EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

(formerly the Examinations of the Institute of Statisticians)

GRADUATE DIPLOMA, 2006

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 calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

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. (i) 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)

(ii) A communication system transmits binary data (sequences of zeros and ones). In the long run, these symbols are equally frequent. When a zero is to be communicated, the system sends a voltage signal at level +1; when a one is to be communicated, it sends a voltage signal at level –1. Due to noise on the transmission channel, the level of the received signal is a random variable, X, where $X \sim \mathrm{N}(1, \sigma^2)$ when a zero is sent and $X \sim \mathrm{N}(-1, \sigma^2)$ when a one is sent.

(a) The receiver decides that a zero was sent if and only if $X > 0$. Derive an expression for the probability that an error is made at the receiver. Evaluate this probability when $\sigma = ½$.

(10)

(b) In an attempt to reduce the error rate, each digit is now to be transmitted independently three times. The receiver will decide that a zero was sent if and only if at least two of the three voltage values received are greater than 0. Find the probability that an error is made at the receiver when $\sigma = ½$.

(6)

2. (i) Suppose that U is a continuous random variable whose probability density function, f_U, is symmetric about 0. Let F_U denote the cumulative distribution function of U. Define the random variable W by $W = |U|$.

(a) Show that W has cumulative distribution function

$$F_W(w) = 2F_U(w) - 1, \qquad w \geq 0.$$

Deduce the probability density function of W in terms of f_U.

(5)

(b) Suppose now that U follows the Normal distribution with expected value 0 and variance τ^2 (> 0). Find the probability density function of W.

Deduce that $E(W) = \sqrt{\frac{2\tau^2}{\pi}}$. Find also the variance of W.

(11)

(ii) If X and Y are independent and identically distributed random variables, then $E\left(\left|X - Y\right|\right)$ is known as the Gini statistic of their common distribution. Use the results of part (i) to find the Gini statistic of the Normal distribution with expected value μ and variance σ^2 (where $\sigma > 0$).

(4)

3. Snapdragon (*Antirrhinum*) plants can have red, pink or white flowers; the flowers on any individual plant are all the same colour. A snapdragon plant grown from a dihybrid cross has probability ¼ of having red flowers, ½ of having pink flowers and ¼ of having white flowers. A gardener grows 20 snapdragon plants, each having been independently produced from a dihybrid cross. Let the random variable X be the number of these plants that have red flowers and Y the number of them that have pink flowers.

(i) Write down an expression for the joint probability distribution, $P(X = x, Y = y)$. Find the probability that the gardener grows exactly five plants with red flowers *and* exactly ten plants with pink flowers.

(5)

(ii) The random variable X follows a binomial distribution. Without doing any algebra, explain why, and state its parameters.

(4)

(iii) Find the probability that the gardener grows no more than one plant with white flowers.

(4)

(iv) The conditional distribution of X, given $Y = y$ (for any possible value y), is also a binomial distribution. Without doing any algebra, explain why. State the parameters of this conditional distribution.

(3)

(v) Assume that the time of flowering does not vary with colour, and that the first five plants to flower are all pink. Find the probability that at least three of the remaining plants will have red flowers.

(4)

Turn over

4. (i) Suppose that X and Y are independent random variables, each following the chi-squared distribution with one degree of freedom. This distribution has probability density function (pdf)

$$f(w) = \frac{1}{\sqrt{2\pi w}} e^{-w/2}, \qquad w > 0 .$$

Define new random variables U and V as follows:

$$U = \frac{X}{Y}, \qquad V = Y .$$

Obtain the joint pdf of U and V, and hence show that U has pdf $\dfrac{1}{\pi(u+1)\sqrt{u}}$ for $u > 0$.

(14)

(ii) Suppose now that W_1 and W_2 are independent $N(0, \sigma^2)$ random variables, for some $\sigma > 0$. Noting that (i) shows that U follows the $F_{1,1}$ distribution, or otherwise, find the pdf of $\left| \dfrac{W_1}{W_2} \right|$.

(6)

5. (i) The continuous random variable X follows the gamma distribution with probability density function

$$f_X(x) = \frac{\theta^\alpha x^{\alpha-1} e^{-\theta x}}{\Gamma(\alpha)}, \qquad x > 0 .$$

Here α and θ are positive constants and $\Gamma(.)$ denotes the gamma function, defined by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$.

Show that X has the following moment generating function (mgf):

$$M_X(t) = \left(\frac{1}{1-(t/\theta)} \right)^\alpha, \qquad t < \theta .$$

Hence find the expected value and variance of X.

(9)

(ii) Using the result of part (i), or otherwise, find the mgf of the standardised random variable

$$Z = \frac{\theta}{\sqrt{\alpha}} \left(X - \frac{\alpha}{\theta} \right).$$

Now let θ be fixed. Find the limiting form of this mgf as $\alpha \to \infty$. [Hint: consider the logarithm of the mgf.]

Hence give an approximation to the distribution of X for large values of α.

(11)

Turn over

6. A random sample of size n is drawn from the uniform distribution on the interval 0 to θ (where $\theta > 0$). The ordered values in this sample are $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$. The sample range is $U = X_{(n)} - X_{(1)}$.

(i) Write down the probability density function (pdf) of $X_{(j)}$, for $j = 1, 2, \ldots, n$. Hence obtain the expected value and variance of $X_{(j)}$. Deduce the expected value of U.

(9)

(ii) Find the joint pdf of $X_{(1)}$ and $X_{(n)}$. Use it to find $E\left(U^2\right)$ and deduce that the variance of U is $\dfrac{2(n-1)\theta^2}{(n+1)^2(n+2)}$.

(7)

(iii) The statistics $\dfrac{n+1}{n}X_{(n)}$ and $\dfrac{n+1}{n-1}U$ both have expected value θ and are both possible estimators of θ. Find their variances. Which do you think is the better estimator?

(4)

7. (i) Take the following values as a random sample from the uniform distribution on the range 0 to 1.

0.1269 0.2473 0.5107 0.9068

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

(a) $P(X = x) = \left(\frac{2}{3}\right)^{x-1}\left(\frac{1}{3}\right), \quad x = 1, 2, 3, \dots$. (6)

(b) $f_X(x) = \frac{1}{(1+x)^2}, \quad x > 0$. (6)

(ii) A certain production process requires the use of two machines of the same kind. The factory maintains a stock of three of these machines, two of which are in use on the production line while the other one is in store. When one of the machines on the production line develops a fault, it is sent for repair and replaced by the machine from store. After repair, a machine is put in store if the other two machines are still working normally, or put onto the production line if one of the other machines has developed a fault in the meantime.

The time (in hours) for which a machine is in operation on the production line before it develops a fault is an exponential random variable with expected value 100. The process of taking a faulty machine out of production and replacing it with a machine from store can be assumed to be instantaneous. The time (in hours) required to repair a machine is an exponential random variable with expected value 2.5.

At time $t = 0$, all three machines are in working order with machines A and B on the production line and machine C in store.

(a) Use the uniform random numbers given above, in the order given, to simulate the amount of time:

- that elapses till the first fault develops on machine A;
- that elapses till the first fault develops on machine B;
- that elapses between machine C replacing the first machine to fail and the first fault developing on machine C;
- that is required to repair the first machine that fails.

Explain your methods carefully. (6)

(b) In your simulation, show whether or not the first machine to fail is repaired before a fault develops on another machine. (2)

8. (i) Let Y be a discrete random variable with

$$P(Y = y) = (1 - \phi)^{y-1}\phi, \qquad y = 1, 2, \dots ,$$

where $0 < \phi < 1$. For all $k = 1, 2, \dots$, show that

$$P(Y = k + y \mid Y > k) = P(Y = y), \qquad y = 1, 2, \dots .$$

(4)

(ii) In a model of how a certain bank with just one teller (server) operates, the number of customers in the system at time n (including any being served) is denoted by X_n ($n = 0, 1, 2, \dots$). The service times of all customers, in whole numbers of time units, are independent (and independent of the arrivals process), and follow the distribution given in (i). Customers arrive singly, arrivals are possible only at the times 0, 1, 2, … ; the probability of an arrival at any time n is θ, ($0 < \theta < 1$), independently of all other times. Use the result in (i) to justify the claim that X_n is a Markov chain, and write down its transition probabilities.

(6)

(iii) Let $[\pi_0, \pi_1, \pi_2, \dots]$ denote the stationary probabilities of this Markov chain. In the special case where $\phi = ½$ and $\theta = ¼$, show that the stationary probabilities are related by the following equations.

$$\pi_0 = \frac{3}{2}\pi_1, \qquad \pi_1 = \frac{1}{2}\pi_0 + \frac{3}{4}\pi_2, \qquad \pi_j = \frac{1}{4}\pi_{j-1} + \frac{3}{4}\pi_{j+1} \quad (j = 2, 3, ...)$$

Show that these equations are satisfied by the following probabilities.

$$\pi_0 = \frac{1}{2}, \qquad \pi_j = \frac{1}{3^j} \quad (j = 1, 2, ...)$$

(10)