EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2016

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* ${}^nC_r$.

This examination paper consists of 12 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. (i) The discrete random variable U is equally likely to take any of the integer values 1, 2, …, k (for $k \geq 1$). Show that $E(U) = \frac{k+1}{2}$ and $\text{Var}(U) = \frac{k^2 - 1}{12}$.

(6)

[You may use without proof the result that $1^2 + 2^2 + \ldots + k^2 = \frac{k(k+1)(2k+1)}{6}$.]

(ii) X and Y are independent geometric random variables with probability distributions

$$P(X = x) = (1-\theta)^{x-1}\theta, \quad x = 1, 2, \ldots$$

and $$P(Y = y) = (1-\theta)^{y-1}\theta, \quad y = 1, 2, \ldots$$

where $0 < \theta < 1$.

(a) Derive the probability distribution of the random variable $S = X + Y$.

(6)

(b) Find $P(X = x \mid S = s)$ for all possible values of x, when $s \geq 2$ is a fixed integer. Write down expressions for $E(X \mid S = s)$ and $\text{Var}(X \mid S = s)$.

(8)

2. [In this question, you may use without proof the result that, for any non-negative integer k,

$$\int_0^\infty t^k e^{-t}\,dt = k!\ ]$$

(i) The continuous random variable U has the gamma distribution with probability density function

$$f(u) = \frac{\theta^n u^{n-1} e^{-\theta u}}{(n-1)!}, \qquad u > 0,$$

where n is a positive integer and $\theta > 0$. Show that $E(U) = \dfrac{n}{\theta}$ and $\mathrm{Var}(U) = \dfrac{n}{\theta^2}$.

(6)

(ii) The continuous random variables X and Y have joint probability density function

$$f(x, y) = \begin{cases} 4xe^{-(x+y)}, & 0 < x < y, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Derive the marginal probability density function of X. Use part (i) to deduce $E(X)$ and $\mathrm{Var}(X)$.

(5)

(b) Derive the conditional probability density function of Y given $X = x$. Find $E(Y \mid X)$.

(6)

(c) Hence find $E(Y)$.

(3)

3. The continuous random variable U, which may take only non-negative values, has cumulative distribution function $F(u)$ and probability density function $f(u)$. The hazard function of U is defined to be

$$h(u) = \frac{f(u)}{1 - F(u)}, \quad u \geq 0.$$

(i) The random variable Z has a Weibull distribution with probability density function given by

$$f(z) = \begin{cases} \alpha \theta z^{\theta - 1} \exp(-\alpha z^{\theta}), & z \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha > 0$ and $\theta > 0$. Derive the hazard function of Z. Write down conditions on θ that give

(a) a hazard function that is constant for all $z \geq 0$,

(b) a hazard function that decreases as z increases.

(6)

A system consists of two components, C_1 and C_2. Failures occur in the two components independently. Let the random variable X_i (i = 1, 2) be the time to first failure of C_i, where X_i has cumulative distribution function $F_i(x_i)$ and probability density function $f_i(x_i)$. Let Y be the time to first failure of the system.

(ii) C_1 and C_2 are connected in series. This means that the system fails as soon as either of the two components fails. Show that Y has cumulative distribution function

$$G(y) = F_1(y) + F_2(y) - F_1(y)F_2(y), \quad y \geq 0.$$

Deduce that the hazard function of Y is the sum of the hazard functions of X_1 and X_2. Hence or otherwise show that, if X_i (i = 1, 2) follows a Weibull distribution with parameters α_i and θ, then Y also follows a Weibull distribution.

(8)

(iii) The components C_1 and C_2 are now connected in parallel to create a new system. The time to first failure of this system is the maximum of the times to first failure of the two components. Explain why the time to first failure of this system, Y, has cumulative distribution function

$$G(y) = F_1(y)F_2(y), \quad y \geq 0.$$

In the case where the two components are identical, and therefore have identically distributed times to first failure, obtain an expression for the hazard function of the system in terms of the cumulative distribution function and probability density function of an individual component. Show that the hazard for the system, $h(y)$, is no greater than the hazard for an individual component, for all values of y.

(6)

4. (a) The continuous random variables X_1 and X_2 jointly have a bivariate Normal distribution. X_1 has expected value 50 and standard deviation 8. X_2 has expected value 45 and standard deviation 10. The correlation between X_1 and X_2 is –0.25.

(i) Obtain the expectation and covariance matrix of the random vector

$$\mathbf{X} = (X_1, X_2)^T.$$

(3)

(ii) In different physical units, the same quantities can be recorded as

$$Y_1 = 0.555(X_1 - 32) \text{ and } Y_2 = 0.447X_2.$$

Obtain the distribution of the random vector $\mathbf{Y} = (Y_1, Y_2)^T$.

(5)

(iii) Use this example to illustrate one reason why correlation is often preferred to covariance as a measure of association between two random variables.

(2)

(b) In a longitudinal study of child growth, measurements X_1, X_2 and X_3 are made of a baby's length at ages 1 month, 2 months and 3 months respectively. These random variables are modelled by a multivariate Normal distribution with

$$E(\mathbf{X}) = \begin{pmatrix} \mu \\ \mu+\alpha \\ \mu+2\alpha \end{pmatrix}, \quad \mathrm{Cov}(\mathbf{X}) = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 \end{pmatrix},$$

where $\sigma^2 > 0$ and $0 < \rho < 1$.

(i) Write down the correlations between all possible pairs of these measurements.

(2)

(ii) The random variable Y is the mean length of an individual child at these three ages. Specify the distribution of Y.

(5)

(iii) Measurements are to be made, independently, on a sample of n children. The random vector of measurements on the ith child (i = 1, 2, …, n) is $\mathbf{X}_i$. State the distribution of the sample mean vector $\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i$, giving its expectation vector and covariance matrix.

(3)

5. The independent random variables X_1, X_2 and X_3 each have the uniform distribution on the interval 0 to 1. Let W denote the median of X_1, X_2 and X_3, and V denote the maximum of X_1, X_2 and X_3.

(i) The joint probability density function of two order statistics, $X_{(i)}$ and $X_{(j)}$, of a sample of n independent and identically distributed random variables, each with probability density function $f(x)$ and distribution function $F(x)$, is given by

$$g(x_{(i)}, x_{(j)}) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x_{(i)})]^{i-1}[F(x_{(j)})-F(x_{(i)})]^{j-i-1}$$
$$\times [1-F(x_{(j)})]^{n-j} f(x_{(i)}) f(x_{(j)}).$$

Use this expression to show that the joint probability density function of W and V is

$$g(w,v) = 6w, \quad 0 \le w \le v \le 1.$$

(3)

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

$$E\left(W^k V^m\right) = \frac{6}{(k+2)(k+m+3)}.$$

Hence or otherwise, find $E(W)$, $\mathrm{Var}(W)$, $E(V)$, $\mathrm{Var}(V)$ and $\mathrm{Cov}(W, V)$.

(12)

(iii) In a certain experiment, the same unknown quantity θ is to be measured three times. One researcher intends to estimate θ by the median of the three measurements while another researcher prefers to estimate θ by the maximum of the three measurements. Assuming that the measurements are $\theta + X_1 - \frac{1}{2}$, $\theta + X_2 - \frac{1}{2}$ and $\theta + X_3 - \frac{1}{2}$, find the expected value and variance of the difference between these two estimators.

(5)

6. (i) The discrete random variable X has the binomial distribution

$$P(X = x) = \binom{m}{x}\theta^x (1-\theta)^{m-x}, \quad x = 0, 1, ..., m,$$

where m is a positive integer and $0 < \theta < 1$. Show that X has moment generating function

$$M_X(t) = \left(1-\theta+\theta e^t\right)^m.$$

Use this moment generating function to find the expected value and variance of X.

(9)

(ii) X_1, X_2, …, X_n are independent random variables, each with the binomial distribution given in part (i) in the special case where $m = 1$. Use moment generating functions to prove that $S = X_1 + X_2 + \ldots + X_n$ is also a binomial random variable.

(5)

(iii) State the *Central Limit Theorem*. Use it, along with the results from parts (i) and (ii), to prove that the binomial 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) The following values are a random sample from the uniform distribution on (0, 1).

0.0885 0.4096 0.7370 0.9384

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

(i) Poisson: $P(X = x) = \dfrac{e^{-2.5}(2.5)^x}{x!}, \quad x = 0, 1, 2, \ldots .$ (5)

(ii) Shifted exponential: $f(x) = 10e^{-10(x-3)}, \quad x > 3.$ (5)

(b) Beetle is a party game played using a standard die. The player attempts to complete a rough sketch of a beetle, consisting of a body, a head, a tail, four legs, two antennae and two eyes [This is a special beetle for the game, with only four legs]. On each turn, the player rolls the die once and may draw one part of the beetle depending on the score on the die as indicated in the following table.

6 = body	3 = leg
5 = head	2 = antenna
4 = tail	1 = eye

If the player has already drawn all of a particular body part (for example, four legs), no body part is added on subsequent turns when the corresponding score (for example, 3) is obtained on the die. In addition, the body itself must be drawn before any other part may be drawn. The head, tail or legs may then be attached to the body, but antennae and eyes may only be added after the head. This means that a turn results in at most one body part being added to the sketch of the beetle, but often no body part is added.

Use the following sequence of random digits, which should be read from left to right across each row, to carry out one simulation of a player's attempt to draw a beetle. Explain your method carefully. State clearly the total number of random digits and the total number of simulated rolls of a die required to complete the beetle.

5	2	1	0	9	0	0	9	6	8	6	8	0	8	9	8	6	3	6	4
2	8	8	4	5	4	1	9	3	9	0	0	6	2	6	6	2	0	6	8
7	0	5	4	9	9	6	7	5	6	5	3	9	4	4	3	5	1	0	9

(6)

Describe how you would extend this simulation to find a plausible range of values for the number of rolls of the die that a player would require in order to complete a beetle.

(4)

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

$$f(x,y) = \begin{cases} \dfrac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} x^{\alpha-1} y^{\beta-1} (1-x-y)^{\gamma-1}, & 0 < x < 1,\ 0 < y < 1,\ x + y < 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are parameters and $\Gamma(.)$ is the gamma function.

(i) Obtain the joint probability density function of $U = 1 - X$, $V = \dfrac{Y}{1-X}$.

(9)

(ii) Explain how you know that U and V are independent random variables. Obtain their marginal probability density functions and identify these marginal distributions.

(8)

(iii) Deduce the marginal distributions of X, Y and $Z \equiv 1 - X - Y$.

(3)

BLANK PAGE

BLANK PAGE