EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

GRADUATE DIPLOMA, 2017

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 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) A_1, A_2 and A_3 are events such that $P(A_2 \cap A_1) > 0$. Starting from the definition of conditional probability, show that

$$P(A_2 \cap A_1) = P(A_2 \mid A_1) P(A_1)$$

and hence that

$$P(A_3 \cap A_2 \cap A_1) = P(A_3 \mid A_2 \cap A_1) P(A_2 \mid A_1) P(A_1).$$

(5)

12% of the electrical components made in a certain factory are defective; the remainder are non-defective. All the electrical components made in this factory are inspected by Inspector 1, who passes 10% of defective components and 85% of non-defective ones but fails all other components. All components passed by Inspector 1 are then inspected independently by Inspector 2, who passes 5% of defective components and 80% of non-defective ones but fails all other components. Only components that pass both inspections are sold.

(ii) What proportion of the electrical components made in this factory are sold?

(6)

(iii) What proportion of the electrical components that are sold are, in fact, defective?

(2)

At present, Inspector 1 carries out more inspections than Inspector 2 (since Inspector 1 inspects all components but Inspector 2 does not inspect components that are failed by Inspector 1). Their manager seeks to change the system so as to equalise their workloads. It will continue to be the case that only components that are passed on a first inspection will be inspected again, but now a proportion θ of all components (chosen at random) will be first inspected by Inspector 1 and then, if necessary, by Inspector 2, while the remaining components will be first inspected by Inspector 2 and then, if necessary, by Inspector 1.

(iv) What value should θ take so that, on average, the two inspectors test the same number of components?

(7)

2. Snapdragon (*Antirrhinum*) plants can have red, pink or white flowers; the flowers on any individual plant are all the same colour. A snapdragon plant grown from a dihybrid cross has probability $\frac{1}{4}$ of having red flowers, $\frac{1}{2}$ of having pink flowers and $\frac{1}{4}$ of having white flowers. A gardener grows 20 snapdragon plants, all independently produced from a dihybrid cross. Let the random variable X be the number of these plants that have red flowers and Y the number of them that have pink flowers.

(i) Write down an expression for the joint probability distribution, $P(X = x, Y = y)$, for appropriate values of x and y which you should specify. Find the probability that the gardener grows exactly five plants with red flowers *and* exactly ten plants with pink flowers.

(4)

(ii) The random variable X has a binomial distribution. Without doing any algebra, explain why, and state the parameters of this distribution.

(4)

(iii) Let $Z = 20 - X - Y$ be the number of plants with white flowers. Write down the (marginal) distributions of Y and Z. Hence find the probability that the gardener grows no more than one plant with white flowers.

(4)

(iv) The conditional distribution of X, given $Y = y$ (for any possible value y), is a binomial distribution. Without doing any algebra, explain why. State the parameters of this conditional distribution.

(3)

(v) Assume that the time of flowering does not vary with colour, and that the first five plants to flower are all pink. Find the probability that at least three of the remaining plants will have red flowers.

(5)

3. X is a continuous random variable. With probability α, where $0 < \alpha < 1$, X is a randomly selected value from a probability distribution with cumulative distribution function $F_1(x)$, expectation μ_1 and variance σ_1^2. Otherwise, X is randomly selected from a probability distribution with cumulative distribution function $F_2(x)$, expectation μ_2 and variance σ_2^2.

(i) Justify the expression

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

for the cumulative distribution function $F(x)$ of X. Deduce that

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

(6)

(ii) By writing down a similar expression for $E(X^2)$, or otherwise, show that

$$\text{Var}(X) = \alpha\sigma_1^2 + (1-\alpha)\sigma_2^2 + \alpha(1-\alpha)(\mu_1 - \mu_2)^2 .$$

(5)

(iii) When $\sigma_1^2 = \sigma_2^2 = \sigma^2$, show that $\text{Var}(X) \geq \sigma^2$. Under what condition does equality hold?

(4)

(iv) $F_i(.)$ (i = 1, 2) is the cumulative distribution function of the exponential distribution with expected value μ_i (for some $\mu_i > 0$), where $\mu_1 \neq \mu_2$. Write down an expression for $\text{Var}(X)$. By considering $[E(X)]^2$, or otherwise, deduce that X does not have an exponential distribution.

(5)

4. (a) A stick of length 1 metre is broken at a single, randomly-chosen point along its length to produce two pieces of stick. The random variable X is the length (in metres) of the shorter of the two pieces.

(i) Obtain the cumulative distribution function of X, and hence its probability density function. Deduce that X has the uniform distribution on the range 0 to $\frac{1}{2}$.

(6)

(ii) Obtain the probability density function of Y, the ratio of the length of the shorter to the longer piece of stick, where

$$Y = \frac{X}{1-X}.$$

(5)

(b) U and W are independent random variables, each with the uniform distribution on the range 0 to 1.

(i) Explain why the random variables $X = \dfrac{U}{W}$ and $Y = W$ have the joint range space $\{(x, y): 0 \le x;\ 0 \le y \le \min(1, 1/x)\}$.

(3)

(ii) Obtain the joint probability density function of X and Y on this joint range space.

(4)

(iii) Hence, or otherwise, obtain the probability density function of $X = \dfrac{U}{W}$.

[You might wish to consider the cases $0 \le x \le 1$ and $1 < x$ separately.]

(2)

5. (i) The continuous random variable, X, has the gamma distribution with probability density function

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

where $\alpha > 0$, $\theta > 0$ and $\Gamma(\alpha)$ is the gamma function defined by

$$\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u}\, du .$$

Show that X has moment generating function

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

Hence find the expected value and variance of X.

(10)

(ii) The random variable $Y = kX$, for some positive constant k. Using moment generating functions, prove that Y also has a gamma distribution and write down its parameters.

(5)

You are given that the gamma distribution with $\alpha = \frac{n}{2}$ and $\theta = \frac{1}{2}$, for positive integer n, is the χ^2 distribution with n degrees of freedom.

(iii) Suppose that X has the gamma distribution with $\alpha = 4$ and $\theta = 3$. Use Table 5, 'Percentage points of the χ^2 distribution', in the Statistical tables provided to find the probability that X is less than 2.227.

(5)

6. The independent continuous random variables $X_1, X_2, \ldots, X_n$ (for $n \geq 2$) are identically distributed, each with cumulative distribution function $F(x)$ and probability density function $f(x)$.

(i) $V = \min(X_1, X_2, \ldots, X_n)$. Explain why, for any value v,

$$P(V \leq v) = 1 - \{1 - F(v)\}^n .$$

Hence write down an expression for the probability density function of V.

(4)

(ii) $W = \max(X_1, X_2, \ldots, X_n)$. Find the cumulative distribution function and probability density function of W in terms of $f(w)$ and $F(w)$.

(3)

(iii) Explain why, for any values v and w such that $v \leq w$,

$$P(V \leq v \text{ and } W \leq w) = P(W \leq w) - P(V > v \text{ and } W \leq w)$$

and why

$$P(V > v \text{ and } W \leq w) = [F(w) - F(v)]^n .$$

Hence show that the joint probability density function of V and W is

$$g(v, w) = n(n-1)\, f(v)\, f(w)[F(w) - F(v)]^{n-2}, \quad v \leq w .$$

(6)

(iv) Suppose that each X_i, $i = 1, 2, \ldots, n$, has the exponential distribution with probability density function $f(x) = \theta \exp(-\theta x)$, $x \geq 0$. By obtaining the joint probability density function of $R = W - V$ and $T = V$, or otherwise, show that the (marginal) probability density function of R is

$$h(r) = (n-1)\theta e^{-\theta r}(1 - e^{-\theta r})^{n-2}, \quad r > 0 .$$

(7)

7. The random variables X_0, X_1, X_2, … are the prices (in £), at consecutive time points 0, 1, 2, …, of a particular share in a company listed on the London Stock Exchange. Time point 0 is today and X_0 is observed to take the value 1. A simple model states that, for some fixed probability θ $(0<\theta<1)$, the share price will either increase from its previous value by a factor of $(1+c)$ or decrease by a factor of $(1-c)$, where $0<c<1$, independently at every future time point. Thus

$$P(X_{n+1}=x_n(1+c) \mid X_n=x_n)=\theta$$

and

$$P(X_{n+1}=x_n(1-c) \mid X_n=x_n)=1-\theta .$$

(i) Write down a table of the probability distribution of X_2. Hence obtain a table of the probability distribution of $Y_2=\log X_2$.

(3)

(ii) The random variable $Y_n=\log X_n$, n = 1, 2, … . Show that Y_n can be expressed in the form a_n+bW_n, where W_n is a binomial random variable. State the values of the constants a_n and b, and the parameters of the distribution of W_n. Deduce $E(Y_n)$ and $\mathrm{Var}(Y_n)$.

(8)

(iii) Use the Central Limit Theorem to explain why the random variable W_n can be approximated by a Normal distribution when n is large enough. Deduce that Y_n is also approximately Normally distributed, for large enough values of n.

(4)

(iv) The random variable U has a $\mathrm{N}(\mu, \sigma^2)$ distribution with moment generating function

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

Use this moment generating function to obtain the expected value and variance of $\exp(U)$ in terms of μ and σ^2. Use this general result to deduce the approximate expected value and variance of the random variable X_n for large n.

(5)

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

0.2546 0.0808 0.9812 0.3228

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

(a) $P(X = x) = \dfrac{\binom{6}{x}\binom{9}{3-x}}{\binom{15}{3}}, \quad x = 0, 1, 2, 3.$

(6)

(b) $f(x) = \dfrac{x^3}{64}, \quad 0 \le x \le 4.$

(4)

(ii) A rural bus service starts at town A, stops in towns B, C and D in turn, then terminates in town E. The journey time (minutes) between any two consecutive towns is a $\mathrm{N}\left(9, \left(\tfrac{1}{2}\right)^2\right)$ random variable. It may be assumed that the journey times between the four pairs of consecutive towns are independent.

The timetable for this service specifies that it will leave A, B, C and D at 11.00, 11.10, 11.20 and 11.30 a.m. respectively, before terminating at E at 11.40 a.m. If the bus reaches B, C or D before it is timetabled to leave that place, then it will wait there and leave at exactly the timetabled time. If the bus reaches a stop after the time when it is due to leave that place, then it will leave again immediately. (It may be assumed that zero time is required for passengers to disembark and board the bus.)

Assuming that the bus leaves A on time, use the uniform random numbers given in part (i), in the order given, to simulate one journey of the bus from A to E. State clearly when the bus arrives at E in your simulation.

Explain briefly how you would use a larger simulation to estimate the expected time at which this service arrives at E and the probability that the service is late arriving at E.

(10)

BLANK PAGE

BLANK PAGE

BLANK PAGE