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

GRADUATE DIPLOMA, 2004

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* ${}^nC_r$.

This examination paper consists of 8 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. Suppose that X and Y are independent, identically distributed geometric random variables, each having probability function given by

$$P(W = w) = (1 - \theta)^{w-1}\theta, \qquad w = 1, 2, \dots,$$

where $0 < \theta < 1$.

(i) Derive the distribution of the random variable $X + Y$.

(8)

(ii) Suppose $u \geq 2$ is a fixed integer. Find $P(X = x \mid X + Y = u)$ for all possible values of x. Name this conditional distribution of X.

(7)

(iii) Suppose that two boys, Andy and Bob, attempt the same new computer game. Random procedures built in to this game ensure that there is probability 0.4 of successfully completing Level 1 every time one of the boys attempts it, and that all their attempts are stochastically independent. Andy goes first, and plays until he completes Level 1 for the first time; then Bob takes over, and plays until he too completes Level 1 for the first time. Find the probability that the two boys require, in total, 6 attempts in order for both of them to complete Level 1 for the first time. *Given* that they require a total of 6 attempts to complete Level 1 for the first time, find the *conditional* probability that Andy requires fewer attempts than Bob.

(5)

2. (i) A random variable is said to follow a discrete uniform distribution on the range (r, s), for integers r and s with $r \le s$, if the random variable is equally likely to take any of the integer values $r, r + 1, \ldots, s$.

(a) Suppose that the random variable U follows a discrete uniform distribution on the range $(1, m)$, where m is a positive integer. Find the expected value and variance of U.

(4)

(b) Suppose that the random variable V follows a discrete uniform distribution on the range $(k + 1, k + m)$, where k is an integer and m is a positive integer. Using the results from part (a), or otherwise, show that V has expected value $k+\dfrac{m+1}{2}$ and variance $\dfrac{1}{12}\left(m^2-1\right)$.

(3)

(ii) A board game is played with a fair, six-sided die whose sides are marked with the values 1, 2, …, 6. A *turn* involves rolling this die repeatedly until a face with a value other than 6 lands uppermost. The score on a turn is calculated as the sum of the values on the faces that land uppermost on all the rolls of the die during that turn.

Let X denote the score on a turn, and let Y be the number of times the die is rolled during a turn. *Given that* $Y = y$, for any possible value y, find the *conditional* expected value and variance of X. Hence find the (unconditional) expected value and variance of X.

[You may wish to use the following results. A geometric random variable with probability function $(1 - \theta)^{w-1}\theta$, $w = 1, 2, \ldots,$ has expected value $\dfrac{1}{\theta}$ and variance $\dfrac{1-\theta}{\theta^2}$. Also $\operatorname{Var}(X)=E\{\operatorname{Var}(X\mid Y)\}+\operatorname{Var}\{E(X\mid Y)\}$.]

(13)

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

$$f(x, y) = 8xe^{-2y} , \qquad y > x > 0.$$

Derive the marginal probability density functions of X and Y. Using the results of part (i), or otherwise, deduce their expected values and variances. Find also the correlation between X and Y.

(13)

4. The continuous random variables X and Y independently follow the uniform distribution on the range 0 to 1. The random variables U and V are defined by

$$U = X - Y , \qquad V = X + Y .$$

(i) Sketch a graph to show the square region of the plane where the joint probability density function of U and V is non-zero. Find the joint probability density function of U and V on this region.

(10)

(ii) Find the marginal probability density functions of U and V. What is the relationship between these distributions?

(5)

(iii) Suppose that the continuous random variables W and Z are independent and that W follows a uniform distribution on the range a to $a + t$ while Z follows a uniform distribution on the range b to $b + t$ ($t > 0$). Use the result of part (ii) to deduce the probability that $W + Z$ does not exceed $a + b + \frac{t}{2}$.

(5)

Turn over

5. (i) The random variable X follows the Normal distribution with expected value μ and variance σ^2 (> 0). Prove that X has moment generating function (mgf)

$$M_X(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}.$$

(7)

(ii) For arbitrary constants a and b, show that the mgf of $aX+b$ is $e^{bt}M_X(at)$. Hence find the mgf of the random variable

$$Z = \frac{X-\mu}{\sigma} \ ,$$

and deduce that Z follows the standard Normal distribution.

(5)

(iii) Use the mgf of Z to find the variance of Z^2.

(8)

6. Any baby born in a certain country is equally likely to be a boy or a girl, independently for all births. Suppose that the birthweight of a baby boy is a continuous random variable with cumulative distribution function $F_1(.)$ and probability density function $f_1(.)$, while the birthweight of a baby girl is a continuous random variable with cumulative distribution function $F_2(.)$ and probability density function $f_2(.)$. Let X denote the birthweight of a baby selected at random from this population.

(i) Show that X has cumulative distribution function

$$F(x) = \frac{1}{2}\{F_1(x) + F_2(x)\} \; .$$

Find also an expression for the probability density function of X in terms of $f_1(x)$ and $f_2(x)$.

(4)

(ii) The birthweights of baby boys and girls have expected values μ_1 and μ_2, respectively, and common variance σ^2. Show that

$$E(X) = \frac{1}{2}(\mu_1 + \mu_2) \; ,$$

$$\text{Var}(X) = \sigma^2 + \frac{1}{4}(\mu_1 - \mu_2)^2 \; .$$

(7)

(iii) A random sample of size $2n$ is to be drawn from all the babies born in this country in a particular year. Let $\overline{X}_r$ denote the mean birthweight of the babies in this sample. Write down an approximation to the distribution of $\overline{X}_r$ when n is large.

(3)

(iv) It is suggested that it would be better to take a stratified random sample of babies, by separately selecting random samples of n boys and n girls. Let $\overline{X}_{st}$ denote the mean birthweight of all $2n$ babies in this sample. Write down the expected value and variance of $\overline{X}_{st}$. Explain why you might prefer $\overline{X}_{st}$ to $\overline{X}_r$ as an estimator of the population mean birthweight.

(6)

Turn over

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

0.3612 0.6789 0.3552 0.2898

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

(a) $P(X = x) = \binom{3}{x}\left(\frac{1}{5}\right)^{x}\left(\frac{4}{5}\right)^{3-x}$, $x = 0, 1, 2, 3.$

(5)

(b) $f_X(x) = 2x^3 \exp\left(-\frac{1}{2}x^4\right)$, $0 \le x.$

(5)

(ii) A small post office opens at 9.00 a.m. The time (minutes) until the first customer arrives, and the times (minutes) between the arrivals of any two consecutive customers thereafter, are all independent exponential random variables with expected value 2.

The post office has just one server. As customers arrive, they join a queue to be served. As soon as the server has finished serving one customer, she starts to serve the first customer in the queue (if there is a queue). If a customer arrives when no previous customer is being served and there is no queue, then the server begins to serve that customer immediately on his or her arrival. The service times (minutes) for individual customers are independent uniform random variables on the range 1.5 to 2.5.

Use pairs of the uniform random numbers given in part (i), in the order given, to simulate the arrival and service times for each of the first two customers to arrive at this post office. Justify your method of simulation, and record the simulated arrival times of these customers and the time at which the server finishes serving them.

(10)

8. The Ehrenfest model of the flow of molecules between two chambers is based on a physical system that consists of two urns, A and B, containing a total of M balls. At each step of a process, one of the M balls is chosen at random and moved from its current urn to the other urn. Let the states of the system be the number of balls in urn A.

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

(4)

(ii) Write down a set of equations that must be satisfied by any stationary distribution of this system. Show that these equations are satisfied by the following probabilities:

$$\pi_j = \binom{M}{j}\left(\frac{1}{2}\right)^M, \qquad j = 0, 1, ..., M.$$

(10)

(iii) Find, approximately, the long-run proportion of the time urn A contains 34 balls, in the case $M = 60$.

(6)