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

GRADUATE DIPLOMA, 2008

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 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. A random sample of size n ($n \geq 2$) is drawn from the standard exponential distribution, whose probability density function (pdf) is $\exp(-x)$ on $x > 0$. Let U and V denote respectively the minimum and maximum values in the sample.

(i) Obtain the joint pdf of U and V, and also the marginal pdf of U.

(5)

(ii) By carrying out a bivariate transformation, or otherwise, deduce the joint distribution of U and the sample range R (where $R = V - U$). Deduce that the marginal pdf of R is given by

$$f_R(r) = (n-1)\exp(-r)(1-\exp(-r))^{n-2}, \qquad r \geq 0.$$

(8)

(iii) Suppose that $n = 2$. Identify the distributions of U and R in this case and hence find their variances. Given that the variance of V is 1.25, use the relation $R = V - U$ to deduce the correlation between U and V.

(7)

2. The independent random variables X and Y each have a discrete uniform distribution on the range $\{1, 2, 3, 4\}$, i.e. their probability mass functions (pmfs) are

$p_X(x) = 0.25, \quad x = 1, 2, 3, 4,$

$p_Y(y) = 0.25, \quad y = 1, 2, 3, 4.$

Let $W = \max(X, Y)$ and $V = \min(X, Y)$.

(i) Tabulate the joint pmf of W and V and obtain the marginal pmfs of W and V.

(7)

(ii) Calculate $E(V)$ and $\mathrm{Var}(V)$.

(5)

(iii) Verify that $P(V = v) = P(W = 5 - v)$ for all possible values v of V, and hence or otherwise find $E(W)$ and $\mathrm{Var}(W)$.

(3)

(iv) Find $\mathrm{Cov}(W, V)$ and hence calculate $\mathrm{Corr}(W, V)$.

(5)

3. The random variables $X_1, X_2, X_3, \ldots$ are independent and identically distributed. Each has probability generating function (pgf) $G_X(s)$. The random variable N is independent of $X_1, X_2, X_3, \ldots$, and has pgf $G_N(s)$. The random variable T is defined as

$$T = X_1 + X_2 + \ldots + X_N \,.$$

You may **use without proof** the result that the pgf of T, $G_T(s)$, is given by $G_T(s) = G_N(G_X(s))$.

(i) Assuming that the means μ_N, μ_X and variances σ_N^2, σ_X^2 of N and X_1, X_2, X_3, … are finite, show that $E(T)$ and $\mathrm{Var}(T)$ are respectively given by

$$\mu_T = \mu_N \mu_X \quad \text{and} \quad \sigma_T^2 = \sigma_N^2 \mu_X^2 + \mu_N \sigma_X^2 \,.$$

(6)

(ii) A child spins a fair coin repeatedly until a tail is obtained. Every time the coin shows a head, she rolls a fair six-sided die and notes the score. When the coin first shows a tail the die is not rolled and the game ends. Let N denote the number of times a head is shown before the game ends, and let T denote the total score on all rolls of the die.

(a) Obtain the probability mass function (pmf) of N, and show that its pgf is given by

$$G_N(s) = \frac{1}{2-s}, \qquad |s| < 2\,.$$

Hence or otherwise obtain $E(N)$ and $\mathrm{Var}(N)$.

(8)

(b) Write down the pmf of the score, X, when a fair six-sided die is rolled, and show that $E(X) = 3.5$ and $\mathrm{Var}(X) = \dfrac{35}{12}$.

(4)

(c) Hence find $\mathrm{E}(T)$ and $\mathrm{Var}(T)$.

(2)

4. An insurance company offers annual motor-car insurance based on a "no claims discount" system with levels of discount 0%, 30% and 60%. A policyholder who makes no claims during the year either moves to the next higher level of discount or remains at the top level. If there is exactly one claim during the year, the policyholder either moves down one level or stays at the bottom level (0%). If there is more than one claim during the year, the policyholder either moves down to or stays at the bottom level. For a particular policyholder, it may be assumed that claims arise in a Poisson process at rate $\lambda > 0$.

(i) Explain why the situation described above is suitable for modelling in terms of a Markov chain with three states, and write down the transition probability matrix in terms of λ.

(6)

(ii) Write down equations for the steady-state (equilibrium) probabilities of occupying the different states, and hence find these probabilities. Use your solution to show that the long-term average level of discount is

$$\frac{0.3\left(e^{\lambda}+1\right)}{e^{2\lambda}-\lambda}\ .$$

(9)

(iii) For policyholders A and B, the parameter λ takes the values 0.1 and 0.2 respectively. Calculate the average discount levels for A and B and comment briefly.

(5)

5. Throughout this question, U(0, 1) denotes the continuous uniform distribution over the range (0, 1).

(a) The Pareto distribution with parameters $\alpha > 0$ and $\lambda > 0$ has probability density function (pdf) given by

$$f_X(x) = \frac{\alpha\lambda^{\alpha}}{(\lambda + x)^{\alpha+1}}, \quad x > 0.$$

It is required to generate pseudorandom numbers that follow this distribution. Show how this may be done by the inversion method, and apply this method to the case $\alpha = 3$, $\lambda = 5$, given that a U(0, 1) random number generator has produced the value $u = 0.327$.

(7)

(b) It is required to generate pseudorandom values of the random variable X which follows the discrete uniform distribution with probability mass function

$$f_X(x) = \frac{1}{n}, \quad x = m, m+1, m+2, \ldots, m+n-1,$$

where m and n are integers and $n > 0$.

(i) Obtain the cumulative distribution function of X, and explain how the discrete analogue of the inversion method may be used to generate a pseudorandom value from this distribution. Illustrate your answer in the case $m = 1$, $n = 20$, given that a U(0, 1) random number generator has produced the value $u = 0.915$.

(7)

(ii) Suppose that U has the U(0, 1) distribution and that $Y = \text{int}(m + nU)$ is the largest integer that does not exceed $m + nU$. Explain carefully why Y has the same distribution as X.

(6)

6. (a) The independent random variables Y and Z each have the exponential distribution with mean 2, and the random variable T is defined as Y/Z. Obtain the joint probability density function (pdf) of T and Z, and deduce that the pdf of T is given by

$$f_T(t) = \frac{1}{(1+t)^2}, \quad t > 0.$$

(8)

(b) Let X and W be independent random variables, each having the Cauchy distribution with pdf

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty,$$

and let $V = (X + W)/2$.

(i) By writing down the joint pdf of W and X and transforming from (X, W) to (V, W), show that the joint pdf of V and W is given by

$$f_{V,W}(v,w) = \frac{2}{\pi^2(1+w^2)\left[1+(2v-w)^2\right]}, \quad -\infty < v,\ w < \infty.$$

(4)

(ii) You are given that this joint pdf may be rewritten in the form

$$f_{V,W}(v,w) = \frac{1}{2\pi^2 v(1+v^2)}\left[\frac{v+w}{1+w^2} + \frac{3v-w}{1+(2v-w)^2}\right].$$

Use this to obtain the marginal pdf of V, and comment briefly.

(8)

7. The random variables X and Y have the (trinomial) probability mass function (pmf)

$$p_{X,Y}(x,y) = \frac{n!}{x!\,y!\,(n-x-y)!} p^x q^y (1-p-q)^{n-x-y},$$

for $x, y = 0, 1, \ldots, n$ subject to $x + y \le n$.

(i) Show that the marginal pmf of X is of binomial form, and identify its parameters.

(5)

(ii) Suppose x ($0 \le x < n$) is given. Show that the conditional distribution of Y, given $X = x$, is of binomial form, and write down $E(Y \mid X = x)$.

(6)

(iii) Using the result that $E(Y) = E_X[E(Y \mid X)]$, obtain the unconditional expectation $E(Y)$.

(2)

(iv) Obtain $E(XY)$ and deduce that $\text{Cov}(X, Y) = -npq$.

(4)

(v) Write down $\text{Var}(X)$ and $\text{Var}(Y)$ and deduce the expression for $\text{Corr}(X, Y)$, simplifying your answer as far as possible.

(3)

Turn over

8. The zero-mean random variables X, Y, Z have a multivariate Normal distribution with covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix},$$

where $|\rho| < 1$.

(i) Given that the marginal joint probability density function (pdf) of X and Y is

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \left(x^2 - 2\rho xy + y^2 \right) \right], \quad -\infty < x, y < \infty,$$

obtain the marginal pdf of X and hence or otherwise show that the conditional distribution of $Y|X = x$ is Normal with mean $x\rho$ and variance $1 - \rho^2$.

[Hint. The substitution $u = y - x\rho$ may be useful.]

(6)

(ii) Deduce the distribution of $Z|X = x$.

(2)

(iii) Let $W = X - \lambda Y$, where λ is a constant. Find λ such that Cov(Y, W) = 0. Show that with this choice of λ, ρ_{zw} = Corr(Z, W) is given by $\rho\sqrt{\dfrac{1-\rho}{1+\rho}}$.

(6)

(iv) Since W and Y are uncorrelated with this choice of λ, the squared multiple correlation of Z on W and Y is given by $\rho^2_{z.wy} = \rho^2_{zw} + \rho^2_{zy}$. Use this result to find $\rho^2_{z.wy}$. Using the fact that $\rho^2_{z.wy} \le 1$, or otherwise, deduce that $\rho \ge -1/2$.

(6)