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

GRADUATE DIPLOMA, 2003

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.*

The notation log *denotes logarithm to base* ***e***.
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* ${}^{n}C_{r}$.

This examination paper consists of 9 printed pages, **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. The discrete random variables X and Y jointly follow a multinomial distribution with

$$P(X = x, Y = y) = \frac{n!}{x!y!(n-x-y)!}\theta_1^x \theta_2^y (1-\theta_1-\theta_2)^{n-x-y},$$

$$x = 0, 1, \ldots, n, \qquad y = 0, 1, \ldots, n-x,$$

where n is a positive integer, $0 < \theta_1 < 1$, $0 < \theta_2 < 1$ and $0 < \theta_1 + \theta_2 < 1$.

(i) Show that, marginally, X follows a binomial distribution and state the parameters of this distribution. State also the distribution of Y.

(8)

(ii) Show that, *conditional* on $Y = y$ (for any possible y), X follows a binomial distribution and state the parameters of this distribution.

(6)

(iii) Suppose that 10 pairs of ordinary fair dice are rolled independently. Let X be the number of pairs of dice on which a double 6 is scored. Let Y be the number of pairs of dice on which no 6 is scored. State the joint distribution of X and Y. Find the expected number of pairs of dice on which a double 6 is scored. Find the expected number of pairs of dice on which a double 6 is scored, *given* that at least one 6 is scored on every pair of dice.

(6)

Turn over

2. (i) (a) The events $E_1, E_2, \ldots, E_n$ partition the sample space, S. Another event A in S has probability $P(A) > 0$. Write down the law of total probability, which expresses $P(A)$ in terms of conditional and unconditional probabilities involving the events $E_1, E_2, \ldots, E_n$. Write down Bayes' Theorem for probabilities of the form $P(E_j \mid A)$.

(4)

(b) The events $F_1, F_2, \ldots, F_m$ also partition S, and $P(E_i \cap F_j) > 0$ for all i and j. Explain why the events $\{E_i \cap F_j, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m\}$ partition S. Prove that

$$P(A) = \sum_{i=1}^{n} \sum_{j=1}^{m} P\left(A \mid E_i \cap F_j\right) P\left(F_j \mid E_i\right) P\left(E_i\right) .$$

(6)

(ii) Haemophilia is a blood condition, in which an essential clotting factor is either partly or completely missing. This causes a person with haemophilia to bleed for longer than normal, which can be fatal. Haemophilia is a genetic disorder, which only arises when passed on from affected parents to their children. Males either suffer from haemophilia or do not. Females may be carriers of the disorder, which means that they do not suffer from haemophilia themselves but can still pass on the genes for haemophilia to their children.

If a (female) carrier of haemophilia and a man who does not suffer from haemophilia have a son, then the boy has probability 0.5 of suffering from haemophilia. If the same couple have a daughter, then the girl has probability 0.5 of being a carrier of haemophilia and probability 0.5 of not being a carrier, in which case she will not be affected at all by the disorder. Children of the same parents inherit or do not inherit the genes for haemophilia independently of one another.

(a) A certain woman knows that her mother was a carrier of haemophilia and that her father did not suffer from haemophilia. By a man who does not suffer from haemophilia, she has three sons and one daughter. None of her sons suffers from haemophilia. Find the conditional probability that this woman is a carrier of haemophilia. Hence find the conditional probability that her daughter is a carrier of haemophilia.

(7)

(b) The daughter marries a man who does not suffer from haemophilia. They have just two children, both girls. Given all the information available about this family, find the conditional probability that at least one of these girls is a carrier of haemophilia.

(3)

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

$$f(x, y) = \begin{cases} 6x, & 0 \le x \le 1, \quad 0 \le y \le 1 - x, \\ 0, & \text{otherwise.} \end{cases}$$

(i) Without carrying out any calculation, explain why X and Y cannot be independent.

(2)

(ii) Show that, for any non-negative integers r and s,

$$E(X^r Y^s) = \frac{6(r+1)!s!}{(r+s+3)!} \ .$$

Hence find the marginal expected values and variances of X and Y, and the correlation between X and Y.

(14)

(iii) For any value w, such that $0 \le w \le 1$, find $P(X + Y \le w)$. Hence write down the cumulative distribution function and probability density function of the random variable $W = X + Y$.

(4)

[You may use the result $\int_0^1 u^a (1-u)^b \, du = \dfrac{a!b!}{(a+b+1)!}$ when a, b are non-negative integers.]

4. A competitor is to shoot at a target point O on a vertical wall. The horizontal and vertical displacements (cm) from O to the point where one of this competitor's bullets hits the target are denoted X and Y respectively. The random variables X and Y follow independent Normal distributions, each with mean 0 and standard deviation σ.

(i) Find the joint probability density function of the random variables R (cm) and ϕ (radians), defined by

$$X = R\cos(\phi),$$
$$Y = R\sin(\phi).$$

(10)

(ii) Hence find the probability density function of the distance (in cm) from O to the point where one of this competitor's bullets hits the target.

(5)

(iii) Find the value k (> 0) such that 50% of the bullets fired by this competitor will lie within a circle of radius $k\sigma$ cm centred at O.

(5)

Turn over

5. (i) For some value $\mu > 0$, the random variable X follows the Poisson distribution with probability function

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

Show that X has moment generating function

$$M_X(t) = \exp\left\{\left(e^t - 1\right)\mu\right\}.$$

Use this result to derive the expected value and variance of X.

(8)

(ii) Using the result of part (i), or otherwise, find the moment generating function of the random variable

$$Z = \frac{X - \mu}{\sqrt{\mu}} .$$

Find the limiting form of this moment generating function as $\mu \to \infty$.

[Hint. Consider taking the logarithm of the moment generating function.]

Using this result, name the limiting distribution of Z.

(8)

(iii) Suppose now that $Y_1, Y_2, \ldots, Y_n$ are independent random variables such that each Y_i follows a Poisson distribution with expected value μ_i. Using the result of part (i), show that

$$Y_1 + Y_2 + \ldots + Y_n$$

also follows a Poisson distribution.

(4)

Turn over

6. Suppose that the continuous random variable U may take only non-negative values and has cumulative distribution function $F(u)$ and probability density function $f(u)$, for $u \geq 0$. The hazard function of U is defined to be

$$h(u) = \frac{f(u)}{1 - F(u)}, \qquad u \geq 0.$$

(i) The random variable W is said to follow a Weibull distribution with parameters $\alpha > 0$ and $\theta > 0$, denoted We(α, θ), if it has probability density function

$$f(w) = \begin{cases} \alpha\theta w^{\theta-1} \exp\left(-\alpha w^{\theta}\right), & w \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Derive the hazard function of W. Write down conditions on α and θ that give (a) a hazard function that is constant for all $w \geq 0$, (b) a hazard function that decreases as w increases.

(6)

(ii) A system consists of two components, C_1 and C_2, connected in series. This means that the system fails as soon as either of the two components fails. Failures occur in the two components independently. Let the random variable X_i be the time to failure of C_i (i = 1, 2), where X_i has cumulative distribution function $F_i(x_i)$ and probability density function $f_i(x_i)$. Let Y be the time to failure of the system. Show that Y has cumulative distribution function

$$G(y) = F_1(y) + F_2(y) - F_1(y)F_2(y), \quad y \geq 0.$$

Deduce that the hazard function of Y is the sum of the hazard functions of X_1 and X_2. Hence show that, if X_i follows a We(α_i, θ) distribution (i = 1, 2), then Y also follows a Weibull distribution.

(8)

(iii) Now suppose that the components C_1 and C_2, described in part (ii), are connected in parallel to create a new system. This system fails only when both components fail. Explain why the time to failure of this system has cumulative distribution function

$$G(y) = F_1(y)F_2(y).$$

In the case where the two components are identical, and therefore have identically distributed times to failure, obtain an expression for the hazard function of the system in terms of the cumulative distribution function and probability density function of an individual component. Show that the hazard for the system, $h(y)$, is no greater than the hazard for an individual component, for all possible values of y.

(6)

7. (i) The following numbers are a random sample of real numbers from a uniform distribution on the range 0 to 1:

0.5398 0.3372 0.9887 0.4920

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

(a) $P(X = x) = \dfrac{\binom{5}{x}\binom{10}{2-x}}{\binom{15}{2}}, \qquad x = 0, 1, 2.$ (6)

(b) $f_X(x) = 3x^2, \qquad 0 \le x \le 1.$ (4)

(ii) A rural bus service starts at town A, stops in towns B, C and D in turn, then terminates in town E. The journey time (minutes) between any two consecutive towns is a $N(9, (½)^2)$ random variable. It may be assumed that the journey times between the four pairs of consecutive towns are independent.

The timetable for this service specifies that the bus will leave A, B, C and D at 11.00, 11.10, 11.20 and 11.30 a.m. respectively, before terminating at E at 11.40 a.m. If the bus reaches B, C or D before it is timetabled to leave that place, then it will wait there and leave at exactly the timetabled time. If the bus reaches a stop after the time when it is due to leave that place, then it will leave again immediately. (It may be assumed that the time required for passengers to get on and off the bus is sufficiently small to be neglected.)

Assuming that the bus leaves A on time, use the uniform random numbers given in part (i), in the order given, to simulate one journey of the bus from A to E. State clearly when the bus arrives at E in your simulation.

Explain how you would use a larger simulation to estimate the expected time at which this service arrives at E and the probability that the service is late arriving at E.

(10)

8. Obesity is an increasing health problem in Western countries, even for young children. A simple model of development suggests that a person who is obese at age i years ($i = 0, 1, \ldots$) has probability θ of also being obese at age $i + 1$ years, while a person who is not obese at age i years has probability ϕ of being obese at age $i + 1$ years, irrespective of that person's condition at previous ages. It is assumed that $0 < \phi < \theta < 1$ and that no baby is obese at birth.

(i) Model this as a Markov Chain and write down its transition matrix.

(3)

(ii) Derive the two-step transition matrix of this Markov Chain. What proportion of children are obese at age 2 years?

(5)

(iii) What proportion of three-year-old children have never been obese, at any age up to and including 3 years?

(2)

(iv) Let p_i denote the proportion of people who are obese at age i years ($i = 0, 1, \ldots$). Find an expression for p_{i+1} in terms of p_i, θ and ϕ. Using an inductive argument, or otherwise, show that

$$p_i = \frac{1-(\theta-\phi)^i}{1-(\theta-\phi)}\phi .$$

(6)

(v) To what limit does p_i tend as i increases? Evaluate this limit for a population in which $\theta = 0.8$ and $\phi = 0.1$. What does this suggest about obesity levels among adults in this population?

(4)