# EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

*(formerly the Examinations of the Institute of Statisticians)*



## GRADUATE DIPLOMA, 2007

## 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* $\begin{pmatrix} n \\ r \end{pmatrix}$ *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.  The independent discrete random variables $X$ and $Y$ have the following Poisson distributions (where $\theta > 0$ and $\lambda > 0$).

$$P(X = x) = \frac{e^{-\theta}\theta^x}{x!}, \quad x = 0, 1, \ldots$$

$$P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, \ldots$$

(i)   Obtain expressions for the following probabilities.

(a)   $P(X = x, Y = y)$,   $x = 0, 1, \ldots$,   $y = 0, 1, \ldots$

(b)   $P(X + Y = 0)$

(c)   $P(X + Y = 1)$

(d)   $P(X + Y = z)$,   $z = 0, 1, \ldots$

Deduce that the random variable $X + Y$ also follows a Poisson distribution, and state its expected value.

(10)

(ii)  Suppose now that $X + Y$ is known to equal $z$, where $z$ is some positive integer. Determine $P(Y = y \mid X + Y = z)$ for all possible values of $y$. Name the conditional distribution of $Y$, given that $X + Y = z$.

(6)

(iii) Suppose that $U_1$, $U_2$, $\ldots$, $U_n$ are independent random variables all having the same distribution as $X$ above. State the distribution of $U_1 + U_2 + \ldots + U_{n-1}$. Deduce the conditional distribution of $U_n$ given that $U_1 + U_2 + \ldots + U_n = z$, where $z > 0$.

(4)

**Turn over**

2. A local authority consists of 20 elected members, of whom 12 represent the Blue party, 4 the Red party and 4 the Yellow Party. Five of the elected members are to be selected at random to form a new licensing committee. Let the random variables $X$ and $Y$, respectively, be the numbers of Blues and Reds who are chosen to serve on the new committee.

(i) Explain why the joint probability distribution of $X$ and $Y$ is

$$P(X = x, Y = y) = \frac{\binom{12}{x}\binom{4}{y}\binom{4}{5-x-y}}{\binom{20}{5}}$$

for appropriate values of $x$ and $y$, which you should specify.

(4)

(ii) Find the probability that the new committee consists of 3 Blues, 1 Red and 1 Yellow.

(2)

(iii) Without doing any algebra, explain why $X$ follows a hypergeometric distribution, and write down an expression for the form of $P(X = x)$.

(4)

(iv) Find the probability that more than one Yellow is chosen to serve on this committee.

(3)

(v) The conditional distribution of $X$, given $Y = y$ (for any possible value $y$), is also a hypergeometric distribution. Without doing any algebra, explain why. Write down the form of $P(X = x \mid Y = 2)$.

(4)

(vi) Find the probability that one Red is chosen <u>given</u> that all four Yellows are chosen to serve on the licensing committee.

(3)

3. (i) Suppose that $X$ and $Y$ are continuous random variables with joint probability density function $f(x, y)$. Let $h(X)$ be any real-valued function of $X$. Prove that

$$E\{h(X).Y\} = E_X\{h(X).E_Y(Y \mid X)\}$$

(6)

(ii) Suppose now that the regression of $Y$ on $X$ is linear, that is

$$E_Y(Y \mid X) = \alpha + \beta X .$$

(a) Using the result of part (i), find expressions for $E(Y)$ and $E(XY)$ in terms of $E(X)$ and $E(X^2)$. Hence find the correlation $\rho$ between $X$ and $Y$ in terms of $\beta$ and the unconditional variances of $X$ and $Y$.

(9)

(b) Suppose also that the conditional variance of $Y$ given $X$ is constant, that is

$$\text{Var}_Y(Y \mid X) = \sigma^2 \qquad \text{(whatever the value of } X\text{)}.$$

Express $\sigma^2$ in terms of $\rho$ and the unconditional variance of $Y$.

[Hint. You may use the following identity without proof:
$$\text{Var}(Y) = \text{Var}\{E(Y \mid X)\} + E\{\text{Var}(Y \mid X)\} .]$$

(5)

4. (i) Suppose that $X$ and $Y$ are independent N(0, 1) random variables. Let

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

Find the joint probability density function of $U$ and $V$.

Deduce that $U$ and $V$ are independent, and identify their marginal distributions.

(12)

(ii) Suppose that $W_1$ and $W_2$ are a random sample of size two from the Normal distribution with expected value $\mu$ and variance $\sigma^2$ ($> 0$). Write down the sample mean. Show that the sample variance is

$$\tfrac{1}{2}(W_1 - W_2)^2 .$$

Use the results proved in part (i) to show that this sample mean and sample variance are independent.

(8)

4

5.  (i)   Suppose that the random variable $X$ has moment generating function (mgf) $M_X(t)$. For arbitrary constants $a$ and $b$, show that the mgf of $aX + b$ is $e^{bt} M_X(at)$.

(3)

(ii)  Suppose that the random variable $X$ follows the Normal distribution with expected value $\mu$ and variance $\sigma^2$ $(> 0)$.

(a)   Show that $X$ has mgf

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

(7)

(b)   Using (i) and (ii)(a), find the mgf of the random variable

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

Deduce that $Z$ follows the standard Normal distribution.

(3)

(c)   If $Y = e^X$, then $Y$ is said to follow a lognormal distribution. Show that, for any positive integer $k$, the expected value of $Y^k$ is $M_X(k)$. Use this result to find the expected value and variance of the lognormal distribution.

(7)

6.    [Throughout this question, you may use without proof the result that, for non-negative integers $r$ and $s$,

$$\int_0^1 u^r (1-u)^s \, du = \frac{r! \, s!}{(r+s+1)!} \, .]$$

A random sample of size $n$ is drawn from the uniform distribution on the interval 0 to 1. The ordered values in this sample are $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$.

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

(8)

(ii)   Deduce the expected value and variance of the sample median when $n$ is an <u>odd</u> number.

(2)

(iii)  In the special case when $n = 4$, write down the joint pdf of $X_{(2)}$ and $X_{(3)}$. Hence find $E\left(X_{(2)} X_{(3)}\right)$.

(5)

(iv)   Using the results of parts (i) and (iii), find the expected value and variance of the sample median when $n = 4$.

(5)

**Turn over**

7.    (i)    The continuous random variable $U$ follows the uniform distribution on the range 0 to 1, and $\lambda > 0$ is given. Define $X$ by

$$X = -\frac{1}{\lambda}\log(1-U).$$

By finding its probability density function, prove that $X$ follows the exponential distribution with expected value $\frac{1}{\lambda}$.

(5)

(ii)    The random variable $Y$ follows the gamma distribution with probability density function

$$f_Y(y) = \lambda^2 y e^{-\lambda y}, \qquad y > 0.$$

Show that $Y$ has cumulative distribution function

$$F_Y(y) = 1 - e^{-\lambda y} - \lambda y e^{-\lambda y}, \qquad y > 0.$$

Explain why the probability integral transform does not provide a suitable method for simulating from this gamma distribution. Suggest a more suitable method that makes use of the result proved in part (i).

(7)

(iii)   Use the table of random digits provided (Table 12 in the accompanying "Statistical tables for use in examinations") to generate 4 random variates from a uniform distribution on the range 0 to 1. Work to three decimal places, and explain your method briefly.

Hence, using Table 1 of "Statistical tables for use in examinations", obtain 4 random variates from the binomial distribution with index $n = 15$ and parameter $p = 0.25$. Explain your method carefully.

(8)

**Turn over**

8.    An insurance firm operates a system of "no claims" bonuses on car insurance.  When drivers are first insured, they pay the basic annual premium.  Drivers who make no claim on their insurance policy in the current year may qualify for a reduction on their premium for the next year.  With one claim-free year they have a small reduction, with two consecutive claim-free years they have a larger reduction, and with three or more consecutive claim-free years they have an even larger reduction.  Drivers who reach this third level of reduction continue at that level in all subsequent years until they make a claim.  A driver who makes at least one claim in the current year loses any "no claims" bonus and returns to the basic premium level for the following year.  The insurance firm believes that a proportion $\theta$ $(0 < \theta < 1)$ of drivers make at least one claim in any given year.

(i)    Stating any necessary assumption, describe how a driver's experience of "no claims" bonuses may be modelled as a Markov chain.  Write down the transition matrix for this model.

(4)

(ii)    Find the two-step and three-step transition matrices for this model.  Suppose that, in the current year, a driver has one year's "no claims" bonus.  Find the probability that this driver will be back in the same state of the model in (a) two years' time, (b) three years' time.  Comment on your answers.

(8)

(iii)    Find the stationary distribution for this model.

(8)