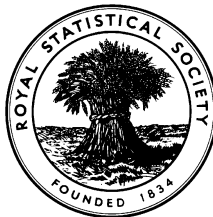


# EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY



## GRADUATE DIPLOMA, 2011

### 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 6 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. (i) The events  $\{E_i\}$  ( $i = 1, 2, \dots, n$ ) partition a sample space, and  $A$  is an event with non-zero probability  $P(A)$ . Write down Bayes' Theorem for the conditional probability  $P(E_1 | A)$ , in terms of the conditional probabilities  $\{P(A | E_i)\}$  and the unconditional probabilities  $\{P(E_i)\}$  ( $i = 1, 2, \dots, n$ ).

(4)

- (ii) The only taxis operating in a certain town are either Blue or Green, and there are four times as many Blue taxis as Green ones. A witness to a hit-and-run accident testifies that a Blue taxi was involved. In tests made under similar lighting conditions, Blue taxis are correctly identified by the witness 80% of the time, and Green taxis are correctly identified 60% of the time. Assuming all the taxis in the town were equally likely to have been involved in the incident, find the probability that the taxi was indeed Blue.

(7)

It is now discovered that three-quarters of the Blue taxis, and half the Green taxis, had been hired for use in a corporate event in another town at the time, and are thus eliminated from suspicion. What proportion of taxis available at the time were Blue, and what is now the probability that the taxi involved in the incident was Blue?

(9)

2. The independent random variables  $X_1, X_2, \dots, X_n$  all have probability density function  $e^{-x}$  on  $0 \leq x < \infty$ , and their order statistics are written  $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ .

- (i) Show that the probability density function of  $X_{(r)}$  is

$$\frac{n!}{(r-1)!(n-r)!} (1-e^{-x})^{r-1} e^{-x(n-r+1)} \quad \text{on } 0 \leq x < \infty, \text{ for } r = 1, 2, \dots, n.$$

(4)

- (ii) Find the joint probability density function of  $X_{(r)}$  and  $X_{(s)}$  when  $1 \leq r < s \leq n$ .

(6)

- (iii) Let  $U = X_{(1)}$  and  $V = X_{(n)}$  be the smallest and largest order statistics. Deduce from part (ii) that their joint probability density function is

$$n(n-1)(e^{-u} - e^{-v})^{n-2} e^{-u-v} \quad \text{on } 0 \leq u < v < \infty.$$

(2)

- (iv) Write  $R = V - U$  as the range of these random variables. Deduce the joint probability density function of  $U$  and  $R$ .

(4)

- (v) Hence find the probability density function of the range.

(4)

3. An insurance company assesses its potential liability by taking the number  $X$  of claims over a year to have a Poisson distribution with mean  $\lambda > 0$ , and the size  $U$  of any claim to follow the gamma distribution with density  $f(u) = Ku^{\alpha-1}e^{-\nu u}$  on  $u > 0$ , independently of all other claims. Here,  $K$  and  $\nu$  ( $\nu > 0$ ) are constants, with  $K$  given by  $K = \frac{\nu^\alpha}{\Gamma(\alpha)}$ , where  $\Gamma(\alpha) = \int_0^\infty w^{\alpha-1}e^{-w}dw$ .

(i) Show that the moment generating function of the given gamma density is  $\left(\frac{\nu}{\nu-t}\right)^\alpha$  when  $t < \nu$ . Deduce the associated mean and variance. (12)

(ii) Use (without proof) the identities

$$E(Y) = E(E(Y|X)) \quad \text{and} \quad \text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

to give expressions for the mean and variance of the value  $Y$  of the total claims on the company over a year. (8)

4. At the recent elections, the 26 places on a Town Council were filled by 12 councillors from the Yellow party, 8 from the Reds, and 6 from the Blues. When a Council committee is being formed, each party may nominate one of its councillors to fill one place, the remaining places (however many) being filled at random from the rest of the Council. All councillors of any party always vote the same way. The chairman is chosen at random from the whole committee, and when a vote of all the committee members is tied, the chairman's vote determines the outcome.

[Leave your answers here as exact fractions, NOT converted to approximating decimals.]

- (i) Suppose the size of a committee is seven. Calculate the probability that the Yellow party has at least four members. (6)
- (ii) Suppose the size of a committee is six. Calculate the probability that the Yellow party will win a vote, no matter how the others vote. (6)
- (iii) Suppose now that the Reds and Blues merge, to form the Purple party. Council officers propose two alternative ways forward:
- (a) all committee memberships and chairmen remain as they were before the merger;
- (b) new committees are formed completely afresh, with each of the two new parties able to nominate one member, the others again chosen at random from the rest of the Council.

For a committee of size six, and the general rules unchanged, calculate how much difference there is between (a) and (b) in the chances that the Purple party could guarantee to win a vote. (8)

5. The continuous random variables  $X$  and  $Y$  have joint probability density  $f(x, y) = Kx^2y$  over the triangle with vertices  $(-1, 0)$ ,  $(0, 1)$  and  $(1, 0)$  (and  $f(x, y) = 0$  elsewhere).

(i) Sketch the region over which the density is non-zero. Using symmetry or otherwise, evaluate the constant  $K$ , and calculate the marginal probability densities of  $X$  and  $Y$ . (8)

(ii) Find the covariance of  $X$  and  $Y$  and say, with a reason, whether or not the two variables are independent. (4)

(iii) Suppose  $-1 < t < 1$ . Explain, with the use of a diagram, why the probability  $P(X + Y \leq t)$  can be expressed as  $\int_0^w \left( \int_u^v f(x, y) dx \right) dy$ , where  $u = y - 1$  and  $v = t - y$ ; and deduce the value of  $w$ . [There is no need to evaluate this integral.] (8)

6. Suppose that the non-negative random variable  $X$ , whose mean is  $\mu$  and whose variance is  $\sigma^2$ , belongs to a family of random variables with  $\sigma^2 = K\mu^r$  for some constants  $K$  and  $r$ . It is desired to find some transformation  $Y = f(X)$  such that the variance of  $Y$  is approximately constant for members of this family.

(i) Use a Taylor series expansion to verify that, if  $r \neq 2$ , then  $f(t) = t^{1-\frac{1}{2}r}$  will achieve the objective, while if  $r = 2$ , then  $f(t) = \log(t)$  may be used. (12)

(ii) Show that a random variable  $V$  that has a Poisson distribution belongs to this family. Hence estimate the variance of the random variable  $\sqrt{V}$ . (4)

(iii) Show that, when  $0 < p < 0.1$ , the random variable  $W$ , which has the geometric distribution given by  $P(W = k) = p(1 - p)^{k-1}$  for  $k = 1, 2, 3, \dots$ , satisfies this mean-variance condition approximately. Hence estimate the variance of  $\log(W)$ . (4)

[You may use the results  $E(W) = \frac{1}{p}$  and  $\text{Var}(W) = \frac{1-p}{p^2}$  without proof.]

7. (i) Suppose  $\{Z_0, Z_1, Z_2, \dots\}$  are independent random variables, each having a standard Normal distribution. In terms of members of this sequence, write down random variables having
- (a) a chi-squared distribution on  $m$  degrees of freedom, (2)
  - (b) a  $t$  distribution with  $n$  degrees of freedom, (2)
  - (c) an  $F$  distribution on  $m$  and  $n$  degrees of freedom. (4)

(ii) Use part (i)(a) to prove that the sum of two independent chi-squared random variables also follows a chi-squared distribution. (4)

(iii) Suppose  $\{X_1, X_2, \dots, X_n\}$  are independent random variables, all having a Normal distribution with mean  $\mu$  and variance  $\sigma^2 > 0$ . Write

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad \text{and} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Name the distributions followed by  $\bar{X}$  and  $\frac{(n-1)S^2}{\sigma^2}$ . Given that  $\bar{X}$  and  $S^2$  are independent, deduce the distribution of  $\frac{(\bar{X} - \mu)\sqrt{n}}{S}$ , showing fully all steps of your argument. (8)

8. (i) A sequence of integers,  $\{u_1, u_2, u_3, \dots, u_{1000}\}$ , all between 0 and 9 inclusive, is claimed to be suitable to use as pseudo-random digits. Outline briefly three ways you might check this claim. (6)

(ii) Now assume that the claim is true. Explain how to use the sequence to generate satisfactory values of random variables having the following distributions.

(a) A continuous uniform distribution over the interval  $(2, 5)$ . (4)

(b) A binomial distribution with parameters  $n = 4$  and  $p = 0.3$  using the table-look-up method. [There is no need to make any numerical calculations here.] (4)

(c) A Cauchy distribution with density  $\frac{1}{\pi(1+x^2)}$  over  $-\infty < x < \infty$ , using the inversion method. [Derive the required formula for this method.] (6)