

THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE EXAMINATION

NEW MODULAR SCHEME

introduced from the examinations in 2007

MODULE 2

SPECIMEN PAPER B

AND SOLUTIONS

The time for the examination is 1½ hours. The paper contains four questions, of which candidates are to attempt **three**. Each question carries 20 marks. An indicative mark scheme is shown within the questions, by giving an outline of the marks available for each part-question. The pass mark for the paper as a whole is 50%.

The solutions should not be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids. For this reason, they do not carry mark schemes. Please note that in many cases there are valid alternative methods and that, in cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of the questions and solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of the questions or solutions.

1. (i) The random variable X follows the Poisson distribution with probability mass function

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

- (a) Show that the mean and variance of X are both equal to λ .
- (b) State the Poisson approximation to the binomial distribution, indicating the circumstances in which it is appropriate.

(5)

- (ii) A civil servant calculates weekly social security payments for unemployed adults. These payments vary according to claimants' circumstances, and errors may occur. Over a long period of time, the probability of a wrong calculation has been found to be 0.0075. Find to 4 decimal places the exact probability that a sample of 200 contains (a) 1 wrong calculation, (b) 4 wrong calculations.

(7)

- (iii) Repeat part (ii) using the Poisson approximation to the binomial distribution. In each case find to two significant figures the percentage error in the Poisson calculation. Comment briefly on your results.

(8)

2. Among the inhabitants of Altamania, height, H say, is distributed Normally with mean 160 cm and standard deviation 4 cm, i.e. $N(160, 16)$.

(i) Find the proportion of the population whose heights are within one standard deviation of the mean. Find also the proportion of the population who are more than 168 cm tall.

(5)

(ii) The Altamanian Police Force (APF) is restricted to persons who are more than 168 cm tall, and may be assumed to consist of a random sample of Altamanians satisfying this condition.

Find

(a) the median height of members of the APF,

(b) the proportion of members of the APF who are more than 170 cm tall.

(10)

(iii) Assume that the mean and standard deviation of height among members of the APF are 169.5 cm and 1.352 cm respectively. Find an approximate value for the probability that the mean height of a random sample of 25 members of the APF is more than 170 cm.

(5)

3. The events A , B and C have respective probabilities $\frac{2}{3}$, $\frac{1}{2}$ and $\frac{1}{4}$, and \bar{A} , \bar{B} and \bar{C} are, respectively, the complements of A , B and C .

(i) Given that A , B and C are mutually independent, find

(a) $P(A \cap \bar{B} \cap \bar{C})$,

(b) $P(A \cap \bar{C} | A \cap \bar{B})$.

(8)

(ii) Now let A and B be independent, A and C be independent and B and C be independent, so that A , B and C are pairwise independent, and let $P(A \cap B \cap C) = x$. Find in terms of x

(a) $P(A \cap \bar{B} \cap \bar{C})$,

(b) $P(A \cap \bar{C} | A \cap \bar{B})$,

(c) $P(A \cup B \cup C)$.

Also find the maximum and minimum possible values of x .

(12)

4. The random variable X has the geometric probability mass function (pmf) given by $f(x)$, where

$$f(x) = (1-p)^x p, \quad x = 0, 1, 2, \dots, \quad 0 < p < 1.$$

- (i) Sketch the graph of $f(x)$ for the case $p = 1/3$, for $0 \leq x \leq 5$. (5)

- (ii) Obtain the mean and variance of X . (4)

- (iii) For any non-negative integer x , show that $P(X \geq x) = (1-p)^x$, and deduce that for any non-negative integers l and m

$$P(X \geq l+m | X \geq l) = P(X \geq m).$$

Interpret this result. (5)

- (iv) The random variable Y has pmf $g(y)$, where

$$g(y) = (1-\theta)^y \theta, \quad y = 0, 1, 2, \dots, \quad 0 < \theta < 1.$$

X and Y are independent, and the random variable Z is defined as the minimum of X and Y , i.e. $Z = \min(X, Y)$. By noting that $P(Z \geq z) = P(X \geq z \text{ and } Y \geq z)$, find an expression for $P(Z \geq z)$, where z is any non-negative integer. By considering $P(Z \geq z) - P(Z \geq z+1)$, or otherwise, show that

$$P(Z = z) = [(1-p)(1-\theta)]^z (p + \theta - p\theta), \quad z = 0, 1, 2, \dots$$

Identify the form of this distribution and hence write down $E(Z)$ and $\text{Var}(Z)$. (6)

SOLUTIONS

Question 1

$$(i) (a) \quad E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=1}^{\infty} \frac{\lambda e^{-\lambda} \lambda^{x-1}}{(x-1)!} = \lambda \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = \lambda \times 1 = \lambda.$$

We have $\text{Var}(X) = E(X^2) - \{E(X)\}^2$. $E(X^2)$ can be found similarly, but it is easier to use $E(X^2) = E[X(X-1)] + E(X)$ and thus $\text{Var}(X) = E[X(X-1)] + E(X) - \{E(X)\}^2$.

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=2}^{\infty} \frac{\lambda^2 e^{-\lambda} \lambda^{x-2}}{(x-2)!} = \lambda^2 \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = \lambda^2 \times 1 = \lambda^2.$$

$$\text{Thus } \text{Var}(X) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

[The probability generating function or moment generating function could also be used – this work is in Module 5.]

(i) (b) The binomial distribution with parameters n and p may be approximated by the Poisson distribution with parameter np if n is large and p is small. As a "rule of thumb", $\frac{1}{2} \leq np \leq 10$ gives an indication of how large n should be and how small p should be. (If $np > 10$, a Normal approximation to the binomial may be better.)

(ii) Let X = number of wrong calculations. We have $X \sim B(200, 0.0075)$.

$$P(X=1) = \binom{200}{1} (0.0075) (0.9925)^{199} = 200 \times 0.0075 \times 0.9925^{199} = 0.3353(2).$$

$$\begin{aligned} P(X=4) &= \binom{200}{4} (0.0075)^4 (0.9925)^{196} \\ &= \frac{200 \times 199 \times 198 \times 197}{4 \times 3 \times 2 \times 1} \times 0.0075^4 \times 0.9925^{196} = 0.0468(0). \end{aligned}$$

Solution continued on next page

(iii) We approximate using $X \sim \text{Poisson}(200 \times 0.0075 = 1.5)$. With this,

$$P(X = 1) = 1.5e^{-1.5} = 0.3347(0)$$

giving a percentage error of $\frac{100(0.33532 - 0.33470)}{0.33532} = 0.18\%$, and

$$P(X = 4) = \frac{e^{-1.5}(1.5)^4}{4!} = 0.0470(7)$$

giving a percentage error of $\frac{100(0.04707 - 0.04680)}{0.04680} = 0.58\%$.

[Note. These percentage errors might come out *slightly* differently if more accuracy is kept in the binomial and Poisson probabilities.]

Both approximations are remarkably accurate, with percentage errors well below 1%. The approximation for $X = 1$ (one wrong calculation) is the more accurate of the two. That approximation is an underestimate; the other is an overestimate.

Question 2

$$H \sim N(160, 16)$$

$$\begin{aligned} \text{(i)} \quad P(156 < H \leq 164) &= P\left(\frac{156-160}{4} < Z \leq \frac{164-160}{4}\right) \quad [\text{where } Z \sim N(0,1)] \\ &= P(-1 < Z \leq 1) = \Phi(1) - \Phi(-1) = \Phi(1) - \{1 - \Phi(1)\} \quad \text{by symmetry, i.e. } 2(\Phi(1)) - 1 \\ &= 0.6826. \quad [\Phi \text{ denotes the cdf of the standard Normal distribution as usual.}] \end{aligned}$$

$$P(H > 168) = P(Z > 2) = 1 - \Phi(2) = 0.0228.$$

- (ii) (a) The relevant portion of the Normal distribution of H is that beginning at 168, which corresponds to the value $Z = 2$. The median value m within this portion has $\frac{1}{2}(0.0228)$ probability above it, i.e. 0.0114, so $\Phi(m) = 1 - 0.0114 = 0.9886$, corresponding to $Z = 2.277$. The corresponding value of H is $\mu + \sigma Z$ which is $160 + (4 \times 2.277) = 169.1$ cm.

(b) $H = 170$ corresponds to $Z = \frac{10}{4} = 2.5$; we have $\Phi(2.5) = 0.9938$, so $1 - \Phi(2.5) = 0.0062$. Conditional on $H > 168$, the probability is $\frac{P(H > 170)}{P(H > 168)}$

$$= \frac{0.0062}{0.0228} = 0.272.$$

(iii) Mean height of 25 members $\sim N\left(169.5, \frac{1.352^2}{25}\right)$.

$$\begin{aligned} P(\text{mean} > 170) &= P\left(Z > \frac{170-169.5}{(1.352/5)}\right) = 1 - \Phi\left(\frac{0.5 \times 5}{1.352}\right) = 1 - \Phi(1.849) \\ &= 1 - 0.9677 = 0.0323. \end{aligned}$$

Question 3

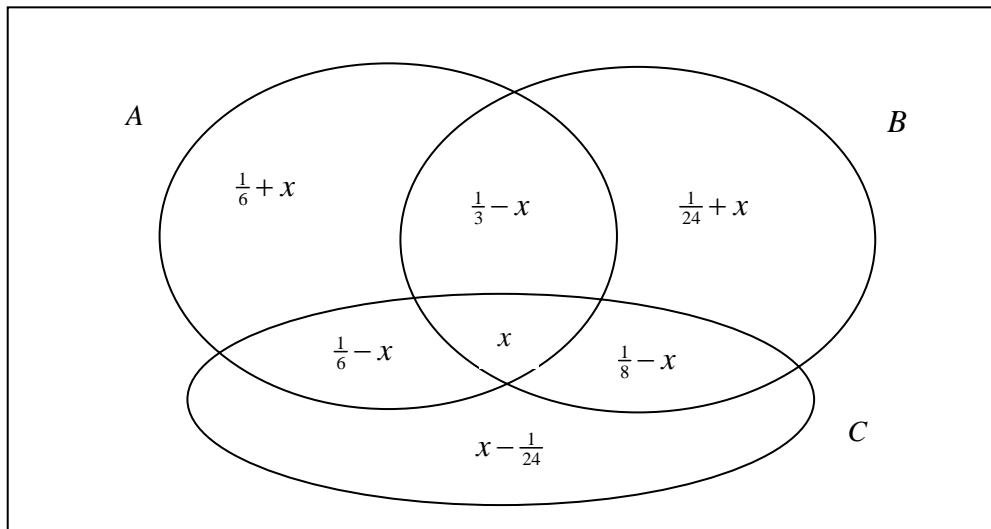
$$P(A) = \frac{2}{3} \quad P(B) = \frac{1}{2} \quad P(C) = \frac{1}{4}$$

(i) (a) By the given independence,

$$P(A \cap \bar{B} \cap \bar{C}) = P(A)P(\bar{B})P(\bar{C}) = \frac{2}{3} \left(1 - \frac{1}{2}\right) \left(1 - \frac{1}{4}\right) = \frac{1}{4}.$$

$$\begin{aligned} \text{(b)} \quad P(A \cap \bar{C} | A \cap \bar{B}) &= \frac{P((A \cap \bar{C}) \cap (A \cap \bar{B}))}{P(A \cap \bar{B})} = \frac{P(A \cap \bar{C} \cap \bar{B})}{P(A \cap \bar{B})} \\ &= \frac{\frac{1}{4}}{\frac{2}{3} \left(1 - \frac{1}{2}\right)} = \frac{3}{4}. \end{aligned}$$

(ii)



Using pairwise independence, the value of $P(A \cap B)$ is $P(A)P(B)$, etc, and hence the values $\frac{1}{3} - x$, $\frac{1}{8} - x$ and $\frac{1}{6} - x$ are found. The others follow using $P(A)$, $P(B)$ and $P(C)$.

$$\text{(a)} \quad P(A \cap \bar{B} \cap \bar{C}) = \frac{1}{6} + x \quad \text{from the diagram.}$$

$$\begin{aligned} \text{(b)} \quad P(A \cap \bar{C} | A \cap \bar{B}) &= \frac{P((A \cap \bar{C}) \cap (A \cap \bar{B}))}{P(A \cap \bar{B})} = \frac{P(A \cap \bar{C} \cap \bar{B})}{P(A \cap \bar{B})} \\ &= \frac{\frac{1}{6} + x}{\left(\frac{1}{6} + x\right) + \left(\frac{1}{6} - x\right)} = 3x + \frac{1}{2}. \end{aligned}$$

$$\text{(c)} \quad P(A \cup B \cup C) = \frac{19}{24} + x.$$

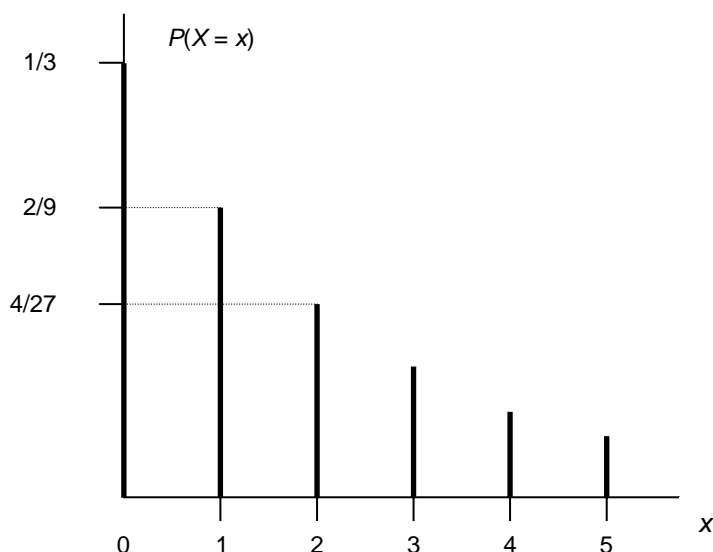
Since all probabilities must lie in $[0,1]$, we have $x \geq \frac{1}{24}$ and $x \leq \frac{1}{8}$, i.e.

$$\frac{1}{24} \leq x \leq \frac{1}{8}.$$

Question 4

Probability mass function: $f(x) = (1-p)^x p$, $x = 0, 1, 2, \dots$, $0 < p < 1$.

(i)



(ii) $E(X) = \sum_{x=0}^{\infty} x(1-p)^x p$. This can be found in several ways; the method shown here (using geometric series) is perhaps not the most efficient, but should be easy to understand. We have

$$\begin{aligned} \sum_{x=0}^{\infty} x(1-p)^x p &= p \{ 0 + (1-p) + 2(1-p)^2 + 3(1-p)^3 + \dots \} \\ &= p \{ 0 + (1-p) + (1-p)^2 + (1-p)^3 + \dots \\ &\quad + (1-p)^2 + (1-p)^3 + \dots \\ &\quad + (1-p)^3 + \dots \\ &\quad + \dots \} \\ &= p \left\{ \frac{1-p}{1-(1-p)} + \frac{(1-p)^2}{1-(1-p)} + \frac{(1-p)^3}{1-(1-p)} + \dots \right\} = \frac{p}{p} \left\{ \frac{1-p}{1-(1-p)} \right\} = \frac{1-p}{p}. \end{aligned}$$

Similarly, $E(X^2) = \sum_{x=0}^{\infty} x^2 (1-p)^x p = \frac{(1-p)^2 + (1-p)}{p^2}$.

Thus $\text{Var}(X) = E(X^2) - \{E(X)\}^2 = \frac{(1-p)^2 + (1-p)}{p^2} - \left(\frac{1-p}{p} \right)^2 = \frac{1-p}{p^2}$.

[The probability generating function or moment generating function could also be used – this work is in Module 5.]

Solution continued on next page

$$(iii) \quad P(X \geq x) = \sum_{r=x}^{\infty} (1-p)^r p \quad [\text{geometric series}] = \frac{(1-p)^x p}{1-(1-p)} = (1-p)^x \quad (\text{for } x =$$

0, 1, 2, ...). We now use $P(A|B) = \frac{P(A \cap B)}{P(B)}$ and take the event A as " $X \geq l + m$ "

and the event B as " $X \geq l$ ", so that $A \cap B = A$. Thus

$$P(X \geq l + m | X \geq l) = \frac{(1-p)^{l+m}}{(1-p)^l} = (1-p)^m = P(X \geq m).$$

This is the "lack of memory" property of a geometric distribution.

$$(iv) \quad \text{By independence, } P(Z \geq z) = P(X \geq z)P(Y \geq z) = (1-p)^z (1-\theta)^z.$$

$$\begin{aligned} P(Z = z) &= P(Z \geq z) - P(Z \geq z+1) = \{(1-p)(1-\theta)\}^z - \{(1-p)(1-\theta)\}^{z+1} \\ &= \{(1-p)(1-\theta)\}^z (1 - (1-p)(1-\theta)) = \{(1-p)(1-\theta)\}^z (p + \theta - p\theta), \text{ for } z = 0, 1, \dots \end{aligned}$$

This is a geometric distribution as given at the start of the question with p replaced by $p + \theta - p\theta$. Hence, from part (ii),

$$E(Z) = \frac{1-p-\theta+p\theta}{p+\theta-p\theta}, \quad \text{Var}(Z) = \frac{1-p-\theta+p\theta}{(p+\theta-p\theta)^2}.$$