *Bayes' Theorem*.

(4)

(b) In a multiple-choice examination, each question is linked with 5 possible answers, of which just 1 is correct.

(i) A particular candidate has probability θ ($0 < \theta < 1$) of knowing the correct answer to a question. If the candidate does not know the correct answer, then he chooses one of the possible answers at random. Show that the probability he answers the question correctly is $\frac{1}{5}(1+4\theta)$.

(4)

(ii) When a candidate gives the correct answer, 1 mark is awarded. When a candidate gives the wrong answer, or no answer at all, a fraction $\frac{1}{n}$ of a mark is *deducted*. Find the value of n that makes the expected mark awarded for one question to the candidate in part (i) equal to θ.

(6)

(iii) If the examination consists of 50 questions and n takes the value calculated in part (ii), verify that a candidate must give 34 correct answers in order to obtain 30 marks.

For the case where θ = 0.75 independently for each question, find *approximately* the probability that this candidate's total mark for the examination is at least 30.

(6)

3. (i) The continuous random variable U follows a gamma distribution with probability density function

$$f(u) = \frac{\theta^{\alpha} u^{\alpha-1} e^{-\theta u}}{\Gamma(\alpha)}, \qquad u > 0 ,$$

where $\alpha > 0$, $\theta > 0$, and $\Gamma(\ )$ denotes the gamma function. Find the expected value and variance of U.

(7)

(ii) The continuous random variables X and Y have joint probability density function

$$f(x, y) = \theta^2 e^{-\theta y} , \qquad y > x > 0 .$$

Draw a sketch to show the region of the (x, y) plane in which this function is defined. Derive the marginal probability density functions of X and Y and use the result of part (i) to deduce their expected values and variances. Find the correlation between X and Y.

(13)

Turn over

4. X is a standard Normal random variable, with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right) , \qquad -\infty < x < \infty .$$

Z is another standard Normal random variable, which is independent of X. The random variable Y is defined by $Y = |Z|$, and so has probability density function

$$f(y) = \frac{\sqrt{2}}{\sqrt{\pi}} \exp\left(-y^2/2\right) , \qquad 0 < y .$$

(i) Find the joint probability density function of the random variables U and V defined by $U = \dfrac{X}{Y}$ and $V = Y$.

(9)

(ii) Show that the marginal probability density function of U is

$$f(u) = \frac{1}{(u^2+1)\pi} , \qquad -\infty < u < \infty .$$

(7)

(iii) This means that U follows the Student's t distribution with 1 degree of freedom. In general, what distributions should X and W follow so that

$$\frac{X}{\sqrt{W/m}}$$

follows a t distribution with m degrees of freedom (m = 1, 2, …)? Explain why the result obtained in parts (i) and (ii) is a special case of this general result.

(4)

5. (i) Suppose that the discrete random variable X is the number of successes in n independent Bernoulli trials with constant success probability θ (where $0 < \theta < 1$). Then X has a binomial distribution, $X \sim B(n, \theta)$. Show that X has moment generating function

$$M_X(t) = \left(1 - \theta + \theta e^t\right)^n \quad .$$

Hence, or otherwise, find the expected value and variance of X.

(9)

(ii) When X is a $B(n, \theta)$ random variable, find the moment generating function of

$$Z = \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}} \quad .$$

Find the limiting form of the moment generating function of Z as $n \to \infty$. [Hint: start by finding the limit of the logarithm of this moment generating function.] By recognising the limiting moment generating function of Z, deduce the limiting distribution of Z as $n \to \infty$.

(11)

6. (i) Consider a (potentially infinite) sequence of independent Bernoulli trials, with constant success probability ϕ (where $0 < \phi < 1$). Let the discrete random variable R be the number of consecutive failures recorded before the first success. Show that R has probability generating function

$$g_R(t) = \frac{\phi}{1 - t(1-\phi)}, \qquad |t| < \frac{1}{1-\phi} \quad .$$

Hence find the expected value and variance of R.

(9)

(ii) The discrete random variable X and the continuous random variable Y are jointly distributed. Marginally, Y has the probability density function

$$f(y) = \theta e^{-\theta y} \quad , \qquad y > 0 \, ,$$

(where $\theta > 0$). Conditional on $Y = y$, X follows a Poisson distribution with expected value y. Show that the marginal distribution of X is the distribution described in part (i), for a particular value of ϕ.

(7)

(iii) Confirm that, for X and Y defined as in part (ii),

$$E(X) = E\{E(X \mid Y)\}$$

$$\text{var}(X) = E\{\text{var}(X \mid Y)\} + \text{var}\{E(X \mid Y)\}$$

(4)

7. (a) Let X be any continuous random variable, and let $F(x)$ be its cumulative distribution function. Suppose that U is a continuous random variable which follows a uniform distribution on the interval (0, 1), and define the new random variable Y by $Y = F^{-1}(U)$, where $F^{-1}(\ )$ is the inverse function of $F(\ )$. By considering the cumulative distribution function of Y, or otherwise, show that Y has the same distribution as X.

(5)

(b) The following values are a random sample of numbers from a uniform distribution on the interval (0, 1).

0.205 0.476 0.879 0.924

Use these values to generate 4 random variates from each of the following distributions, carefully explaining the method you use in each case.

(i) Geometric : $P(X = x) = \left(\frac{1}{2}\right)^x$, $x = 1, 2, \ldots$.

(6)

(ii) Pareto : $f(x) = \frac{18}{x^3}$, $x > 3$.

(5)

(iii) Standard Normal : $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right)$, $-\infty < x < \infty$.

(4)

Turn over

8. Two urns each contain n balls. Of the total of $2n$ balls, n are red and n are black. At each step of a random process, one of the balls in each urn is chosen at random and these two balls are then exchanged (so that each urn continues to contain n balls). Let the states of the system be indexed by the number, r, of red balls in the first urn.

(i) Write down the transition probabilities for a Markov Chain model of this process.

(8)

(ii) Write down a system of equations that must be satisfied by the stationary distribution, $\underline{\Pi} = [\Pi_0, \Pi_1, ..., \Pi_n]$, of this model.

(6)

(iii) For the case $n = 3$, solve the stationary equations in order to find Π_0, Π_1, Π_2 and Π_3.

(6)