EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2014

MODULE 1 : Probability distributions

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.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. (i) Let X be a continuous random variable and A an event with probability θ, where $0 < \theta < 1$. Conditional on A, X has cumulative distribution function $F_1(x)$ and expectation μ_1 while, conditional on A' (the complement of A), X has cumulative distribution function $F_2(x)$ and expectation μ_2. Justify the expression

$$F(x) = \theta F_1(x) + (1-\theta)F_2(x)$$

for the cumulative distribution function $F(x)$ of X and hence deduce that

$$E(X) = \theta\mu_1 + (1-\theta)\mu_2 .$$

(5)

Small chocolate biscuits of a certain brand are sold in packets of 6 with a nominal weight of 25 g. The weight (g) of an individual biscuit is a N(4.5, 0.25) random variable, and the weights of different biscuits are independent.

(ii) Find the probability that, in total, 6 of these biscuits weigh less than 25 g.

(5)

(iii) When 6 biscuits are put together to form a packet, if their total weight is found to be less than 25 g then a seventh biscuit is added to the packet. Find the mean weight of a packet of biscuits.

(4)

(iv) If it is simply known that the weight of an individual biscuit is a N(4.5, σ^2) random variable, for what values of σ is the probability less than 0.01 that 6 biscuits weigh less than 25 g?

(6)

2. The continuous random variable X has a probability density function, $f(x)$, that is symmetrical about $x = 0$, i.e. $f(-x) = f(x)$ for all x, and all moments of X exist.

(i) Use integration to prove that the median of X is 0.

(4)

(ii) Use integration to prove that $E(X) = 0$.

(3)

Let $Y = X^2$.

(iii) Show that $E(XY) = 0$. Deduce that X and Y are uncorrelated. Are X and Y independent? Justify your answer.

(7)

(iv) Show that Y has probability density function $g(y)$ given by

$$g(y) = \frac{1}{\sqrt{y}} f\left(\sqrt{y}\right), \quad y \geq 0.$$

(6)

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

$$f(x, y) = \frac{1}{x} e^{-x}, \qquad 0 < y < x.$$

(i) Show that X has an exponential distribution. Hence show that, conditional on $X = x$, Y has the uniform distribution on the interval $(0, x)$.

(7)

(ii) Show that, for non-negative integers m and n,

$$E(X^m Y^n) = \frac{(m+n)!}{n+1}.$$

(5)

(iii) Use the result proved in part (ii) to obtain $E(X)$, $\text{Var}(X)$, $E(Y)$ and $\text{Var}(Y)$. Find and interpret the value of the correlation between X and Y.

(8)

[You may use without proof the result that, for any non-negative integer r,

$$\int_0^\infty u^r e^{-u}\, du = r!]$$

4. (a) The discrete random variable X_1 takes the value 0 with probability 0.2 and the value 1 with probability 0.8. The discrete random variable X_2 takes the value 0 with probability 0.4 and the value 1 with probability 0.6. X_1 and X_2 are correlated and $P(X_1 = 1, X_2 = 1) = 0.5$.

(i) Produce a table that shows all the values of the joint probability distribution of (X_1, X_2).

(4)

(ii) Let $Y = 2X_1 + X_2$. Using the table of random digits in the *Statistical tables for use in examinations*, simulate 10 values from the distribution of Y. Explain your method carefully.

(6)

(iii) Use the simulated values from part (ii) to obtain 10 simulations from the joint distribution of (X_1, X_2).

(2)

(b) The following values are a random sample from a uniform distribution on the range 0 to 1.

0.209 0.363 0.516 0.970

(i) Use these values to generate 4 random variates from the standard Normal distribution, explaining your method carefully. (Give your answers to two decimal places.)

(4)

(ii) Use the results of part (i) to generate 4 random variates from the χ^2 distribution with one degree of freedom.

(2)

(iii) Use the results of part (ii) to generate 1 random variate from the χ^2 distribution with four degrees of freedom.

(2)

5. (a) The continuous random variables X_1 and X_2 jointly have the bivariate Normal distribution with expectation $(-1 \quad 1)^T$ and covariance matrix $\begin{pmatrix} 1 & 2 \\ 2 & 16 \end{pmatrix}$.

(i) Find the correlation between X_1 and X_2.

(2)

(ii) Let $Y_1 = X_1 + X_2$. Write down $E(Y_1)$ and $\text{Var}(Y_1)$.

(3)

(iii) Let $Y_2 = kX_1 + X_2$ where k is a constant. Find the value of k such that Y_2 is independent of Y_1.

(5)

(b) Let $X_1, X_2, \ldots, X_n$ ($n \geq 3$) be independent random variables, each with the $N(0, \sigma^2)$ distribution for some $\sigma^2 > 0$. The random variable Y_t is defined by

$$Y_t = \tfrac{1}{3}(X_t + X_{t+1} + X_{t+2}), \quad t = 1, 2, \ldots, n-2.$$

Find the mean vector and covariance matrix of $Y_1, Y_2, \ldots, Y_{n-2}$. What is the joint distribution of $Y_1, Y_2, \ldots, Y_{n-2}$?

(10)

6. (i) The continuous random variable X has the gamma distribution with parameters $\alpha > 0$ and $\theta > 0$, so that X has probability density function

$$f(x) = \frac{\theta^{\alpha} x^{\alpha-1} e^{-\theta x}}{\Gamma(\alpha)}, \quad x > 0,$$

where $\Gamma(\alpha)$ is the gamma function defined by $\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u}\, du$.

Show that the moment generating function of X is

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

Hence show that $E(X) = \dfrac{\alpha}{\theta}$ and $\mathrm{Var}(X) = \dfrac{\alpha}{\theta^2}$.

(9)

(ii) The exponential distribution is the special case of the gamma distribution with $\alpha = 1$. Let $X_1, X_2, \ldots, X_n$ ($n \geq 1$) be independent random variables each with the exponential distribution with parameter θ. Using moment generating functions, show that $\sum_{i=1}^{n} X_i$ has the gamma distribution with parameters n and θ.

(5)

(iii) State the *Central Limit Theorem*. Use it, along with the results from parts (i) and (ii), to prove that the gamma distribution with parameters n and θ can be approximated by a Normal distribution for large enough n. State clearly the parameters of this Normal distribution.

(6)

7. A bag contains m red beads and $n - m$ blue beads (where $m \geq 1$ and $n - m \geq 1$). k beads are chosen at random without replacement from all the n beads in the bag. Let the random variable X_i ($i = 1, 2, \ldots, k$) take the value 1 if the ith bead chosen is red and the value 0 if it is blue.

(i) For $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, k$, where $i \neq j$, explain why $P(X_i = 1) = \frac{m}{n}$ and $P(X_i = X_j = 1) = \frac{m(m-1)}{n(n-1)}$. Use these results to find the expected value and variance of X_i and to show that the covariance between X_i and X_j ($i \neq j$) is

$$\text{Cov}(X_i, X_j) = -\frac{m(n-m)}{n^2(n-1)}.$$

(13)

(ii) The random variable S is the total number of red beads removed from the bag. Use the results derived in part (i) to find the expected value of S and show that

$$\text{Var}(S) = k\frac{m}{n}\left(1 - \frac{m}{n}\right)\left(\frac{n-k}{n-1}\right).$$

(7)

8. A competitor is to shoot at a vertical target, with centre point O. X and Y are, respectively, the horizontal and vertical displacements (in cm) from O to the point where a bullet fired by this competitor hits the target. R and Q are defined by

$$X = R\cos Q, \quad Y = R\sin Q.$$

Here, $R > 0$ is the distance (in cm) from O to the point where a bullet hits the target and Q (in the range 0 to 2π radians) is the angle from the horizontal axis to the ray through O on which the bullet lies, measuring counter-clockwise.

(i) X and Y are modelled as independent Normal random variables, each with expected value 0 and standard deviation σ. Find the joint probability density function of R and Q.

(11)

(ii) Explain how you know that R and Q are independent. Find the marginal probability density function of R.

(4)

(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)

BLANK PAGE