EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2014

MODULE 2 : Probability models

Time allowed: One and a half hours

Candidates should answer **THREE** *questions.*

Each question carries 20 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 4 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. A combination lock consists of four rings each labelled with the digits 1, 2, 3, 4, 5, 6. The rings may be rotated individually and independently, so that all 4-digit combinations of the digits 1, …, 6 (with repetition) can be shown. A customer buys such a lock. The instructions that come with the lock give the correct combination for opening the lock and state that this combination has been chosen at random from all possible combinations.

(i) Evaluate k, the total number of combinations that can be shown. (2)

(ii) Find the probability that the purchased lock has a combination

(a) with all digits equal, (1)

(b) with all digits different, (2)

(c) with a pair of digits equal, the other two digits being different from each other and from the pair, (6)

(d) with exactly three digits equal, (5)

(e) with two pairs of equal digits (but not all four digits the same). (4)

2. The continuous random variable X has probability density function (pdf) $f(x)$ given by

$$f(x) = \frac{k}{x^4}, \quad x \geq \theta,$$

where θ is a positive constant.

(i) (a) Find k in terms of θ. For the case $\theta = 1$, sketch the graph of $f(x)$, marking the value of $f(1)$ on your graph. (6)

(b) Show that $E(X) = \frac{3}{2}\theta$ and $\text{Var}(X) = \frac{3}{4}\theta^2$. (5)

(ii) Let $\bar{X}$ denote the mean of n independent random variables, $X_1, \ldots, X_n$, each of which has the pdf $f(x)$.

(a) Write down an approximation to the distribution of $\bar{X}$ based on the central limit theorem. How would you expect the success of the approximation to vary with n? (2)

(b) Use this distribution to show that $P\left(\left|\bar{X} - E(X)\right| \leq 1.96\theta\sqrt{\frac{3}{4n}}\right) = 0.95$ approximately. Hence find an approximation to the least value of n such that $P\left(\left|\bar{X} - E(X)\right| < 0.1\theta\right) \geq 0.95$. (7)

3. The probability that a given character is miscopied when I send an email is 0.001, independently of all other characters.

(i) If I send an email of 2000 characters, state

(a) the exact distribution,

(b) a suitable approximate distribution,

for the number, X, of miscopied characters. Use the approximate distribution to find $P(X = 0)$ and $P(X > 2)$.

(5)

(ii) If I send a second email, consisting of 3000 characters and independent of the first, state corresponding approximate distributions

(a) for the number, Y, of miscopied characters in the second email,

(b) for the total number, Z, of miscopied characters in the two emails combined.

Use this distribution of Z to find $P(Z = 4)$ and then use the approximate distributions of X, Y and Z to find the conditional probability $P(X = 2 \mid Z = 4)$.

(8)

(iii) In the course of a week I send 50 emails, all independent and consisting of 100 000 characters in total. State

(a) the exact distribution,

(b) a suitable approximation,

for the total number, W, of miscopied characters. Use the approximate distribution to find $P(W > 115)$.

(7)

4. Let X and Y be independent standard Normal random variables and let $\Phi(.)$ denote the cumulative distribution function of the standard Normal random variable.

(i) Write down the distribution of $4X - 3Y$ and hence find $P(4X > 3Y + 2)$.

(5)

(ii) Let $W = \max(X, Y)$.

(a) Write down $P(X \le x, Y \le x)$ in terms of $\Phi(x)$ and hence explain why the cumulative distribution function of W is given by

$$F_W(w) = [\Phi(w)]^2, \quad -\infty < w < \infty .$$

(4)

(b) Find $Q1$ and $Q3$, the lower and upper quartiles of W.

(5)

(iii) A random sample of 400 observations of W is taken. Write down the distribution of the number K of observations in the sample that lie within the interval $(Q1, Q3)$. Use a suitable approximation to calculate $P(K \le 210)$.

(6)