THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE EXAMINATION

NEW MODULAR SCHEME

introduced from the examinations in 2007

MODULE 2

SPECIMEN PAPER A

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. In a hi-tech company, the members of three research groups (A, B and C) are individually invited to enter a prize competition for the best solution to a technical problem. Group A has 2 staff, B has 3 and C has 5. It is assumed that all staff decide independently whether or not to enter. Members of groups A, B and C enter with respective probabilities 1/2, 1/4 and 1/5.

(i) For each group separately, find the probability of (a) no entries, (b) one entry.

(8)

(ii) Given that there is just one entry in total, show that the probability that it comes from a member of group A is 8/17.

(6)

(iii) Explain (but without doing the calculations) the steps that are needed to calculate the probability that there are exactly two entries in total.

(6)

2. The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} kx^2(1-x)^2, & 0 \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(i) Find k and sketch the graph of $f(x)$.

(7)

(ii) Find $E(X)$ and $\mathrm{Var}(X)$, and show that $P\left(X \le \frac{1}{3}\right) = \frac{17}{81}$.

(8)

(iii) A random sample of size 5 is taken from this distribution. Find, correct to 4 decimal places, the probability that all 5 observations exceed 1/3.

(3)

(iv) Find, correct to 4 decimal places, the variance of the mean of a random sample of size 5.

(2)

3. My cycle journey to work is 3 km, and my cycling time (in minutes) if there are no delays is distributed N(15, 1), i.e. Normally with mean 15 and variance 1.

(i) Find the probability that, if there are no delays, I get to work in at most 17 minutes.

(2)

(ii) On my route there are three sets of traffic lights. Each time I meet a red traffic light, I am delayed by a random time that is distributed N(0.7, 0.09). These lights operate independently. Find the probability of my getting to work in at most 17 minutes

(a) if just one light is set at red when I reach it,

(b) if just two lights are set at red when I reach them,

(c) if all three lights are set at red when I reach them.

(9)

(iii) Suppose that, for each set of lights, the chance of delay is 0.5. Deduce that the mean value of T, my total journey time, is 16.05 minutes.

(4)

(iv) Given that Var(T) = 1.5025, use a suitable approximation to calculate the probability that, over 10 journeys, my average journey time to work is at most 17 minutes.

(5)

4. The random variable X follows the binomial B(n, p) distribution with probability mass function

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad , \quad x = 0, 1, ..., n \quad , \quad 0 < p < 1 \quad ,$$

where $q = 1 - p$. Show that $E(X) = np$ and $\text{Var}(X) = npq$.

(5)

A mathematics class in a school is divided into set A with 12 students and set B with 25 students. Both groups are given a test consisting of 16 short questions. For any student in set A, the score (that is, the number of correct answers) is distributed as B(16, 0.75); for any student in set B, the score is distributed as B(16, 0.5). All students answer independently.

(i) Find the probability that

(a) a given set A student gets all 16 questions right,

(3)

(b) at least one student in set A gets all 16 questions right.

(3)

(ii) Use an appropriate approximation to find the probability that a given set B student scores more than a given set A student.

(5)

(iii) Let $\overline{X}$ and $\overline{Y}$ denote the mean scores of students in set A and set B respectively. Write down $E(\overline{X})$ and $E(\overline{Y})$, and show that $\text{Var}(\overline{X}) = 1/4$ and $\text{Var}(\overline{Y}) = 4/25$.

(4)

SOLUTIONS

Question 1

(i) A: (a) $P(0 \text{ entries}) = \left(\frac{1}{2}\right)^2 = \frac{1}{4} = 0.25$.

(b) $P(1 \text{ entry}) = 2 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2} = 0.5$.

B: (a) $P(0 \text{ entries}) = \left(\frac{3}{4}\right)^3 = \frac{27}{64} = 0.4219$.

(b) $P(1 \text{ entry}) = 3 \times \frac{1}{4} \times \left(\frac{3}{4}\right)^2 = \frac{27}{64} = 0.4219$.

C: (a) $P(0 \text{ entries}) = \left(\frac{4}{5}\right)^5 = \frac{1024}{3125} = 0.3277$.

(b) $P(1 \text{ entry}) = 5\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^4 = \frac{256}{625} = 0.4096$.

(ii) P(1 entry in total)

= P(1 from A, 0 from B and C) + P(1 from B, 0 from A and C)
+ P(1 from C, 0 from A and B)

$$= \frac{1}{2} \times \frac{27}{64} \times \frac{1024}{3125} + \frac{27}{64} \times \frac{1}{4} \times \frac{1024}{3125} + \frac{256}{625} \times \frac{1}{4} \times \frac{27}{64} = \frac{459}{3125}.$$

[If worked in decimals, this is 0.1469.]

P(1 from A | 1 in total) = P(1 from A and 1 in total) / P(1 in total)

= P(1 from A, 0 from B and C) / P(1 in total)

$$= \frac{\frac{1}{2} \times \frac{27}{64} \times \frac{1024}{3125}}{\frac{459}{3125}} = \frac{8}{17}.$$

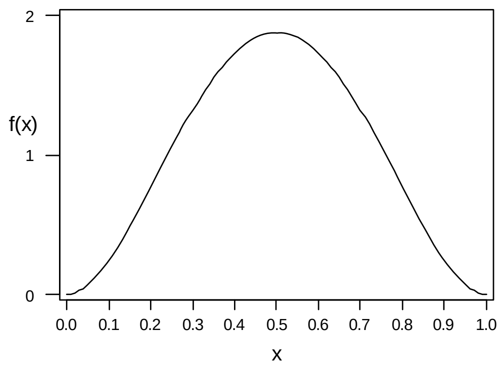
(iii) Denote the numbers of entries from A, B, C as (0, 0, 0) etc. Then we need P(2, 0, 0) + P(0, 2, 0) + P(0, 0, 2) + P(1, 1, 0) + P(1, 0, 1) + P(0, 1, 1). Since entries from each group are independent, we have, as an example, P(1, 1, 0) = P(1 from A).P(1 from B).P(0 from C).

<u>Question 2</u>

(i) We have $k\int_0^1 x^2(1-x)^2 dx = 1$, so $k\int_0^1 \left(x^2 - 2x^3 + x^4\right)dx = 1$. This gives

$$1 = k\left[\frac{1}{3}x^3 - \frac{1}{2}x^4 + \frac{1}{5}x^5\right]_0^1 = k\left(\frac{1}{3} - \frac{1}{2} + \frac{1}{5}\right), \quad \text{so } k = 30.$$

$f(x) = 0$ at $x = 0$ and at $x = 1$. $f(x)$ is symmetrical about $x = ½$. The sketch is as follows.



(ii) $E(X) = \frac{1}{2}$ by symmetry [or by direct integration: $\int_0^1 xf(x)dx$].

$$E(X^2) = 30\int_0^1 x^4(1-x)^2 dx = 30\int_0^1 \left(x^4 - 2x^5 + x^6\right)dx$$

$$= 30\left[\frac{1}{5}x^5 - \frac{1}{3}x^6 + \frac{1}{7}x^6\right]_0^1 = 30\left(\frac{1}{5} - \frac{1}{3} + \frac{1}{7}\right) = 30 \times \frac{1}{105} = \frac{2}{7}.$$

$$\therefore \text{Var}(X) = E(X^2) - \{E(X)\}^2 = \frac{2}{7} - \left(\frac{1}{2}\right)^2 = \frac{1}{28}.$$

$$P\left(X \le \frac{1}{3}\right) = \int_0^{1/3} 30\left(x^2 - 2x^3 + x^4\right)dx = 30\left[\frac{1}{3}x^3 - \frac{1}{2}x^4 + \frac{1}{5}x^5\right]_0^{1/3}$$

$$= 30\left(\frac{1}{3^4} - \frac{1}{2}.\frac{1}{3^4} + \frac{1}{5}.\frac{1}{3^5}\right) = \frac{30}{81}\left(1 - \frac{1}{2} + \frac{1}{15}\right) = \frac{30}{81} \times \frac{17}{30} = \frac{17}{81} \quad (= 0.2099).$$

(iii) The required probability is $\left(1 - \frac{17}{81}\right)^5 = \left(\frac{64}{81}\right)^5 = 0.3079.$

(iv) The variance of $\bar{X}$ for a sample of size 5 is $\frac{\text{Var}(X)}{5} = \frac{1/28}{5} = \frac{1}{140} = 0.00714.$

Question 3

Let X represent cycling time without delays: $X \sim N(15, 1)$.

(i) $P(X \le 17) = \Phi\left(\frac{17-15}{1}\right) = \Phi(2) = 0.9772$.

[Φ denotes the cdf of the standard Normal distribution as usual.]

(ii) Adding in the delay times, also Normally distributed [N(0.7, 0.09)], and letting T denote the total time:

(a) $T \sim N(15.7, 1.09)$, so $P(T \le 17) = \Phi\left(\frac{17-15.7}{\sqrt{1.09}}\right) = \Phi(1.245) = 0.8934$;

(b) $T \sim N(16.4, 1.18)$, so $P(T \le 17) = \Phi\left(\frac{17-16.4}{\sqrt{1.18}}\right) = \Phi(0.552) = 0.7096$;

(c) $T \sim N(17.1, 1.27)$, so $P(T \le 17) = \Phi\left(\frac{17-17.1}{\sqrt{1.27}}\right) = \Phi(-0.0887) = 0.4646$.

(iii) The number of delays is distributed as B(3, ½). Hence the situations in (i), (ii)(a), (ii)(b) and (ii)(c) arise with probabilities 1/8, 3/8, 3/8 and 1/8 respectively, so the (unconditional) mean of the total journey time is

$$E(T) = \frac{1}{8}\times 15 + \frac{3}{8}\times 15.7 + \frac{3}{8}\times 16.4 + \frac{1}{8}\times 17.1 = \frac{128.4}{8} = 16.05 \text{ minutes.}$$

(iv) Mean time $\bar{T} \sim N\left(16.05, \frac{1.5025}{10}\right)$.

$$P(\bar{T} \le 17) = \Phi\left(\frac{17-16.05}{\sqrt{0.15025}}\right) = \Phi(2.451) = 0.9929.$$

Question 4

An easy method is to consider X as $\sum X_i$, where X_i are a set of n Bernoulli variables with $P(X_i = 1) = p, P(X_i = 0) = (1-p)$. Then $E[X_i] = p,$ so $E[X] = np$.

Also $E\left[X_i^2\right] = p,$ so $\operatorname{Var}(X_i) = p - p^2$ and $\operatorname{Var}(X) = n\left(p - p^2\right) = npq.$

ALTERNATIVELY: $E[X] = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = \sum_{x=1}^{n} \frac{n!\, p^x q^{n-x}}{(x-1)!(n-x)!}$

$$= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{(n-1)-(x-1)} = np\,.$$

Similarly, $\operatorname{Var}(\mathrm{X}) = E\left[X(X-1)\right] + E[X] - \left(E[X]\right)^2$, and we have

$$E\left[X(X-1)\right] = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x q^{n-x} = \sum_{x=2}^{n} x(x-1)\binom{n}{x} p^x q^{n-x}$$

$$= n(n-1)\,p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{(n-2)-(x-2)} = n(n-1)\,p^2,$$

and hence $\operatorname{Var}(\mathrm{X}) = n(n-1)\,p^2 + np - n^2 p^2 = np - np^2 = npq$.

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

(i) (a) $0.75^{16} \approx 0.0100226 = 0.0100$ approx.

(b) 1 – P(no one gets all 16 right), probability is $1 - \left\{1 - 0.75^{16}\right\}^{12}$

$= 1 - \{0.9899774\}^{12} = 0.1139$.

Solution continued on next page

(ii) $P(B - A > 0)$ can be studied using a Normal approximation to the difference $B - A$, i.e. $\mathrm{N}\left[(16\times0.5)-(16\times0.75)\,,\;\;(16\times0.5\times0.5)+(16\times0.75\times0.25)\right]$, i.e. N(–4, 7).

The probability is then found using this approximation and a continuity correction (since $B - A$ takes discrete values) as $P\left(B-A>0.5\right)$.

Hence it is found as

P(N(–4, 7) > 0.5) = 1 – P(N(–4, 7) < 0.5)

$$= 1-\Phi\left(\frac{0.5-(-4)}{\sqrt{7}}\right)=1-\Phi\left(\frac{4.5}{\sqrt{7}}\right)=1-\Phi\left(1.7008\right)\approx 0.0445\,.$$

[Φ denotes the cdf of the standard Normal distribution as usual.]

[Note: this would be 0.0653 without the continuity correction.]

(iii) $E\left[\,\bar{X}\,\right]=E\left[X\right]=np=16\times0.75=12$ in set A.

Similarly, $E\left[\bar{Y}\right]=16\times0.5=8$ in set B.

There are 12 students in A and 25 in B, so that

$$\mathrm{Var}\left(\bar{X}\right)=\frac{16\times0.75\times0.25}{12}=\frac{1}{4}\quad\text{in set } A$$

$$\mathrm{Var}\left(\bar{Y}\right)=\frac{16\times0.5\times0.5}{25}=\frac{4}{25}\quad\text{in set } B.$$