SL Paper 2

The random variable X has cumulative distribution function

$$F(x)=\begin{cases}0 & x<0\\ \left(\frac{x}{a}\right)^3 & 0\leqslant x\leqslant a\\ 1 & x>a\end{cases}$$

where a is an unknown parameter. You are given that the mean and variance of X are $\frac{3a}{4}$ and $\frac{3a^2}{80}$ respectively. To estimate the value of a , a random sample of n independent observations, $X_1, X_2, \ldots X_n$ is taken from the distribution of X .

a. (i) Find an expression for c in terms of n such that $U=c\sum_{i=1}^{n}X_i$ is an unbiased estimator for a . [4]

(ii) Determine an expression for $\mathrm{Var}(U)$ in this case.

b. (i) Show that $\mathrm{P}(Y\leq y)=\left(\frac{y}{a}\right)^{3n}, 0\leq y\leq a$ and deduce an expression for the probability density function of Y . [16]

(ii) Find $\mathrm{E}(Y)$.

(iii) Show that $\mathrm{Var}(Y)=\frac{3na^2}{(3n+2)(3n+1)^2}$.

(iv) Find an expression for d in terms of n such that $V=dY$ is an unbiased estimator for a .

(v) Determine an expression for $\mathrm{Var}(V)$ in this case.

c. Show that $\frac{\mathrm{Var}(U)}{\mathrm{Var}(V)}=\frac{3n+2}{5}$ and hence state, with a reason, which of U or V is the more efficient estimator for a . [2]

Markscheme

a. (i) $\mathrm{E}(U)=c\times n\times\frac{3a}{4}=a\Rightarrow c=\frac{4}{3n}$ ***M1A1***

(ii) $Var(U)=\frac{16}{9n^2}\times n\times\frac{3a^2}{80}=\frac{a^2}{15n}$ ***M1A1***

[4 marks]

b. (i) $\mathrm{P}(Y\leq y)=\mathrm{P}(\text{all } Xs\leq y)$ ***M1***

$=[\mathrm{P}\ (X\leq y)]^n$ ***(A1)***

$=\left(\left(\frac{y}{a}\right)^3\right)^n$ ***(A1)***

Note: Only one of the two ***A1*** marks above may be implied.

$=\left(\frac{y}{a}\right)^{3n}$ ***AG***

$g(y)=\frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{y}{a}\right)^{3n}=\frac{3ny^{3n-1}}{a^{3n}}, (0<y<a)$ ***M1A1***

$(g(y)=0\text{ otherwise})$

(ii) $\mathrm{E}(Y) = \int_0^a \frac{3ny^{3n}}{a^{3n}}\mathrm{d}y$ ***M1***

$= \left[\frac{3ny^{3n+1}}{(3n+1)a^{3n}}\right]_0^a$ ***A1***

$= \frac{3na}{3n+1}$ ***A1***

(iii) $\mathrm{Var}(Y) = \int_0^a \frac{3ny^{3n+1}}{a^{3n}}\mathrm{d}y - \left(\frac{3na}{3n+1}\right)^2$ ***M1***

$= \left[\frac{3ny^{3n+2}}{(3n+2)a^{3n}}\right]_0^a - \left(\frac{3na}{3n+1}\right)^2$ ***A1***

$= \frac{3na^2}{3n+2} - \frac{9n^2a^2}{(3n+1)^2}$ ***M1***

$= \frac{3na^2(9n^2+6n+1)-9n^2a^2(3n+2)}{(3n+2)(3n+1)}$ ***A1***

$= \frac{3na^2}{(3n+2)(3n+1)^2}$ ***AG***

(iv) $\mathrm{E}(V) = d \times \frac{3na}{3n+1} = a \Rightarrow d = \frac{3n+1}{3n}$ ***M1A1***

(v) $Var(V) = \left(\frac{3n+1}{3n}\right)^2 \times \frac{3na^2}{(3n+2)(3n+1)^2}$ ***M1***

$= \frac{a^2}{3n(3n+2)}$ ***A1***

[16 marks]

c. $\frac{\mathrm{Var}(U)}{\mathrm{Var}(V)} = \frac{\frac{a^2}{15n}}{\frac{a^2}{3n(3n+2)}}$ ***A1***

$= \frac{3n+2}{5}$ ***AG***

V is the more efficient estimator because $3n + 2 > 5$ (for $n > 1$) ***R1***

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]

Gillian is throwing a ball at a target. The number of throws she makes before hitting the target follows a geometric distribution, $X \sim \mathrm{Geo}(p)$. When she uses a cricket ball the number of throws she makes follows a geometric distribution with $p = \frac{1}{4}$. When she uses a tennis ball the number of throws she makes follows a geometric distribution with $p = \frac{3}{4}$. There is a box containing a large number of balls, 80% of which are cricket balls and the remainder are tennis balls. The random variable A is the number of throws needed to hit the target when a single ball is chosen at random from this box and used for all throws.

a. Find $\mathrm{E}(A)$. [4]

b. Show that $\mathrm{P}(A=r)=\frac{1}{5}\times\left(\frac{3}{4}\right)^{r-1}+\frac{3}{20}\times\left(\frac{1}{4}\right)^{r-1}$. [4]

c. Find $\mathrm{P}(A\leqslant 5|A>3)$. [7]

Markscheme

a. $\mathrm{E}(X\ tennis)=\frac{1}{\frac{3}{4}}=\frac{4}{3}$ ***(A1)***

$\mathrm{E}(X\ cricket)=\frac{1}{\frac{1}{4}}=4$ ***(A1)***

$\mathrm{E}(A)=\frac{4}{3}\times\frac{1}{5}+4\times\frac{4}{5}=\frac{52}{15}$ ***M1A1***

b. $\mathrm{P}(X=r|cricket)=\frac{1}{4}\cdot\left(\frac{3}{4}\right)^{r-1}$ ***(A1)***

$\mathrm{P}(X=r|tennis)=\frac{3}{4}\cdot\left(\frac{1}{4}\right)^{r-1}$ ***(A1)***

$\mathrm{P}(A=r)=\mathrm{P}(X=r|cricket)\times\mathrm{P}(cricket)+\mathrm{P}(X=r|tennis)\times\mathrm{P}(tennis)$ ***(M1)***

$=\frac{1}{4}\times\left(\frac{3}{4}\right)^{r-1}\times\frac{4}{5}+\frac{3}{4}\times\left(\frac{1}{4}\right)^{r-1}\times\frac{1}{5}$ ***A1***

$=\frac{1}{5}\times\left(\frac{3}{4}\right)^{r-1}+\frac{3}{20}\times\left(\frac{1}{4}\right)^{r-1}$ ***AG***

c. $\mathrm{P}(A\leqslant 5)|(A<3)=\frac{\mathrm{P}(A=4\text{ or }5)}{1-\mathrm{P}(A=1\text{ or }2\text{ or }3)}$ ***M1A1A1***

$\mathrm{P}(A=1)=\frac{7}{20}$

$\mathrm{P}(A=2)=\frac{15}{80}$

$\mathrm{P}(A=3)=\frac{39}{320}$

$\mathrm{P}(A=4)=\frac{111}{1280}$

$\mathrm{P}(A=5)=\frac{327}{5120}$

$\Rightarrow\mathrm{P}(A>3)|(A\leqslant 5)=\frac{\frac{771}{5120}}{\frac{1744}{5120}}$ ***A1A2***

Note: Award ***A1*** for correct working for numerator and Award ***A2*** for correct working for denominator

$=\frac{771}{1744}\quad(=0.442)$ ***A1***

Examiners report

a. It was pleasing to see many correct answers to parts a) and b) with candidates correctly recognising how to work with the distribution.

b. It was pleasing to see many correct answers to parts a) and b) with candidates correctly recognising how to work with the distribution.

c. Part c) caused more problems. Although a number of wholly correct solutions were seen, many candidates were unable to work meaningfully with the conditional probability.

The discrete random variable X has the following probability distribution.

$\mathrm{P}(X=x)=\frac{kx}{3^x}$, where $x\in\mathbb{Z}^+$ and k is a constant.

a.i. Write down the first three terms of the infinite series for $G(t)$, the probability generating function for X. [2]

a.ii. Determine the radius of convergence of this infinite series. [4]

a.iii. By considering $\left(1-\frac{t}{3}\right)G(t)$, show that [3]

$$G(t)=\frac{3kt}{(3-t)^2}.$$

a.iv. Hence show that $k=\frac{4}{3}$. [1]

b.i. Show that $\ln G(t)=\ln 4+\ln t-2\ln(3-t)$. [1]

b.ii. By differentiating both sides of this equation, determine the values of $G'(1)$ and $G''(1)$. [6]

b.iii. Hence find $\mathrm{Var}(X)$. [1]

Markscheme

a.i. $G(t)=\frac{kt}{3}+\frac{2kt^2}{3^2}+\frac{3kt^3}{3^3}+\ldots$ ***M1A1***

[??? marks]

a.ii. $\frac{T_{n+1}}{T_n}=\frac{(n+1)kt^{n+1}}{3^{n+1}}\times\frac{3^n}{nkt^n}$ ***M1A1***

$\rightarrow\frac{t}{3}$ as $n\rightarrow\infty$ ***A1***

for convergence, $\left|\frac{t}{3}\right|<1$ so radius of convergence $=3$ ***A1***

[??? marks]

a.iii. $G(t)=\frac{kt}{3}+\frac{2kt^2}{3^2}+\frac{3kt^3}{3^3}+\ldots$

$\frac{t}{3}G(t)=\frac{kt^2}{3^2}+\frac{2kt^3}{3^3}+\ldots$

$\left(1-\frac{t}{3}\right)G(t)=\frac{kt}{3}+\frac{kt^2}{3^2}+\frac{kt^3}{3^3}+\ldots$ ***M1A1***

$=\frac{\frac{kt}{3}}{\left(1-\frac{t}{3}\right)}$ ***A1***

$G(t)=\frac{\frac{kt}{3}}{\left(1-\frac{t}{3}\right)^2}=\frac{3kt}{(3-t)^2}$ ***AG***

[??? marks]

a.iv. $G(1)=1$ ***M1***

so $k=\frac{4}{3}$ ***AG***

[??? marks]

b.i. $\ln G(t)=\ln 4t-\ln{(3-t)}^2$ ***M1***

$\ln G(t) = \ln 4 + \ln t - 2\ln(3-t)$ ***AG***

[??? marks]

b.ii. $\frac{G'(1)}{G(1)} = \frac{1}{t} + \frac{2}{3-t}$ ***M1A1***

putting $t = 1$

$G'(1) = 2$ ***A1***

$\frac{G''(t)G(t)-[G'(t)]^2}{[G(t)]^2} = -\frac{1}{t^2} + \frac{2}{(3-t)^2}$ ***M1A1***

putting $t = 1$

$G''(1) - 4 = -1 + \frac{1}{2}$

$G''(1) = \frac{7}{2}$ ***A1***

[??? marks]

b.iii. $\mathrm{Var}(X) = G''(1) + G'(1) - [G'(1)]^2 = \frac{3}{2}$ ***A1***

[??? marks]

Examiners report

a.i. [N/A]
a.ii. [N/A]
a.iii. [N/A]
a.iv. [N/A]
b.i. [N/A]
b.ii. [N/A]
b.iii. [N/A]

A machine fills containers with grass seed. Each container is supposed to weigh 28 kg. However the weights vary with a standard deviation of 0.54 kg. A random sample of 24 bags is taken to check that the mean weight is 28 kg.

A.a. Assuming the series for e^x , find the first five terms of the Maclaurin series for [3]

$$\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}.$$

A.b. (i) Use your answer to (a) to find an approximate expression for the cumulative distributive function of $\mathrm{N}(0,1)$. [6]

(ii) **Hence** find an approximate value for $\mathrm{P}(-0.5 \le Z \le 0.5)$, where $Z \sim \mathrm{N}(0,1)$.

B.a. State and justify an appropriate test procedure giving the null and alternate hypotheses. [5]

B.b. What is the critical region for the sample mean if the probability of a Type I error is to be 3.5%? [7]

B.c. If the mean weight of the bags is actually 28.1 kg, what would be the probability of a Type II error? [2]

Markscheme

A.a $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots$

$e^{\frac{-x^2}{2}} = 1 + \left(-\frac{x^2}{2}\right) + \frac{\left(-\frac{x^2}{2}\right)^2}{2!} + \frac{\left(-\frac{x^2}{2}\right)^3}{3!} + \frac{\left(-\frac{x^2}{2}\right)^4}{4!} + \ldots$ ***M1A1***

$\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} = \frac{1}{\sqrt{2\pi}}\left(1 - \frac{x^2}{2} + \frac{x^4}{8} - \frac{x^6}{48} + \frac{x^8}{384}\right)$ ***A1***

[3 marks]

A.b (i) $\frac{1}{\sqrt{2\pi}} \int_0^x 1 - \frac{t^2}{2} + \frac{t^4}{8} - \frac{t^6}{48} + \frac{t^8}{384} \mathrm{d}t$ ***M1***

$= \frac{1}{\sqrt{2\pi}}\left(x - \frac{x^3}{6} + \frac{x^5}{40} - \frac{x^7}{336} + \frac{x^9}{3456}\right)$ ***A1***

$\mathrm{P}(Z \le x) = 0.5 + \frac{1}{\sqrt{2\pi}}\left(x - \frac{x^3}{6} + \frac{x^5}{40} - \frac{x^7}{336} + \frac{x^9}{3456} - \ldots\right)$ ***R1A1***

(ii) $\mathrm{P}(-0.5 \le Z \le 0.5) = \frac{2}{\sqrt{2\pi}}\left(0.5 - \frac{0.5^3}{6} + \frac{0.5^5}{40} - \frac{0.5^7}{336} + \frac{0.5^9}{3456} - \ldots\right)$ ***M1***

$= 0.38292 = 0.383$ ***A1***

[6 marks]

B.a this is a two tailed test of the sample mean $\overline{x}$

we use the central limit theorem to justify assuming that ***R1***

$\overline{X} \sim \mathrm{N}\left(28, \frac{0.54^2}{24}\right)$ ***R1A1***

$\mathrm{H}_0 : \mu = 28$ ***A1***

$\mathrm{H}_1 : \mu \ne 28$ ***A1***

[5 marks]

B.b since $\mathrm{P}(\text{Type I error}) = 0.035$, critical value 2.108 ***(M1)A1***

and ($\overline{x} \le 28 - 2.108\sqrt{\frac{0.54^2}{24}}$ or $\overline{x} \ge 28 + 2.108\sqrt{\frac{0.54^2}{24}}$) ***(M1)(A1)(A1)***

$\overline{x} \le 27.7676$ or $\overline{x} \ge 28.2324$

so $\overline{x} \le 27.8$ or $\overline{x} \ge 28.2$ ***A1A1***

[7 marks]

B.c if $\mu = 28.1$

$\overline{X} \sim \mathrm{N}\left(28.1, \frac{0.54^2}{24}\right)$ ***R1***

$\mathrm{P}(\text{Type II error}) = \mathrm{P}(27.7676 < \overline{x} < 28.2324)$

$= 0.884$ ***A1***

Note: Depending on the degree of accuracy used for the critical region the answer for part (c) can be anywhere from 0.8146 to 0.879.

[2 marks]

Examiners report

A.a The derivation of a series from a given one by substitution seems not to be well known. This made finding the required series from (e^x) in part (a) to be much more difficult than it need have been. The fact that this part was worth only 3 marks was a clear hint that an easy derivation was possible.

A.b In part (b)(i) the 0.5 was usually missing which meant that this part came out incorrectly.

B.a The conditions required in part (a) were rarely stated correctly and some candidates were unable to state the hypotheses precisely. There was some confusion with "less than" and "less than or equal to".

B.b There was some confusion with "less than" and "less than or equal to".

B.c Levels of accuracy in the body of the question varied wildly leading to a wide range of answers to part (c).

---

a. The weights, X grams, of tomatoes may be assumed to be normally distributed with mean μ grams and standard deviation σ grams. Barry weighs 21 tomatoes selected at random and calculates the following statistics. [8]

$$\sum x = 1071\backslash);\backslash(\sum x^2 = 54705$$

(i) Determine unbiased estimates of μ and σ^2 .

(ii) Determine a 95% confidence interval for μ .

b. The random variable Y has variance σ^2 , where $\sigma^2 > 0$. A random sample of n observations of Y is taken and S_{n-1}^2 denotes the unbiased estimator for σ^2 . [5]

By considering the expression

$$\text{Var}(S_{n-1}) = \text{E}(S_{n-1}^2) - \{E\ (S_{n-1})\}^2 ,$$

show that S_{n-1} is not an unbiased estimator for σ .

Markscheme

a. (i) $\overline{x} = \frac{1071}{21} = 51$ ***A1***

$S_{n-1}^2 = \frac{54705}{20} - \frac{1071^2}{20\times 21} = 4.2$ ***M1A1***

(ii) degrees of freedom $= 20$; t-value $= 2.086$ ***(A1)(A1)***

95% confidence limits are

$51 \pm 2.086\sqrt{\frac{4.2}{21}}$ ***(M1)(A1)***

leading to $[50.1, 51.9]$ ***A1***

[8 marks]

b. $\text{Var}(S_{n-1}) > 0$ ***A1***

$E(S^2_{n-1}) = \sigma^2$ ***(A1)***

substituting in the given equation,

$\sigma^2 - E(S_{n-1}) > 0$ ***M1***

it follows that

$E(S_{n-1}) < \sigma$ ***A1***

this shows that S_{n-1} is not an unbiased estimator for σ since that would require = instead of $<$ ***R1***

[5 marks]

Examiners report

a. Most candidates attempted (a) although some used the normal distribution instead of the t-distribution.

b. Many candidates were unable even to start (b) and many of those who did filled several pages of algebra with factors such as n / $(n-1)$ prominent. Few candidates realised that the solution required only a few lines.

The weights of apples, in grams, produced on a farm may be assumed to be normally distributed with mean μ and variance σ^2 .

a. The farm manager selects a random sample of 10 apples and weighs them with the following results, given in grams. [5]

$$82, 98, 102, 96, 111, 95, 90, 89, 99, 101$$

(i) Determine unbiased estimates for μ and σ^2 .

(ii) Determine a 95% confidence interval for μ .

b. The farm manager claims that the mean weight of apples is 100 grams but the buyer from the local supermarket claims that the mean is less than this. To test these claims, they select a random sample of 100 apples and weigh them. Their results are summarized as follows, where x is the weight of an apple in grams. [5]

$$\sum x = 9831; \sum x^2 = 972578$$

(i) State suitable hypotheses for testing these claims.

(ii) Determine the p-value for this test.

(iii) At the 1% significance level, state which claim you accept and justify your answer.

Markscheme

a. (i) from the GDC,

unbiased estimate for $\mu = 96.3$ ***A1***

unbiased estimate for $\sigma^2 = 8.028\ldots^2 = 64.5$ ***(M1)A1***

(ii) 95% confidence interval is [90.6, 102] ***A1A1***

Note: Accept 102.0 as the upper limit.

[5 marks]

b. (i) $H_0 : \mu = 100; H_1 : \mu < 100$ ***A1***

(ii) $\overline{x} = 98.31, S_{n-1} = 7.8446\ldots$ ***(A1)***

p-value $= 0.0168$ ***A1***

(iii) the farm manager's claim is accepted because $0.0168 > 0.01$ ***A1R1***

[5 marks]

Examiners report

a. [N/A]
b. [N/A]

The discrete random variable *X* follows a geometric distribution Geo(p) where

$$\mathrm{P}\left(X = x\right) = \begin{cases} pq^{x-1}, \text{ for } x = 1,\ 2\ldots \\ 0, \text{ otherwise} \end{cases}$$

a.i. Show that the probability generating function of *X* is given by [3]

$$G\left(t\right) = \frac{pt}{1 - qt}.$$

a.ii. Deduce that $\mathrm{E}\left(X\right) = \frac{1}{p}$. [4]

b. Two friends A and B play a ball game with the following rules. [5]

Each player starts with zero points. Player A serves first and then the players have alternate pairs of serves so that the service order is A, B, B, A, A, … When player A serves, the probability of her scoring 1 point is p_A and the probability of B scoring 1 point is q_A, where $q_A = 1 - p_A$.

When player B serves, the probability of her scoring 1 point is p_B and the probability of A scoring 1 point is q_B, where $q_B = 1 - p_B$.

Show that, after the first 6 serves, the probability that each player has 3 points is

$$\sum_{x=0}^{x=3} \begin{pmatrix} 3 \\ x \end{pmatrix}^2 (p_A)^x (p_B)^x (q_A)^{3-x} (q_B)^{3-x}.$$

c. After 6 serves the score is 3 points each. Play continues and the game ends when one player has scored two more points than the other player. [3]

Let *N* be the number of further serves required before the game ends. Given that p_A = 0.7 and p_A = 0.6 find P(*N* = 2).

d. Let $M = \frac{1}{2}N$. Show that *M* has a geometric distribution and hence find the value of E(*N*). [7]

Markscheme

a.i. $\mathrm{P}(X = x) = pq^{x-1}$, for $x = 1, 2\ldots$

$G(t) = \sum_{x=1}^{\infty} t^x pq^{x-1}$ ***M1***

$= pt\sum_{x=1}^{\infty} (tq)^{x-1}$ ***A1***

$= pt\left(1 + tq + (tq)^2 \ldots\right)$ ***M1***

$= \frac{pt}{1-tq}$ ***AG***

[3 marks]

a.ii. $G'(t) = \frac{(1-tq)p - pt(-q)}{(1-tq)^2}$ ***M1A1***

$\mathrm{E}(X) = G'(1)$ ***M1***

$= \frac{(1-q)p+pq}{(1-q)^2}$ ***A1***

$= \frac{1}{p}$ ***AG***

[4 marks]

b. after 6 serves (3 serves each) we have *ABBAAB*

A serves *B* serves

3 wins 0 losses $p_1 = {}^3C_3 p_A^3 q_A^0 \, {}^3C_0 p_B^3 q_B^0$ ***M1A1***

2 wins 1 loss $p_2 = {}^3C_2 p_A^2 q_A^1 \, {}^3C_1 p_B^2 q_B^1$ ***A1***

1 win 2 losses $p_3 = {}^3C_1 p_A^1 q_A^2 \, {}^3C_2 p_B^1 q_B^2$ ***A1***

0 wins 3 losses $p_4 = {}^3C_0 p_A^0 q_A^3 \, {}^3C_3 p_B^0 q_B^3$ ***A1***

since ${}^3C_0 = {}^3C_3,\ {}^3C_1 = {}^3C_2$

$\sum_{x=0}^{x=3} \binom{3}{x}^2 (p_A)^x (p_B)^x (q_A)^{3-x} (q_B)^{3-x}$ ***AG***

[5 marks]

c. for *N* = 2 serves are *B*, *A* respectively

P(*N* = 2) = P(*B* wins twice) + P(*A* wins twice) ***(M1)***

= 0.6 × 0.3 + 0.4 × 0.7 ***A1***

= 0.46 ***A1***

[3 marks]

d. for $M = \frac{1}{2}N$

$\mathrm{P}(M = 1) = \mathrm{P}(N = 2) = p_M$ ***M1***

$\mathrm{P}(M = 2) = \mathrm{P}(N = 4)$

$= \mathrm{P}\left(\begin{array}{c}\text{game does not end after}\\ \text{first two serves}\end{array}\right) \times \mathrm{P}\left(\begin{array}{c}\text{game ends after}\\ \text{next two serves}\end{array}\right) = (1 - p_M)\, p_M$ ***A1***

similarly $\mathrm{P}(M = 3) = (1 - p_M)^2 p_M$ ***(A1)***

hence $\mathrm{P}(M = r) = (1 - p_M)^{r-1} p_M$ ***A1***

hence *M* has a geometric distribution ***AG***

$\mathrm{P}(M = 1) = \mathrm{P}(N = 2) = p_M = 0.46$ ***A1*** hence $\mathrm{E}(M) = \frac{1}{p} = \frac{1}{0.46} = 2.174$

$\mathrm{E}(N) = \mathrm{E}(2M) = 2\mathrm{E}(M)$ ***M1***

= 4.35 ***A1***

[7 marks]

Examiners report

a.i. [N/A]
a.ii. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]

The independent random variables X and Y are given by $X \sim \mathrm{N}\left(\mu_1,\ \sigma_1^2\right)$ and $Y \sim \mathrm{N}\left(\mu_2,\ \sigma_2^2\right)$.

Two independent random variables X_1 and X_2 each have a normal distribution with a mean 3 and a variance 9. Four independent random variables Y_1, Y_2, Y_3, Y_4 each have a normal distribution with mean 2 and variance 25. Each of the variables Y_1, Y_2, Y_3, Y_4 is independent of each of the variables X_1, X_2. Find

a. Write down the distribution of $aX + bY$ where $a, b \in \mathbb{R}$. [2]

b.i. $\mathrm{P}(X_1 + Y_1 < 11)$. [3]

b.ii. $\mathrm{P}(3X_1 + 4Y_1 > 15)$. [4]

b.iii. $\mathrm{P}(X_1 + X_2 + Y_1 + Y_2 + Y_3 + Y_4 < 30)$. [3]

c. Given that $\bar{X}$ and $\bar{Y}$ are the respective sample means, find $\mathrm{P}\left(\bar{X} > \bar{Y}\right)$. [5]

Markscheme

a. $aX + bY \sim \mathrm{N}\left(a\mu_1 + b\mu_2,\ a^2\sigma_1^2 + b^2\sigma_2^2\right)$ ***A1A1***

Note: ***A1*** for N and the mean, ***A1*** for the variance.

[2 marks]

b.i. $X_1 + Y_1 \sim \mathrm{N}(5,34)$ ***(A1)(A1)***

$\Rightarrow \mathrm{P}(X_1 + Y_1 < 11) = 0.848$ ***A1***

[3 marks]

b.ii. $3X_1 + 4Y_1 \sim \mathrm{N}(9 + 8, 9 \times 9 + 16 \times 25)$ ***(A1)(M1)(A1)***

Note: Award ***(A1)*** for correct expectation, ***(M1)(A1)*** for correct variance.

$\sim \mathrm{N}(17,481)$

$\Rightarrow \mathrm{P}(3X_1 + 4Y_1 > 15) = 0.536$ ***A1***

[4 marks]

b.iii. $X_1 + X_2 + Y_1 + Y_2 + Y_3 + Y_4 \sim \text{N}(6 + 8, 2 \times 9 + 4 \times 25)$ ***(A1)(A1)***

$\sim \text{N}(14, 118)$

$\Rightarrow \text{P}(X_1 + X_2 + Y_1 + Y_2 + Y_3 + Y_4 < 30) = 0.930$ ***A1***

[3 marks]

c. consider $\bar{X} - \bar{Y}$ ***(M1)***

$\text{E}\left(\bar{X} - \bar{Y}\right) = 3 - 2 = 1$ ***A1***

$\text{Var}\left(\bar{X} - \bar{Y}\right) = \frac{9}{2} + \frac{25}{4}(= 10.75)$ ***(M1)A1***

$\Rightarrow \text{P}\left(\bar{X} - \bar{Y} > 0\right) = 0.620$ ***A1***

[5 marks]

Examiners report

a. [N/A]
b.i. [N/A]
b.ii. [N/A]
b.iii. [N/A]
c. [N/A]

An automatic machine is used to fill bottles of water. The amount delivered, \(Y\) ml , may be assumed to be normally distributed with mean μ ml and standard deviation 8 ml . Initially, the machine is adjusted so that the value of μ is 500. In order to check that the value of μ remains equal to 500, a random sample of 10 bottles is selected at regular intervals, and the mean amount of water, $\overline{y}$, in these bottles is calculated. The following hypotheses are set up.

$\text{H}_0 : \mu = 500$; $\text{H}_1 : \mu \ne 500$

The critical region is defined to be $\left(\overline{y} < 495\right) \cup \left(\overline{y} > 505\right)$.

(i) Find the significance level of this procedure.

(ii) Some time later, the actual value of μ is 503. Find the probability of a Type II error.

Markscheme

(i) Under H_0 , the distribution of $\overline{y}$ is N(500, 6.4) . ***(A1)***

Significance level $= \text{P}\overline{y} < 495 \text{ or } > 505|\text{H}_0$ ***M2***

$= 2 \times 0.02405$ ***(A1)***

$= 0.0481$ ***A1 N5***

Note: Using tables, answer is 0.0478.

(ii) The distribution of $\overline{y}$ is now N(503, 6.4) . ***(A1)***

P(Type II error) $= \text{P}(495 < \overline{y} < 505)$ ***(M1)***

$= 0.785$ ***A1 N3***

Note: Using tables, answer is 0.784.

[8 marks]

Examiners report

[N/A]

The function f is defined by $f(x) = \ln(1 + \sin x)$.

When a scientist measures the concentration μ of a solution, the measurement obtained may be assumed to be a normally distributed random variable with mean μ and standard deviation 1.6.

A.a Show that $f''(x) = \frac{-1}{1+\sin x}$. [4]

A.b Determine the Maclaurin series for $f(x)$ as far as the term in x^4 . [6]

A.c Deduce the Maclaurin series for $\ln(1 - \sin x)$ as far as the term in x^4 . [2]

A.d By combining your two series, show that $\ln \sec x = \frac{x^2}{2} + \frac{x^4}{12} + \ldots$. [4]

A.e Hence, or otherwise, find $\lim\limits_{x \to 0} \frac{\ln \sec x}{x\sqrt{x}}$. [2]

B.a He makes 5 independent measurements of the concentration of a particular solution and correctly calculates the following confidence interval for μ . [5]

$$[22.7\ ,\ 26.1]$$

Determine the confidence level of this interval.

B.b He is now given a different solution and is asked to determine a 95% confidence interval for its concentration. The confidence interval is required to have a width less than 2. Find the minimum number of independent measurements required. [5]

Markscheme

A.a $f'(x) = \frac{\cos x}{1+\sin x}$ ***M1A1***

$f''(x) = \frac{-\sin x(1+\sin x) - \cos^2 x}{(1+\sin x)^2}$ ***M1***

$= \frac{-\sin x - 1}{(1+\sin x)^2}$ ***A1***

$= \frac{-1}{1+\sin x}$ ***AG***

[4 marks]

A.b $f'''(x) = \frac{\cos x}{(1+\sin x)^2}$ ***A1***

$f^{iv}(x) = \frac{-\sin x(1+\sin x)^2 - 2(1+\sin x)\cos^2 x}{(1+\sin x)^4}$ ***A1***

$f(0) = 0$, $f'(0) = 1$, $f''(0) = -1$, $f'''(0) = 1$, $f^{iv}(0) = -2$ ***(A2)***

Note: Award ***A1*** for 2 errors and ***A0*** for more than 2 errors.

$\ln(1 + \sin x) = x - \frac{x^2}{2} + \frac{x^3}{6} - \frac{x^4}{12} + \ldots$ ***M1A1***

[6 marks]

A.c. $\ln(1 - \sin x) = \ln(1 + \sin(-x)) = -x - \frac{x^2}{2} - \frac{x^3}{6} - \frac{x^4}{12} + \ldots$ ***M1A1***

[2 marks]

A.d. Adding, ***M1***

$\ln(1 - \sin^2 x) = \ln \cos^2 x$ ***A1***

$= -x^2 - \frac{x^4}{6} + \ldots$ ***A1***

$\ln \cos x = -\frac{x^2}{2} - \frac{x^4}{12} + \ldots$ ***A1***

$\ln \sec x = \frac{x^2}{2} + \frac{x^4}{12} + \ldots$ ***AG***

[4 marks]

A.e. $\frac{\ln \sec x}{x\sqrt{x}} = \frac{\sqrt{x}}{2} + \frac{x^2\sqrt{x}}{12} + \ldots$ ***M1***

Limit $= 0$ ***A1***

[2 marks]

B.a. Interval width $= 26.1 - 22.7 = 3.4$

So $3.4 = 2z \times \frac{1.6}{\sqrt{5}}$ ***M1A1***

$z = 2.375\ldots$ ***A1***

Probability $= 0.9912$ ***A1***

Confidence level $= 2 \times 0.4912 = 98.2\%$ ***A1***

[5 marks]

B.b. z-value $= 1.96$ ***A1***

We require

$2 \times \frac{1.96 \times 1.6}{\sqrt{n}} < 2$ ***M1A1***

Whence $n > 9.83$ ***A1***

So we need $n = 10$ ***A1***

Note: Accept $=$ signs throughout.

[5 marks]

Examiners report

A.a. [N/A]
A.b. [N/A]
A.c. [N/A]
A.d. [N/A]
A.e. [N/A]
B.a. [N/A]
[N/A]

B.b.

The random variable X has probability density function given by

$$f(x) = \begin{cases} x\mathrm{e}^{-x}, & \text{for } x \geqslant 0, \\ 0, & \text{otherwise} \end{cases}.$$

A sample of size 50 is taken from the distribution of X.

a. Use l'Hôpital's rule to show that $\lim\limits_{x \to \infty} \frac{x^3}{\mathrm{e}^x} = 0$. [3]

b. (i) Find $\mathrm{E}(X^2)$. [10]

(ii) Show that $\mathrm{Var}(X) = 2$.

c. State the central limit theorem. [2]

d. Find the probability that the sample mean is less than 2.3. [2]

Markscheme

a. attempt to apply l'Hôpital's rule ***M1***

$\lim\limits_{x \to \infty} \frac{3x^2}{\mathrm{e}^x}$ ***A1***

then $\lim\limits_{x \to \infty} \frac{6x}{\mathrm{e}^x}$

then $\lim\limits_{x \to \infty} \frac{6}{\mathrm{e}^x}$ ***A1***

$= 0$ ***AG***

[3 marks]

b. (i) $\mathrm{E}(X^2) = \lim\limits_{R \to \infty} \int\limits_0^R x^3\mathrm{e}^{-x}\mathrm{d}x$ ***M1***

attempt at integration by parts ***M1***

the integral $= [-x^3\mathrm{e}^{-x}]_0^R + \int\limits_0^R 3x^2\mathrm{e}^{-x}\mathrm{d}x$ ***A1A1***

$= [-x^3\mathrm{e}^{-x}]_0^R + [-3x^2\mathrm{e}^{-x}]_0^R + \int\limits_0^R 6x\mathrm{e}^{-x}\mathrm{d}x$ ***M1***

$= [-x^3\mathrm{e}^{-x}]_0^R + [-3x^2\mathrm{e}^{-x}]_0^R + [-6x\mathrm{e}^{-x}]_0^R + \int\limits_0^R 6\mathrm{e}^{-x}\mathrm{d}x$ ***A1***

$= [-x^3\mathrm{e}^{-x}]_0^R + [-3x^2\mathrm{e}^{-x}]_0^R + [-6x\mathrm{e}^{-x}]_0^R + [-6\mathrm{e}^{-x}]_0^R$ ***A1***

$= 6$ when $R \to \infty$ ***R1***

(ii) $\mathrm{E}(X) = 2$ ***A1***

$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 6 - 2^2$ ***M1***

$= 2$ ***AG***

[10 marks]

c. if a random sample of size n is taken from ***any*** distribution X, with $\mathrm{E}(X)=\mu$ and $\mathrm{Var}(X)=\sigma^2$, then, for ***large n***, ***A1***

the sample mean $\bar{X}$ has approximate distribution $\mathrm{N}\left(\mu,\ \frac{\sigma^2}{n}\right)$ ***A1***

[2 marks]

d. $\bar{X} \sim \mathrm{N}\left(2,\ \frac{2}{50}=(0.2)^2\right)$ ***(A1)***

$\mathrm{P}(\bar{X}<2.3)=(\mathrm{P}(Z<1.5))=0.933$ ***A1***

[2 marks]

Examiners report

a. [N/A]

b. In part (b) the infinite upper limit was rarely treated rigorously.

c. In answering part (c) many failed to say that the Central Limit Theorem is valid for large samples and for any initial distribution. The parameters of the distribution were often not stated.

d. [N/A]

In a large population of sheep, their weights are normally distributed with mean μ kg and standard deviation σ kg. A random sample of 100 sheep is taken from the population.

The mean weight of the sample is $\bar{X}$ kg.

a. State the distribution of $\bar{X}$, giving its mean and standard deviation. [2]

b. The sample values are summarized as $\sum x=3782$ and $\sum x^2=155341$ where x kg is the weight of a sheep. [6]

(i) Find unbiased estimates for μ and σ^2.

(ii) Find a 95% confidence interval for μ.

c. Test, at the 1% level of significance, the null hypothesis $\mu=35$ against the alternative hypothesis that $\mu>35$. [5]

Markscheme

a. $\bar{X} \sim N\left(\mu,\ \frac{\sigma^2}{100}\right)$ ***A1A1***

Note: Award ***A1*** for N, ***A1*** for the parameters.

b. (i) $\bar{x}=\frac{\sum x}{n}=\frac{3782}{100}=37.8$ ***A1***

$s_{n-1}^2=\frac{155341}{99}-\frac{3782^2}{9900}=124$ ***M1A1***

(ii) $95\%CI=37.82 \pm 1.98\sqrt{\frac{124.3006}{100}}$ ***(M1)(A1)***

$=(35.6,\ 40.0)$ ***A1***

c. **METHOD 1**

one tailed t-test ***A1***

testing 37.82 ***A1***

99 degrees of freedom

reject if $t > 2.36$ ***A1***

t-value being tested is 2.5294 ***A1***

since $2.5294 > 2.36$ we reject the null hypothesis and accept the alternative hypothesis ***R1***

METHOD 2

one tailed t-test ***(A1)***

$p = 0.00650$ ***A3***

since p - $\text{value} < 0.01$ we reject the null hypothesis and accept the alternative hypothesis ***R1***

Examiners report

a. Almost all candidates recognised the sample distribution as normal but were not always successful in stating the mean and the standard deviation. Similarly almost all candidates knew how to find an unbiased estimator for μ, but a number failed to find the correct answer for the unbiased estimator for σ^2. Most candidates were successful in finding the 95% confidence interval for μ. In part c) many fully correct answers were seen but a significant number of candidates did not recognise they were working with a t-distribution.

b. Almost all candidates recognised the sample distribution as normal but were not always successful in stating the mean and the standard deviation. Similarly almost all candidates knew how to find an unbiased estimator for μ, but a number failed to find the correct answer for the unbiased estimator for σ^2. Most candidates were successful in finding the 95% confidence interval for μ. In part c) many fully correct answers were seen but a significant number of candidates did not recognise they were working with a t-distribution.

c. Almost all candidates recognised the sample distribution as normal but were not always successful in stating the mean and the standard deviation. Similarly almost all candidates knew how to find an unbiased estimator for μ, but a number failed to find the correct answer for the unbiased estimator for σ^2. Most candidates were successful in finding the 95% confidence interval for μ. In part c) many fully correct answers were seen but a significant number of candidates did not recognise they were working with a t-distribution.

The random variable X has the binomial distribution $\mathrm{B}(n,\ p)$, where $n > 1$.

Show that

(a) $\frac{X}{n}$ is an unbiased estimator for p;

(b) $\left(\frac{X}{n}\right)^2$ is **not** an unbiased estimator for p^2;

(c) $\frac{X(X-1)}{n(n-1)}$ is an unbiased estimator for p^2.

Markscheme

(a) $\mathrm{E}\left(\frac{X}{n}\right)=\frac{1}{n}\mathrm{E}(X)$ ***M1***

$=\frac{1}{n}\times np=p$ ***AG***

therefore unbiased ***AG***

[2 marks]

(b) $\mathrm{E}\left[\left(\frac{X}{n}\right)^2\right]=\frac{1}{n^2}\left(\mathrm{Var}(X)+[\mathrm{E}(X)]^2\right)$ ***M1A1***

$=\frac{1}{n^2}\left(np(1-p)+n^2p^2\right)$ ***A1***

$\neq p^2$ ***A1***

therefore not unbiased ***AG***

[4 marks]

(c) $\mathrm{E}\left[\left(\frac{X(X-1)}{n(n-1)}\right)\right]=\frac{\mathrm{E}(X^2)-\mathrm{E}(X)}{n(n-1)}$ ***M1***

$=\frac{np(1-p)+n^2p^2-np}{n(n-1)}$ ***A1***

$=\frac{np^2(n-1)}{n(n-1)}$ ***A1***

$=p^2$

therefore unbiased ***AG***

[3 marks]

Examiners report

[N/A]

---

a. The continuous random variable X takes values only in the interval $[a, b]$ and F denotes its cumulative distribution function. Using integration by parts, show that: [4]

$$E(X)=b-\int_a^b F(x)\mathrm{d}x.$$

b. The continuous random variable Y has probability density function f given by: [14]

$$\begin{aligned} f(y)&=\cos y, \quad 0\leqslant y\leqslant \tfrac{\pi}{2}\\ f(y)&=0, \qquad \text{elsewhere.}\end{aligned}$$

(i) Obtain an expression for the cumulative distribution function of Y, valid for $0\le y\le \frac{\pi}{2}$. Use the result in (a) to determine $E(Y)$.

(ii) The random variable U is defined by $U=Y^n$, where $n\in\mathbb{Z}^+$. Obtain an expression for the cumulative distribution function of U valid for $0\le u\le\left(\frac{\pi}{2}\right)^n$.

(iii) The medians of U and Y are denoted respectively by m_u and m_y. Show that $m_u=m_y^n$.

Markscheme

a. $E(X)=\int_a^b xf(x)\mathrm{d}x$ ***M1***

$= [xF(x)]_a^b - \int_a^b F(x)\mathrm{d}x$ ***A1***

$= bF(b) - aF(a) - \int_a^b F(x)\mathrm{d}x$ ***A1***

$= b - \int_a^b F(x)\mathrm{d}x$ because $F(a) = 0$ and $F(b) = 1$ ***A1***

[4 marks]

b. (i) let G denote the cumulative distribution function of Y

$G(y) = \int_0^y \cos t\mathrm{d}t$ ***M1***

$= [\sin t]_0^y$ ***(A1)***

$= \sin y$ ***A1***

$E(Y) = \frac{\pi}{2} - \int_0^{\frac{\pi}{2}} \sin y\mathrm{d}y$ ***M1***

$= \frac{\pi}{2} + [\cos y]_0^{\frac{\pi}{2}}$ ***A1***

$= \frac{\pi}{2} - 1$ ***A1***

(ii) CDF of $U = P(U \le u)$ ***M1***

$= P(Y^n \le u)$ ***A1***

$= P(Y \le u^{\frac{1}{n}})$ ***A1***

$= G(u^{\frac{1}{n}})$ ***(A1)***

$= \sin\left(u^{\frac{1}{n}}\right)$ ***A1***

(iii) m_y satisfies the equation $\sin m_y = \frac{1}{2}$ ***A1***

m_u satisfies the equation $\sin\left(m_u^{\frac{1}{n}}\right) = \frac{1}{2}$ ***A1***

therefore $m_y = m_u^{\frac{1}{n}}$ ***A1***

$m_u = m_y^n$ ***AG***

[14 marks]

Examiners report

a. Solutions to (a) were often unconvincing. Candidates were expected to include in their solution the fact that $F(a) = 0$ and $F(b) = 1$.

b. In (b) (i), it was not enough to state that $G(y) = \int \cos y\mathrm{d}y = \sin y$ although that, fortuitously, gave the correct answer on this occasion. The correct approach was either to state that $G(y) = \int_0^y \cos t\mathrm{d}t = \sin y$ or that $G(y) = \int \cos y\mathrm{d}y = \sin x + C$ and then show that $C = 0$ because $F(0) = 0$ or $F\left(\frac{\pi}{2}\right) = 1$. Solutions to (b) (ii) and (iii) were often disappointing, giving the impression that many of the candidates were not familiar with dealing with cumulative distribution functions.

A random variable X has probability density function f given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{for } x \geqslant 0 \text{ where } \lambda > 0 \\ 0, & \text{for } x < 0. \end{cases}$$

a. (i) Find an expression for $\mathrm{P}(X > a)$, where $a > 0$. [6]

A chicken crosses a road. It is known that cars pass the chicken's crossing route, with intervals between cars measured in seconds, according to the random variable X , with $\lambda = 0.03$. The chicken, which takes 10 seconds to cross the road, starts to cross just as one car passes.

(ii) Find the probability that the chicken will reach the other side of the road before the next car arrives.

Later, the chicken crosses the road again just after a car has passed.

(iii) Show that the probability that the chicken completes both crossings is greater than 0.5.

b. A rifleman shoots at a circular target. The distance in centimetres from the centre of the target at which the bullet hits, can be modelled by X with $\lambda = 0.4$. The rifleman scores 10 points if $X \le 1$, 5 points if $1 < X \le 5$, 1 point if $5 < X \le 10$ and no points if $X > 10$. [10]

(i) Find the expected score when one bullet is fired at the target.

A second rifleman, whose shooting can also be modelled by X , wishes to find his value of λ .

(ii) Given that his expected score is 6.5, find his value of λ .

Markscheme

a. (i) $\mathrm{P}(X > a) = \int_a^\infty \lambda e^{-\lambda x}\mathrm{d}x$ ***M1***

$\left[-e^{-\lambda x}\right]_a^\infty$ ***A1***

$= e^{-\lambda a}$ ***A1***

(ii) $\mathrm{P}(X > 10) = e^{-0.3} (= 0.74\ldots)$ ***(M1)A1***

(iii) probability of a safe double crossing $= e^{-0.6}\ (= 0.74^2) = 0.55$ ***A1***

which is greater than 0.5 ***AG***

[6 marks]

b. (i) $\mathrm{P}(X \le 1) = 0.3296\ldots$ ***(A1)***

$\mathrm{P}(1 \le X \le 5) = 0.5349\ldots$ ***(A1)***

$\mathrm{P}(5 \le X \le 10) = 0.1170\ldots$ ***(A1)***

$\mathrm{E}(\text{score}) = 10 \times 0.3296\ldots + 5 \times 0.5349\ldots + 1 \times 0.1170\ldots$ ***M1A1***

$= 6.09$ ***A1***

Note: Accept probabilities in exponential form until the final decimal answer.

(ii) $\mathrm{E}(\text{score})$ for ***X*** with unknown parameter can be expressed as $10 \times (1 - e^{-\lambda}) + 5 \times (e^{-\lambda} - e^{-5\lambda}) + (e^{-5\lambda} - e^{-10\lambda})$ ***(M1)(A1)***

attempt to solve $\mathrm{E}(\text{score}) = 6.5$ ***(M1)***

obtain $\lambda = 0.473$ ***A1***

[10 marks]

Examiners report

a. This question was generally well done.

b. This question was generally well done.

The discrete random variable X follows the distribution Geo(p).

Arthur tosses a biased coin each morning to decide whether to walk or cycle to school; he walks if the coin shows a head.

The probability of obtaining a head is 0.55.

a. (i) Write down the mode of X . [3]

(ii) Find the exact value of p if $\mathrm{Var}(X) = \frac{28}{9}$.

b. (i) Find the smallest value of n for which the probability of Arthur walking to school on the next n days is less than 0.01. [6]

(ii) Find the probability that Arthur cycles to school for the third time on the last of eight successive days.

Markscheme

a. (i) the mode is 1 ***A1***

(ii) attempt to solve $\frac{1-p}{p^2} = \frac{28}{9}$ ***M1***

obtain $p = \frac{3}{7}$ ***A1***

Note: $p = 0.429$ is awarded ***M1A0***.

[3 marks]

b. (i) require least n such that

$0.55^n < 0.01$ ***(M1)***

EITHER

listing values: 0.55, 0.3025, 0.166, 0.091, 0.050, 0.028, 0.015, 0.0084 ***(M1)***

obtain $n = 8$ ***A1***

OR

$n > \frac{\ln 0.01}{\ln 0.55} = 7.70\ldots$ ***(M1)***

obtain $n = 8$ ***A1***

(ii) recognition of negative binomial ***(M1)***

$X \sim NB(3, 0.45)$

$\mathrm{P}(X=8)=\binom{7}{2}\times 0.45^{3}\times 0.55^{5}$ ***(A1)***

$=0.0963$ ***A1***

Note: If 0.45 and 0.55 are mixed up, count it as a misread – probability in that case is 0.0645.

[6 marks]

Examiners report

a. (a)(i) A surprising number of candidates were unaware of the definition of the mode of a distribution.

(a)(ii) Generally well done, although a few candidates gave a decimal answer.

b. (b) Generally well done, and it was pleasing that most were familiar with the direct use of the negative binomial distribution in (ii).