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

GRADUATE DIPLOMA, 2000

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* ${}^nC_r$ *and that* ln *stands for* $\log_e$.

This examination paper consists of 8 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. (a) State *Bayes' Theorem*.

(4)

(b) *Helicobacter pylori* is a bacterium that is harmful to human beings. It lodges in the lining of the stomach and intestine, causing gastritis and peptic ulcers. Two combination drug therapies (Treatment A and Treatment B) for eradicating *Helicobacter pylori* have recently been tested. The effectiveness of these treatments depends on whether or not a patient is resistant to the effect of a particular chemical compound (Metronidazole), but a patient's resistance status is not routinely determined before beginning treatment to eradicate *Helicobacter pylori*. It is estimated that Treatment A successfully eradicates *Helicobacter pylori* in 92% of resistant patients and in 87% of non-resistant patients. The corresponding proportions for Treatment B are 75% and 95%.

Suppose that, in a certain population of individuals infected with *Helicobacter pylori*, the proportion who are resistant is θ $(0 < \theta < 1)$. If a patient from this population is unsuccessfully treated with Treatment B, find the probability that this patient is resistant.

(5)

For what values of θ would a greater proportion of patients be successfully treated by Treatment B than by Treatment A?

(5)

Suppose now that $\theta = 0.25$. If 20 patients, selected at random from this population, are treated with Treatment B, find the probability that at least 18 of them will be treated successfully.

(6)

2. The continuous random variables X and Y have the joint probability density function

$$\frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} x^{\alpha-1} y^{\beta-1}(1-x-y)^{\gamma-1}, \qquad 0 < x < 1,\ 0 < y < 1,\ 0 < x + y < 1,$$

where $\alpha > 0$, $\beta > 0$, $\gamma > 0$ and $\Gamma(\ )$ is the gamma function.

(i) Let r and s be non-negative integers. Show that the expected value of $X^r Y^s$ is

$$E(X^r Y^s) = \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \cdot \frac{\Gamma(\beta+s)}{\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha+\beta+\gamma+r+s)}$$

(8)

(ii) Hence determine the expected value and variance of X.

(6)

(iii) Find the correlation between X and Y.

(6)

3. (i) The discrete random variable X follows a hypergeometric distribution, with probability function

$$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, \ldots, n,$$

where N, M and n are positive integers such that $N > M$, $M \geq n$ and $N - M \geq n$.

Show that X has expected value $E(X) = n\dfrac{M}{N}$.

It can also be shown that $E(X(X-1)) = \dfrac{n(n-1)M(M-1)}{N(N-1)}$.

[Note: you do not need to prove this result.]

Use these expressions for $E(X)$ and $E(X(X-1))$ to prove that the variance of X is

$$\text{var}(X) = n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right).$$

(12)

(ii) A bag contains 100 small sweets that are identical apart from their colour; 60 of the sweets are red, 20 are green and 20 are yellow. A child randomly selects 5 sweets from the bag (without replacement). Let X be the number of red sweets and Y be the number of green sweets that the child selects.

Write down the joint probability function of X and Y.

Find the conditional probability function of X, given that $Y = y$ ($y = 0, \ldots, 5$). Hence evaluate the conditional expectation and variance of X, given that $Y = y$.

(8)

4. (i) Suppose that the discrete random variable X has the probability function

$$P(X = x) = (1-\theta)^{x-1}\,\theta \;, \qquad\qquad x = 1, 2, \ldots \,.$$

Show that X has moment generating function

$$M_X(t) = \frac{e^t\theta}{1-e^t(1-\theta)}\,, \qquad\qquad t < -\ln(1-\theta)\,.$$

Hence show that the expected value of X is $\dfrac{1}{\theta}$ and that the variance of X is $\dfrac{1-\theta}{\theta^2}$.

(9)

(ii) Consider a (potentially infinite) sequence of Bernoulli trials, each with success probability θ. Let the discrete random variable Y be the number of trials up to and including that on which the nth success occurs ($n \geq 1$). Show that Y may be decomposed as

$$Y = X_1 + X_2 + \ldots + X_n \;,$$

where $X_1, X_2, \ldots, X_n$ are independent and each has the same distribution as X in part (i).

Hence state an approximation to the distribution of Y for large values of n.

(6)

(iii) In an investigation of Senile Dementia, it is required to recruit 400 elderly people from the general population as 'controls'. The investigators have access to a large pool of elderly people, through social clubs, but they believe that only 80% of all the elderly people in this pool are both suitable for inclusion in the study and willing to take part. Find the approximate probability that they will require to contact at least 520 elderly people in order to recruit 400 suitable 'controls'.

(5)

5. The random variables X and Y independently follow χ^2 distributions with (respectively) r and s degrees of freedom. So, for example, X has the probability density function

$$f_X(x) = \frac{x^{(r-2)/2}\, e^{-x/2}}{2^{r/2}\, \Gamma\left(\dfrac{r}{2}\right)}, \qquad x > 0 \ ,$$

where $r > 0$ and $\Gamma(\ )$ is the gamma function. Define the random variables U and V by

$$U = \frac{\dfrac{X}{r}}{\dfrac{Y}{s}} \qquad \text{and} \qquad V = \frac{Y}{s}\,.$$

(i) Show that U and V have joint probability density function

$$f(u,v) = k\, u^{(r-2)/2}\, v^{(r+s-2)/2}\, e^{-v(ru+s)/2}, \qquad u > 0, v > 0$$

where k is a constant that you should evaluate.

(9)

(ii) Find the marginal probability density function of U, and name this distribution.

(6)

(iii) State which exponential distribution is also a χ^2 distribution. Use your answer to part (ii) to deduce the distribution of $\dfrac{X}{Y}$ when X and Y independently follow the exponential distribution with expected value $1/\theta$.

[Note: you may assume without proof that, if a random variable W has an exponential distribution and c is a positive constant, then cW also has an exponential distribution.]

(5)

6. A random sample of size n is drawn from the uniform distribution over the interval $(-\theta, \theta)$, where $\theta > 0$. The ordered values in the sample are denoted $U_1 \leq U_2 \leq \ldots \leq U_n$.

(i) Derive the joint probability density function of U_1 and U_n.

(6)

(ii) Show that the range $R = U_n - U_1$ has probability density function

$$f(r) = \frac{n(n-1)r^{n-2}(2\theta - r)}{(2\theta)^n}, \qquad 0 < r < 2\theta \ .$$

(8)

(iii) Find the expected value of R, and comment on the possible use of $\frac{1}{2}R$ as an estimator of θ.

(6)

7. (a) Suppose that the continuous random variable X follows the Weibull distribution with probability density function

$$f_X(x) = \alpha\,\theta\, x^{\alpha-1} \exp\left(-\theta x^{\alpha}\right), \qquad x > 0$$

(where $\alpha > 0$ and $\theta > 0$). Find the probability density function of the random variable $Y = \theta X^{\alpha}$, and identify the distribution of Y.

(5)

(b) The following numbers are a random sample of real numbers from a uniform distribution between 0 and 1:

0.398 0.085 0.986 0.538.

Use these values to generate four random variates from each of the following distributions, and explain carefully the method you use in each case:

(i) $\mathrm{P}(X = x) = \dfrac{\dbinom{5}{x}\dbinom{5}{3-x}}{\dbinom{10}{3}}, \qquad x = 0, 1, 2, 3$ (6)

(ii) $f_X(x) = \exp(-x), \qquad x > 0$ (4)

(iii) $f_X(x) = \dfrac{1}{\sqrt{x}} \exp\left(-2\sqrt{x}\right), \qquad x > 0$ (5)

[<u>Hint</u>: for part (b)(iii), you might find it helpful to refer to part (a).]

Turn over

8. The purpose of a dose-response study is to examine the relationship between the dose of a therapeutic agent and its potential toxic effect ("side effect"). Durham, Flournoy and Rosenberger (*Biometrics*, 1997) describe the following sequential scheme for allocating patients to dose levels in a study of this kind.

Each patient is allocated one of the dose levels $d_1, d_2, \ldots, d_N$ where $d_1 < d_2 < \ldots < d_N$. At dose level d_i, a patient experiences toxic effects with probability ϕ_i, where $0 \le \phi_1 \le \phi_2 \le \ldots \le \phi_N \le 1$. Suppose that the jth patient is allocated dose level d_i. If this patient experiences toxic effects, then the $(j+1)$th patient is allocated dose level d_{i-1} (when $i = 2, 3, \ldots, N$) or d_1 again (when $i = 1$). If the jth patient does not experience toxic effects, then the $(j+1)$th patient is randomly allocated either to dose level d_{i+1}, with probability β $(0 < \beta < 0.5)$, or to dose level d_i, with probability $1 - \beta$ (when $i = 1, 2, \ldots, N-1$). When the jth patient is allocated dose level d_N, and does not experience toxic effects, then the $(j+1)$th patient is always allocated dose level d_N.

In this allocation scheme, the dose levels of the successive patients form a Markov Chain, and the states of this chain can be described by the available dose levels $d_1, d_2, \ldots, d_N$. Write down the matrix of one-step transition probabilities for a Markov Chain model of this process.

(6)

Write down a set of equations that must be satisfied by the stationary distribution, $\underline{\Pi}$, of this system. Verify that these equations are satisfied by the probabilities

$$\Pi_{i+1} = \beta \frac{1-\phi_i}{\phi_{i+1}} \Pi_i, \qquad i = 0, 1, ..., N\text{-}1.$$

(8)

Show that, if $\phi_i \le \dfrac{\beta}{1+\beta}$, then $\Pi_{i-1} \le \Pi_i$, and that, if $\phi_i > \dfrac{\beta}{1+\beta}$, then $\Pi_{i+1} < \Pi_i$. Hence find a value for β such that the stationary distribution has a mode as close as possible to the dose level at which the toxicity probability $\phi = 0.1$.

(6)