HL Paper 3

The weight of tea in *Supermug* tea bags has a normal distribution with mean 4.2 g and standard deviation 0.15 g. The weight of tea in *Megamug* tea bags has a normal distribution with mean 5.6 g and standard deviation 0.17 g.

a. Find the probability that a randomly chosen *Supermug* tea bag contains more than 3.9 g of tea. [2]

b. Find the probability that, of two randomly chosen *Megamug* tea bags, one contains more than 5.4 g of tea and one contains less than 5.4 g of tea. [4]

c. Find the probability that five randomly chosen *Supermug* tea bags contain a total of less than 20.5 g of tea. [4]

d. Find the probability that the total weight of tea in seven randomly chosen *Supermug* tea bags is more than the total weight in five randomly chosen *Megamug* tea bags. [5]

Markscheme

a. let *S* be the weight of tea in a random *Supermug* tea bag

$S \sim \mathrm{N}(4.2, 0.15^2)$

$\mathrm{P}(S > 3.9) = 0.977$ ***(M1)A1***

[2 marks]

b. let *M* be the weight of tea in a random *Megamug* tea bag

$M \sim \mathrm{N}(5.6, 0.17^2)$

$\mathrm{P}(M > 5.4) = 0.880\ldots$ ***(A1)***

$\mathrm{P}(M < 5.4) = 1 - 0.880\ldots = 0.119\ldots$ ***(A1)***

required probability $= 2 \times 0.880\ldots \times 0.119\ldots = 0.211$ ***M1A1***

[4 marks]

c. $\mathrm{P}(S_1 + S_2 + S_3 + S_4 + S_5 < 20.5)$

let $S_1 + S_2 + S_3 + S_4 + S_5 = A$ ***(M1)***

$\mathrm{E}(A) = 5\mathrm{E}(S)$

= 21 ***A1***

$\mathrm{Var}(A) = 5\mathrm{Var}(S)$

= 0.1125 ***A1***

$A \sim \mathrm{N}(21, 0.1125)$

$\mathrm{P}(A < 20.5) = 0.0680$ ***A1***

[4 marks]

d. $\mathrm{P}(S_1+S_2+S_3+S_4+S_5+S_6+S_7-(M_1+M_2+M_3+M_4+M_5)>0)$

let $S_1+S_2+S_3+S_4+S_5+S_6+S_7-(M_1+M_2+M_3+M_4+M_5)=B$ ***(M1)***

$\mathrm{E}(B)=7\mathrm{E}(S)-5\mathrm{E}(M)$

= 1.4 ***A1***

Note: Above ***A1*** is independent of first ***M1***.

$\mathrm{Var}(B)=7\mathrm{Var}(S)+5\mathrm{Var}(M)$ ***(M1)***

= 0.302 ***A1***

$\mathrm{P}(B>0)=0.995$ ***A1***

[5 marks]

Examiners report

a. For most candidates this was a reasonable start to the paper with many candidates gaining close to full marks. The most common error was in (b) where, surprisingly, many candidates did not realise the need to multiply the product of the two probabilities by 2 to gain the final answer. Weaker candidates often found problems in understanding how to correctly find the variance in both (c) and (d).

b. For most candidates this was a reasonable start to the paper with many candidates gaining close to full marks. The most common error was in (b) where, surprisingly, many candidates did not realise the need to multiply the product of the two probabilities by 2 to gain the final answer. Weaker candidates often found problems in understanding how to correctly find the variance in both (c) and (d).

c. For most candidates this was a reasonable start to the paper with many candidates gaining close to full marks. The most common error was in (b) where, surprisingly, many candidates did not realise the need to multiply the product of the two probabilities by 2 to gain the final answer. Weaker candidates often found problems in understanding how to correctly find the variance in both (c) and (d).

d. For most candidates this was a reasonable start to the paper with many candidates gaining close to full marks. The most common error was in (b) where, surprisingly, many candidates did not realise the need to multiply the product of the two probabilities by 2 to gain the final answer. Weaker candidates often found problems in understanding how to correctly find the variance in both (c) and (d).

Two species of plant, A and B, are identical in appearance though it is known that the mean length of leaves from a plant of species A is 5.2 cm, whereas the mean length of leaves from a plant of species B is 4.6 cm. Both lengths can be modelled by normal distributions with standard deviation 1.2 cm.

In order to test whether a particular plant is from species A or species B, 16 leaves are collected at random from the plant. The length, x, of each leaf is measured and the mean length evaluated. A one-tailed test of the sample mean, $\bar{X}$, is then performed at the 5% level, with the hypotheses: $H_0:\mu=5.2$ and $H_1:\mu<5.2$.

a. Find the critical region for this test. [3]

c. It is now known that in the area in which the plant was found 90% of all the plants are of species A and 10% are of species B. [2]

Find the probability that $\bar{X}$ will fall within the critical region of the test.

d. If, having done the test, the sample mean is found to lie within the critical region, find the probability that the leaves came from a plant of species A. [3]

Markscheme

a. $\bar{X} \sim N\left(5.2,\ \frac{1.2^2}{16}\right)$ ***(M1)***

critical value is $5.2 - 1.64485\ldots \times \frac{1.2}{4} = 4.70654\ldots$ ***(A1)***

critical region is $]-\infty,\ 4.71]$ ***A1***

Note: Allow follow through for the final ***A1*** from their critical value.

Note: Follow through previous values in (b), (c) and (d).

[3 marks]

c. $0.9 \times 0.05 + 0.1 \times (1 - 0.361\ldots) = 0.108875997\ldots = 0.109$ ***M1A1***

Note: Award ***M1*** for a weighted average of probabilities with weights $0.1, 0.9$.

[2 marks]

d. attempt to use conditional probability formula ***M1***

$\frac{0.9\times0.05}{0.108875997\ldots}$ ***(A1)***

$= 0.41334\ldots = 0.413$ ***A1***

[3 marks]

Total [10 marks]

Examiners report

a. Solutions to this question were generally disappointing.

In (a), the standard error of the mean was often taken to be $\sigma(1.2)$ instead of $\frac{\sigma}{\sqrt{n}}(0.3)$ and the solution sometimes ended with the critical value without the critical region being given.

c. In (c), the question was often misunderstood with candidates finding the weighted mean of the two means, ie $0.9 \times 5.2 + 0.1 \times 4.6 = 5.14$ instead of the weighted mean of two probabilities.

d. Without having the solution to (c), part (d) was inaccessible to most of the candidates so that very few correct solutions were seen.

The random variable $X \sim \mathrm{Po}(m)$. Given that $\mathrm{P}(X = k-1) = \mathrm{P}(X = k+1)$, where k is a positive integer,

a. show that $m^2 = k(k+1)$; [2]

b. hence show that the mode of X is k . [6]

Markscheme

a. $\frac{m^{k-1}e^{-m}}{(k-1)!} = \frac{m^{k+1}e^{-m}}{(k+1)!}$ ***M1***

$\Rightarrow 1 = \frac{m^2}{(k+1)k}$ ***A1***

Note: Award ***A1*** for any correct intermediate step.

$\Rightarrow m^2 = (k+1)k$ ***AG***

[2 marks]

b. $\frac{\mathrm{P}(X=k)}{\mathrm{P}(X=k-1)} = \frac{e^{-m}\times\frac{m^k}{k!}}{e^{-m}\times\frac{m^{k-1}}{(k-1)!}}$ ***M1***

$= \frac{m}{k}$ ***A1***

$= \frac{\sqrt{k(k+1)}}{k}$ ***M1***

$= \sqrt{\frac{k+1}{k}} > 1$ ***R1***

so $\mathrm{P}(X=k) > \mathrm{P}(X=k-1)$ ***R1***

similarly $\mathrm{P}(X=k) > \mathrm{P}(X=k+1)$ ***R1***

hence k is the mode ***AG***

[6 marks]

Examiners report

a. Most candidates were able to complete part (a). The remainder of the question involved some understanding of the shape of the distribution and some facility with algebraic manipulation.

b. Most candidates were able to complete part (a). The remainder of the question involved some understanding of the shape of the distribution and some facility with algebraic manipulation.

A traffic radar records the speed, v kilometres per hour ($\mathrm{km\,h^{-1}}$), of cars on a section of a road.

The following table shows a summary of the results for a random sample of 1000 cars whose speeds were recorded on a given day.

Speed	Number of cars
$50 \le v < 60$	5
$60 \le v < 70$	13
$70 \le v < 80$	52
$80 \le v < 90$	68
$90 \le v < 100$	98
$100 \le v < 110$	105
$110 \le v < 120$	289
$120 \le v < 130$	142
$130 \le v < 140$	197
$140 \le v < 150$	31

a. Using the data in the table, [4]

(i) show that an estimate of the mean speed of the sample is 113.21 $\mathrm{km\,h^{-1}}$;

(ii) find an estimate of the variance of the speed of the cars on this section of the road.

b. Find the 95% confidence interval, I, for the mean speed. [2]

c. Let J be the 90% confidence interval for the mean speed. [2]

Without calculating J, explain why $J \subset I$.

Markscheme

a. (i) $\bar{v} = \frac{1}{1000}(55 \times 5 + 65 \times 13 + \ldots + 145 \times 31)$ ***A1M1***

Note: ***A1*** for mid-points, ***M1*** for use of the formula.

$= \frac{113\ 210}{1000} = 113.21$ ***AG***

(ii) $s^2 = \frac{(55-113.21)^2 \times 5 + (65-113.21)^2 \times 13 + \ldots + (145-113.21)^2 \times 31}{999}$ ***(M1)***

$= \frac{362\ 295.9}{999} = 362.6585\ldots = 363$ ***A1***

Note: Award ***A1*** if answer rounds to 362 or 363.

Note: Condone division by 1000.

[4 marks]

b. $\bar{v} \pm \frac{t_{0.025} \times s}{\sqrt{n}}$ ***(M1)***

hence the confidence interval $I = [112.028,\ 114.392]$ ***A1***

Note: Accept answers which round to 112 and 114.

Note: Condone the use of $z_{0.025}$ for $t_{0.025}$ and σ for s.

[2 marks]

c. less confidence implies narrower interval ***R2***

Note: Accept equivalent statements or arguments having a meaningful diagram and/or relevant percentiles.

hence the confidence interval I at the 95% level contains the confidence interval J at the 90% level ***AG***

[2 marks]

Examiners report

a. In (a)(i), the candidates were required to show that the estimate of the mean is 113.21 so that those who stated simply 'Using my GDC, mean = 113.21' were given no credit. Candidates were expected to indicate that the interval midpoints were used and to show the appropriate formula. In (a)(ii), division by either 999 or 1000 was accepted, partly because of the large sample size and partly because the question did not ask for an unbiased estimate of variance.

b.

c. Solutions to (c) were often badly written, often quite difficult to understand exactly what was being stated.

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

(a) (i) Show that $\frac{\mathrm{P}(X=x)}{\mathrm{P}(X=x-1)} = \frac{(n-x+1)p}{x(1-p)}$.

(ii) Deduce that if $\mathrm{P}(X = x) > \mathrm{P}(X = x - 1)$ then $x < (n+1)p$.

(iii) Hence, determine the value of x which maximizes $\mathrm{P}(X = x)$ when $(n+1)p$ is not an integer.

(b) Given that $n = 19$, find the set of values of p for which X has a unique mode of 13.

Markscheme

(a) (i) $\frac{\mathrm{P}(X=x)}{\mathrm{P}(X=x-1)} = \frac{\left(\frac{n!}{(n-x)!x!} \times p^x \times (1-p)^{n-x}\right)}{\left(\frac{n!}{(n-x+1)!(x-1)!} \times p^{x-1} \times (1-p)^{n-x+1}\right)}$ ***M1A1***

$= \frac{(n-x+1)p}{x(1-p)}$ ***AG***

(ii) if $\mathrm{P}(X = x) > \mathrm{P}(X = x - 1)$ then

$(n - x + 1)p > x(1 - p)$ ***(M1)A1***

$np - xp + p > x - px$ ***A1***

$x < (n+1)p$ ***AG***

(iii) to maximise the probability we also need

$\mathrm{P}(X = x) > \mathrm{P}(X = x + 1)$ ***(M1)***

$\frac{(n-(x+1)+1)p}{(x+1)(1-p)} < 1$

$np - xp < x - xp + 1 - p$

$p(n+1) < x + 1$ ***A1***

hence $p(n+1) > x > p(n+1) - 1$ ***(A1)***

so x is the integer part of $(n+1)p$ *i.e.* the largest integer less than $(n+1)p$ ***A1***

[9 marks]

(b) the mode is the value which maximises the probability ***(R1)***

$20p > 13 > 20p - 1$ ***M1***

$\Rightarrow p > \frac{13}{20} = 0.65$, and $p < \frac{7}{10} = 0.70$ ***A1A1***

(it follows that $0.65 < p < 0.7$)

[4 marks]

Total [13 marks]

Examiners report

Many candidates made a reasonable attempt at (a)(i) and (ii) but few were able to show that the mode is the integer part of $(n+1)p$. Part (b) also proved difficult for most candidates with few correct solutions seen.

Each week the management of a football club recorded the number of injuries suffered by their playing staff in that week. The results for a 52-week period were as follows:

Number of injuries per week	0	1	2	3	4	5	6
Number of weeks	6	14	15	9	5	2	1

a. Calculate the mean and variance of the number of injuries per week. [2]

b. Explain why these values provide supporting evidence for using a Poisson distribution model. [1]

Markscheme

a. mean = 2.06 ***A1***

variance = 1.94 ***A1***

[2 marks]

b. a Poisson distribution has the property that its mean and variance are the same ***R1***

[1 mark]

Examiners report

a. Many candidates picked up good marks for this question, but lost marks because of inattention to detail. The mean of the data was usually given correctly, but sometimes the variance was wrong. It may seem a small point, but the correct hypotheses should not mention the value of the estimated mean. Some candidates did not notice that some columns needed to be combined.

b. Many candidates picked up good marks for this question, but lost marks because of inattention to detail. The mean of the data was usually given correctly, but sometimes the variance was wrong. It may seem a small point, but the correct hypotheses should not mention the value of the estimated mean. Some candidates did not notice that some columns needed to be combined.

If X is a random variable that follows a Poisson distribution with mean $\lambda > 0$ then the probability generating function of X is $G(t) = e^{\lambda(t-1)}$.

a. (i) Prove that $\mathrm{E}(X) = \lambda$. [6]

(ii) Prove that $\mathrm{Var}(X) = \lambda$.

b. Y is a random variable, independent of X, that also follows a Poisson distribution with mean λ. [3]

If $S = 2X - Y$ find

(i) $\mathrm{E}(S)$;

(ii) $\mathrm{Var}(S)$.

c. Let $T = \frac{Y}{2} + \frac{Y}{2}$. [3]

(i) Show that T is an unbiased estimator for λ.

(ii) Show that T is a more efficient unbiased estimator of λ than S.

d. Could either S or T model a Poisson distribution? Justify your answer. [1]

e. By consideration of the probability generating function, $G_{X+Y}(t)$, of $X+Y$, prove that $X+Y$ follows a Poisson distribution with mean 2λ. [3]

f. Find [2]

(i) $G_{X+Y}(1)$;

(ii) $G_{X+Y}(-1)$.

g. Hence find the probability that $X+Y$ is an even number. [3]

Markscheme

a. (i) $G'(t) = \lambda e^{\lambda(t-1)}$ ***A1***

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

$= \lambda$ ***AG***

(ii) $G''(t) = \lambda^2 e^{\lambda(t-1)}$ ***M1***

$\Rightarrow G''(1) = \lambda^2$ ***(A1)***

$\mathrm{Var}(X) = G''(1) + G'(1) - \left(G'(1)\right)^2$ ***(M1)***

$= \lambda^2 + \lambda - \lambda^2$ ***A1***

$= \lambda$ ***AG***

[6 marks]

b. (i) $\mathrm{E}(S) = 2\lambda - \lambda = \lambda$ ***A1***

(ii) $\mathrm{Var}(S) = 4\lambda + \lambda = 5\lambda$ ***(A1)A1***

Note: First ***A1*** can be awarded for either 4λ or λ.

[3 marks]

c. (i) $\mathrm{E}(T) = \frac{\lambda}{2} + \frac{\lambda}{2} = \lambda$ (so T is an unbiased estimator) ***A1***

(ii) $\mathrm{Var}(T) = \frac{1}{4}\lambda + \frac{1}{4}\lambda = \frac{1}{2}\lambda$ ***A1***

this is less than $\mathrm{Var}(S)$, therefore T is the more efficient estimator ***R1AG***

Note: Follow through their variances from (b)(ii) and (c)(ii).

[3 marks]

d. no, mean does not equal the variance ***R1***

[1 mark]

e. $G_{X+Y}(t) = e^{\lambda(t-1)} \times e^{\lambda(t-1)} = e^{2\lambda(t-1)}$ ***M1A1***

which is the probability generating function for a Poisson with a mean of 2λ ***R1AG***

[3 marks]

f. (i) $G_{X+Y}(1) = 1$ ***A1***

(ii) $G_{X+Y}(-1) = e^{-4\lambda}$ ***A1***

[2 marks]

g. $G_{X+Y}(1) = p(0) + p(1) + p(2) + p(3) \ldots$

$G_{X+Y}(-1) = p(0) - p(1) + p(2) - p(3) \ldots$

so $2\mathrm{P}(\text{even}) = G_{X+Y}(1) + G_{X+Y}(-1)$ ***(M1)(A1)***

$\mathrm{P}(\text{even}) = \frac{1}{2}(1 + e^{-4\lambda})$ ***A1***

[3 marks]

Total [21 marks]

Examiners report

a. Solutions to the different parts of this question proved to be extremely variable in quality with some parts well answered by the majority of the candidates and other parts accessible to only a few candidates. Part (a) was well answered in general although the presentation was sometimes poor with some candidates doing the differentiation of $G(t)$ and the substitution of $t = 1$ simultaneously.

b. Part (b) was well answered in general, the most common error being to state that $\mathrm{Var}(2X - Y) = \mathrm{Var}(2X) - \mathrm{Var}(Y)$.

c. Parts (c) and (d) were well answered by the majority of candidates.

d. Parts (c) and (d) were well answered by the majority of candidates.

e. Solutions to (e), however, were extremely disappointing with few candidates giving correct solutions. A common incorrect solution was the following:

$$G_{X+Y}(t) = G_X(t)G_Y(t)$$

Differentiating,

$$G'_{X+Y}(t) = G'_X(t)G_Y(t) + G_X(t)G'_Y(t)$$

$$\mathrm{E}(X+Y) = G'_{X+Y}(1) = \mathrm{E}(X) \times 1 + \mathrm{E}(Y) \times 1 = 2\lambda$$

This is correct mathematics but it does not show that $X + Y$ is Poisson and it was given no credit. Even the majority of candidates who showed that $G_{X+Y}(t) = \mathrm{e}^{2\lambda(t-1)}$ failed to state that this result proved that $X + Y$ is Poisson and they usually differentiated this function to show that $\mathrm{E}(X+Y) = 2\lambda$.

f. In (f), most candidates stated that $G_{X+Y}(1) = 1$ even if they were unable to determine $G_{X+Y}(t)$ but many candidates were unable to evaluate $G_{X+Y}(-1)$. Very few correct solutions were seen to (g) even if the candidates correctly evaluated $G_{X+Y}(1)$ and $G_{X+Y}(-1)$.

g. [N/A]

Engine oil is sold in cans of two capacities, large and small. The amount, in millilitres, in each can, is normally distributed according to Large $\sim \mathrm{N}(5000,\ 40)$ and Small $\sim \mathrm{N}(1000,\ 25)$.

a. A large can is selected at random. Find the probability that the can contains at least 4995 millilitres of oil. [2]

b. A large can and a small can are selected at random. Find the probability that the large can contains at least 30 milliliters more than five times the amount contained in the small can. [6]

c. A large can and five small cans are selected at random. Find the probability that the large can contains at least 30 milliliters less than the total amount contained in the small cans. [5]

Markscheme

a. $\mathrm{P}(L \geq 4995) = 0.785$ ***(M1)A1***

Note: Accept any answer that rounds correctly to 0.79.

Award ***M1A0*** for 0.78.

Note: Award ***M1A0*** for any answer that rounds to 0.55 obtained by taking $\mathrm{SD} = 40$.

[2 marks]

b. we are given that $L \sim \mathrm{N}(5000,\ 40)$ and $S \sim \mathrm{N}(1000,\ 25)$

consider $X = L - 5S$ (ignore ± 30) ***(M1)***

$\mathrm{E}(X) = 0$ (± 30 consistent with line above) ***A1***

$\mathrm{Var}(X) = \mathrm{Var}(L) + 25\mathrm{Var}(S) = 40 + 625 = 665$ ***(M1)A1***

require $\mathrm{P}(X \geq 30) \quad (\text{or } \mathrm{P}(X \geq 0) \text{ if } -30 \text{ above})$ ***(M1)***

obtain 0.122 ***A1***

Note: Accept any answer that rounds correctly to 2 significant figures.

[6 marks]

c. consider $Y = L - (S_1 + S_2 + S_3 + S_4 + S_5)$ (ignore ± 30) ***(M1)***

$\mathrm{E}(Y) = 0$ (± 30 consistent with line above) ***A1***

$\mathrm{Var}(Y) = 40 + 5 \times 25 = 165$ ***A1***

require $\mathrm{P}(Y \leq -30)$ $(\text{or } \mathrm{P}(Y \leq 0) \text{ if } +30 \text{ above})$ ***(M1)***

obtain 0.00976 ***A1***

Note: Accept any answer that rounds correctly to 2 significant figures.

Note: Condone the notation $Y = L - 5S$ if the variance is correct.

[5 marks]

Total [13 marks]

Examiners report

a. Most candidates solved (a) correctly. In (b) and (c), however, many candidates made the usual error of confusing $\sum_{i=1}^{n} X_i$ and nX. Indeed some candidates even use the second expression to mean the first. This error leads to an incorrect variance and of course an incorrect answer. Some candidates had difficulty in converting the verbal statements into the correct probability statements, particularly in (c).

b. Most candidates solved (a) correctly. In (b) and (c), however, many candidates made the usual error of confusing $\sum_{i=1}^{n} X_i$ and nX. Indeed some candidates even use the second expression to mean the first. This error leads to an incorrect variance and of course an incorrect answer. Some candidates had difficulty in converting the verbal statements into the correct probability statements, particularly in (c).

c. Most candidates solved (a) correctly. In (b) and (c), however, many candidates made the usual error of confusing $\sum_{i=1}^{n} X_i$ and nX. Indeed some candidates even use the second expression to mean the first. This error leads to an incorrect variance and of course an incorrect answer. Some candidates had difficulty in converting the verbal statements into the correct probability statements, particularly in (c).

When Andrew throws a dart at a target, the probability that he hits it is $\frac{1}{3}$; when Bill throws a dart at the target, the probability that he hits the it is $\frac{1}{4}$. Successive throws are independent. One evening, they throw darts at the target alternately, starting with Andrew, and stopping as soon as one of their darts hits the target. Let X denote the total number of darts thrown.

a. Write down the value of $\mathrm{P}(X=1)$ and show that $\mathrm{P}(X=2)=\frac{1}{6}$. [2]

b. Show that the probability generating function for X is given by [6]

$$G(t)=\frac{2t+t^2}{6-3t^2}.$$

c. Hence determine $\mathrm{E}(X)$. [4]

Markscheme

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

$\mathrm{P}(X=2)=\frac{2}{3}\times\frac{1}{4}$ ***A1***

$=\frac{1}{6}$ ***AG***

[2 marks]

b. $G(t)=\frac{1}{3}t+\frac{2}{3}\times\frac{1}{4}t^2+\frac{2}{3}\times\frac{3}{4}\times\frac{1}{3}t^3+\frac{2}{3}\times\frac{3}{4}\times\frac{2}{3}\times\frac{1}{4}t^4+\ldots$ ***M1A1***

$=\frac{1}{3}t\left(1+\frac{1}{2}t^2+\ldots\right)+\frac{1}{6}t^2\left(1+\frac{1}{2}t^2+\ldots\right)$ ***M1A1***

$=\frac{\frac{t}{3}}{1-\frac{t^2}{2}}+\frac{\frac{t^2}{6}}{1-\frac{t^2}{2}}$ ***A1A1***

$=\frac{2t+t^2}{6-3t^2}$ ***AG***

[6 marks]

c. $G'(t)=\frac{(2+2t)(6-3t^2)+6t(2t+t^2)}{(6-3t^2)^2}$ ***M1A1***

$\mathrm{E}(X)=G'(1)=\frac{10}{3}$ ***M1A1***

[4 marks]

Examiners report

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

The weights of the oranges produced by a farm may be assumed to be normally distributed with mean 205 grams and standard deviation 10 grams.

a. Find the probability that a randomly chosen orange weighs more than 200 grams. [2]

b. Five of these oranges are selected at random to be put into a bag. Find the probability that the combined weight of the five oranges is less than 1 kilogram. [4]

c. The farm also produces lemons whose weights may be assumed to be normally distributed with mean 75 grams and standard deviation 3 grams. Find the probability that the weight of a randomly chosen orange is more than three times the weight of a randomly chosen lemon. [5]

Markscheme

a. $z = \frac{200-205}{10} = -0.5$ ***(M1)***

probability = 0.691 (accept 0.692) ***A1***

Note: Award ***M1A0*** for 0.309 or 0.308

[2 marks]

b. let X be the total weight of the 5 oranges

then $\mathrm{E}(X) = 5 \times 205 = 1025$ ***(A1)***

$\mathrm{Var}(X) = 5 \times 100 = 500$ ***(M1)(A1)***

$\mathrm{P}(X < 1000) = 0.132$ ***A1***

[4 marks]

c. let $Y = B - 3C$ where B is the weight of a random orange and C the weight of a random lemon ***(M1)***

$\mathrm{E}(Y) = 205 - 3 \times 75 = -20$ ***(A1)***

$\mathrm{Var}(Y) = 100 + 9 \times 9 = 181$ ***(M1)(A1)***

$\mathrm{P}(Y > 0) = 0.0686$ ***A1***

[5 marks]

Note: Award ***A1*** for 0.0681 obtained from tables

Examiners report

a. As might be expected, (a) was well answered by many candidates, although those who gave 0.6915 straight from tables were given an arithmetic penalty. Parts (b) and (c), however, were not so well answered with errors in calculating the variances being the most common source of incorrect solutions. In particular, some candidates are still uncertain about the difference between nX and $\sum_{i=1}^{n} X_i$.

b. As might be expected, (a) was well answered by many candidates, although those who gave 0.6915 straight from tables were given an arithmetic penalty. Parts (b) and (c), however, were not so well answered with errors in calculating the variances being the most common source of incorrect solutions. In particular, some candidates are still uncertain about the difference between nX and $\sum_{i=1}^{n} X_i$.

c. As might be expected, (a) was well answered by many candidates, although those who gave 0.6915 straight from tables were given an arithmetic penalty. Parts (b) and (c), however, were not so well answered with errors in calculating the variances being the most common source of incorrect solutions. In particular, some candidates are still uncertain about the difference between nX and $\sum_{i=1}^{n} X_i$.

---

a. Determine the probability generating function for $X \sim \mathrm{B}(1,\ p)$. [4]

b. Explain why the probability generating function for $\mathrm{B}(n,\ p)$ is a polynomial of degree n. [2]

c. Two independent random variables X_1 and X_2 are such that $X_1 \sim \mathrm{B}(1,\ p_1)$ and $X_2 \sim \mathrm{B}(1,\ p_2)$. Prove that if $X_1 + X_2$ has a binomial distribution then $p_1 = p_2$. [5]

Markscheme

a. $\mathrm{P}(X = 0) = 1 - p(= q);\ \mathrm{P}(X = 1) = p$ ***(M1)(A1)***

$\mathrm{G}_x(t) = \sum_r \mathrm{P}(X = r)t^r$ (or writing out term by term) ***M1***

$= q + pt$ ***A1***

[4 marks]

b. **METHOD 1**

PGF for $B(n,\ p)$ is $(q + pt)^n$ ***R1***

which is a polynomial of degree n ***R1***

METHOD 2

in n independent trials, it is not possible to obtain more than n successes (or equivalent, *eg*, $\mathrm{P}(X > n) = 0$) ***R1***

so $a_r = 0$ for $r > n$ ***R1***

[2 marks]

c. let $Y = X_1 + X_2$

$G_Y(t) = (q_1 + p_1 t)(q_2 + p_2 t)$ ***A1***

$G_Y(t)$ has degree two, so if Y is binomial then

$Y \sim \mathrm{B}(2,\ p)$ for some p ***R1***

$(q + pt)^2 = (q_1 + p_1 t)(q_2 + p_2 t)$ ***A1***

Note: The LHS could be seen as $q^2 + 2pqt + p^2t^2$.

METHOD 1

by considering the roots of both sides, $\frac{q_1}{p_1} = \frac{q_2}{p_2}$ ***M1***

$\frac{1-p_1}{p_1} = \frac{1-p_2}{p_2}$ ***A1***

so $p_1 = p_2$ ***AG***

METHOD 2

equating coefficients,

$p_1p_2 = p^2,\ q_1q_2 = q^2$ or $(1-p_1)(1-p_2) = (1-p)^2$ ***M1***

expanding,

$p_1 + p_2 = 2p$ so $p_1,\ p_2$ are the roots of $x^2 - 2px + p^2 = 0$ ***A1***

so $p_1 = p_2$ ***AG***

[5 marks]

Total [11 marks]

Examiners report

a. Solutions to (a) were often disappointing with some candidates simply writing down the answer. A common error was to forget the possibility of X being zero so that $G(t) = pt$ was often seen.

b. Explanations in (b) were often poor, again indicating a lack of ability to give a verbal explanation.

c. Very few complete solutions to (c) were seen with few candidates even reaching the result that $(q_1 + p_1t)(q_2 + p_2t)$ must equal $(q + pt)^2$ for some p.

The random variable X is assumed to have probability density function f, where

$$f(x) = \begin{cases} \frac{x}{18}, & 0 \leqslant x \leqslant 6 \\ 0, & \text{otherwise.} \end{cases}$$

Show that if the assumption is correct, then

$$\mathrm{P}(a \leqslant X \leqslant b) = \frac{b^2 - a^2}{36},\ \text{for } 0 \leqslant a \leqslant b \leqslant 6.$$

Markscheme

$\mathrm{P}(a \leqslant X \leqslant b) = \int_a^b \frac{x}{18}\mathrm{d}x$ ***M1A1***

$= \left[\frac{x^2}{36}\right]_a^b$ ***A1***

$= \frac{b^2-a^2}{36}$ ***AG***

[3 marks]

Examiners report

This was the best answered question on the paper, helped probably by the fact that rounding errors in finding the expected frequencies were not an issue. In (a), some candidates thought, incorrectly, that all they had to do was to show that $\int_0^6 f(x)\mathrm{d}x = 1$.

The discrete random variable X has the following probability distribution, where $0 < \theta < \frac{1}{3}$.

x	1	2	3
$\mathrm{P}(X = x)$	θ	2θ	$1-3\theta$

a. Determine $\mathrm{E}(X)$ and show that $\mathrm{Var}(X) = 6\theta - 16\theta^2$. [4]

b. In order to estimate θ, a random sample of n observations is obtained from the distribution of X. [10]

(i) Given that $\bar{X}$ denotes the mean of this sample, show that

$$\hat{\theta}_1 = \frac{3 - \bar{X}}{4}$$

is an unbiased estimator for θ and write down an expression for the variance of $\hat{\theta}_1$ in terms of n and θ.

(ii) Let Y denote the number of observations that are equal to 1 in the sample. Show that Y has the binomial distribution $\mathrm{B}(n,\ \theta)$ and deduce that $\hat{\theta}_2 = \frac{Y}{n}$ is another unbiased estimator for θ. Obtain an expression for the variance of $\hat{\theta}_2$.

(iii) Show that $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$ and state, with a reason, which is the more efficient estimator, $\hat{\theta}_1$ or $\hat{\theta}_2$.

Markscheme

a. $\mathrm{E}(X) = 1 \times \theta + 2 \times 2\theta + 3(1 - 3\theta) = 3 - 4\theta$ ***M1A1***

$\mathrm{Var}(X) = 1 \times \theta + 4 \times 2\theta + 9(1 - 3\theta) - (3 - 4\theta)^2$ ***M1A1***

$= 6\theta - 16\theta^2$ ***AG***

[4 marks]

b. (i) $\mathrm{E}(\hat{\theta}_1) = \frac{3-\mathrm{E}(\bar{X})}{4} = \frac{3-(3-4\theta)}{4} = \theta$ ***M1A1***

so $\hat{\theta}_1$ is an unbiased estimator of θ ***AG***

$\mathrm{Var}(\hat{\theta}_1) = \frac{6\theta-16\theta^2}{16n}$ ***A1***

(ii) each of the n observed values has a probability θ of having the value 1 ***R1***

so $Y \sim \mathrm{B}(n,\ \theta)$ ***AG***

$\mathrm{E}(\hat{\theta}_2) = \frac{\mathrm{E}(Y)}{n} = \frac{n\theta}{n} = \theta$ ***A1***

$\mathrm{Var}(\hat{\theta}_2) = \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}$ ***M1A1***

(iii) $\mathrm{Var}(\hat{\theta}_1) - \mathrm{Var}(\hat{\theta}_2) = \frac{6\theta-16\theta^2-16\theta+16\theta^2}{16n}$ ***M1***

$= \frac{-10\theta}{16n} < 0$ ***A1***

$\hat{\theta}_1$ is the more efficient estimator since it has the smaller variance ***R1***

[10 marks]

Examiners report

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

Bill also has a box with 10 biscuits in it. 4 biscuits are chocolate and 6 are plain. Bill takes a biscuit from his box at random, looks at it and replaces it in the box. He repeats this process until he has looked at 5 biscuits in total. Let B be the number of chocolate biscuits that Bill takes and looks at.

d. State the distribution of B . [1]

e. Find $\mathrm{P}(B = 3)$. [2]

f. Find $\mathrm{P}(B = 5)$. [2]

Markscheme

d. B has the binomial distribution $\left(B\left(5, \frac{4}{10}\right)\right)$ ***A1***

[1 mark]

e. $\mathrm{P}(B = 3) = \left(\binom{5}{3}\left(\frac{4}{10}\right)^3\left(\frac{6}{10}\right)^2 =\right)\frac{144}{625}(= 0.2304)$ ***(M1)A1***

Note: Accept 0.230.

[2 marks]

f. $\mathrm{P}(B = 5) = \left(\left(\frac{4}{10}\right)^5 =\right)\frac{32}{3125}(= 0.01024)$ ***(M1)A1***

Note: Accept 0.0102.

[2 marks]

Examiners report

d. This was generally well answered. Some students did not read the question carefully enough and see the comparisons made between the Hypergeometric distribution and the Binomial distribution, with 5 trials (some candidates went to 10 trials) in each case. Part (h) caused the most problems and it was very rare to see a script that gained the reasoning mark for saying that A and B were independent events. This question was a good indicator of the standard of the rest of the paper.

e. This was generally well answered. Some students did not read the question carefully enough and see the comparisons made between the Hypergeometric distribution and the Binomial distribution, with 5 trials (some candidates went to 10 trials) in each case. Part (h) caused the most problems and it was very rare to see a script that gained the reasoning mark for saying that A and B were independent events. This question was a good indicator of the standard of the rest of the paper.

f. This was generally well answered. Some students did not read the question carefully enough and see the comparisons made between the Hypergeometric distribution and the Binomial distribution, with 5 trials (some candidates went to 10 trials) in each case. Part (h) caused the most problems and it was very rare to see a script that gained the reasoning mark for saying that *A* and *B* were independent events. This question was a good indicator of the standard of the rest of the paper.

The random variable *X* has probability distribution Po(8).

a. (i) Find $\mathrm{P}(X = 6)$. [5]

(ii) Find $\mathrm{P}(X = 6 | 5 \leqslant X \leqslant 8)$.

b. $\bar{X}$ denotes the sample mean of $n > 1$ independent observations from X. [3]

(i) Write down $\mathrm{E}(\bar{X})$ and $\mathrm{Var}(\bar{X})$.

(ii) Hence, give a reason why $\bar{X}$ is not a Poisson distribution.

c. A random sample of 40 observations is taken from the distribution for X. [6]

(i) Find $\mathrm{P}(7.1 < \bar{X} < 8.5)$.

(ii) Given that $\mathrm{P}\left(|\bar{X} - 8| \leqslant k\right) = 0.95$, find the value of k.

Markscheme

a. (i) $\mathrm{P}(X = 6) = 0.122$ ***(M1)A1***

(ii) $\mathrm{P}(X = 6 | 5 \leqslant X \leqslant 8) = \frac{\mathrm{P}(X=6)}{\mathrm{P}(5 \leqslant X \leqslant 8)} = \frac{0.122\ldots}{0.592\ldots - 0.0996\ldots}$ ***(M1)(A1)***

$= 0.248$ ***A1***

[5 marks]

b. (i) $\mathrm{E}(\bar{X}) = 8$ ***A1***

$\mathrm{Var}(\bar{X}) = \frac{8}{n}$ ***A1***

(ii) $\mathrm{E}(\bar{X}) \neq \mathrm{Var}(\bar{X})$ (for $n > 1$) ***R1***

Note: Only award the ***R1*** if the two expressions in (b)(i) are different.

[3 marks]

c. (i) **EITHER**

$\bar{X} \sim \mathrm{N}(8, 0.2)$ ***(M1)A1***

Note: ***M1*** for normality, ***A1*** for parameters.

$\mathrm{P}(7.1 < \bar{X} < 8.5) = 0.846$ ***A1***

OR

The expression is equivalent to

$\mathrm{P}(283 \leqslant \sum X \leqslant 339)$ where $\sum X$ is $\mathrm{Po}(320)$ ***M1A1***

$= 0.840$ ***A1***

Note: Accept 284, 340 instead of 283, 339

Accept any answer that rounds correctly to 0.84 or 0.85.

(ii) **EITHER**

$k = 1.96\frac{\sigma}{\sqrt{n}}$ or $1.96\ \mathrm{std}(\bar{X})$ ***(M1)(A1)***

$k = 0.877$ or $1.96\sqrt{0.2}$ ***A1***

OR

The expression is equivalent to

$P(320 - 40k \leqslant \sum X \leqslant 320 + 40k) = 0.95$ ***(M1)***

$k = 0.875$ ***A2***

Note: Accept any answer that rounds to 0.87 or 0.88.

Award ***M1A0*** if modulus sign ignored and answer obtained rounds to 0.74 or 0.75

[6 marks]

Examiners report

a. [N/A]

b. [N/A]

c. [N/A]

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} 2x, & 0 \leqslant x \leqslant 0.5, \\ \frac{4}{3} - \frac{2}{3}x, & 0.5 \leqslant x \leqslant 2 \\ 0, & \text{otherwise.} \end{cases}$$

a. Sketch the function f and show that the lower quartile is 0.5. [3]

b. (i) Determine $\mathrm{E}(X)$. [4]

(ii) Determine $\mathrm{E}(X^2)$.

c. Two independent observations are made from X and the values are added. [5]

The resulting random variable is denoted Y.

(i) Determine $\mathrm{E}(Y - 2X)$.

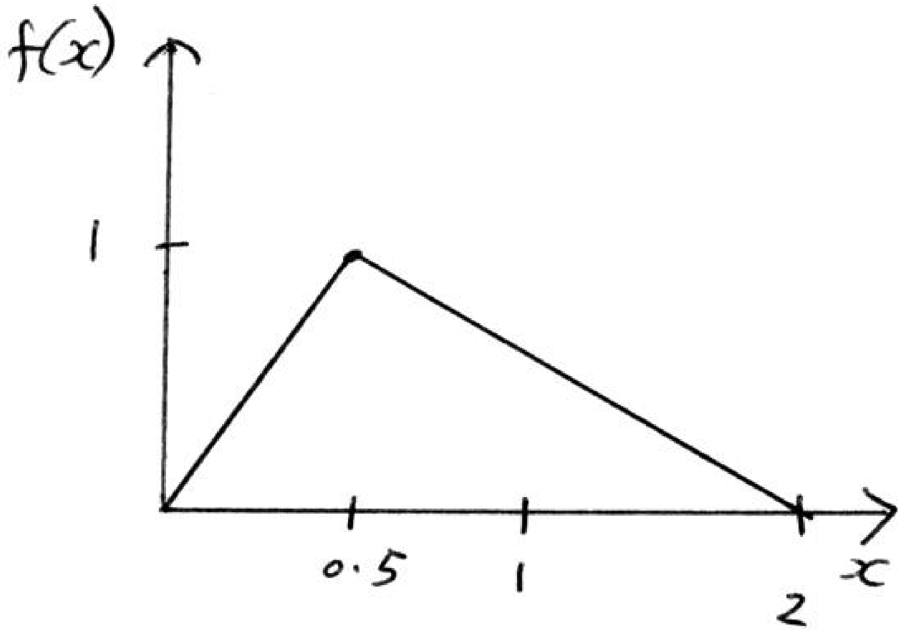
(ii) Determine $\mathrm{Var}\ (Y - 2X)$.

d. (i) Find the cumulative distribution function for X. [7]

(ii) Hence, or otherwise, find the median of the distribution.

Markscheme

a. piecewise linear graph



correct shape ***A1***
with vertices (0, 0), (0.5, 1) and (2, 0) ***A1***
LQ: $x = 0.5$, because the area of the triangle is 0.25 ***R1***
[3 marks]

b. (i) $\mathrm{E}(X) = \int_0^{0.5} x \times 2x\mathrm{d}x + \int_{0.5}^2 x \times \left(\frac{4}{3} - \frac{2}{3}x\right)\mathrm{d}x = \frac{5}{6}\ (= 0.833...)$ ***(M1)A1***

(ii) $\mathrm{E}(X^2) = \int_0^{0.5} x^2 \times 2x\mathrm{d}x + \int_{0.5}^2 x^2 \times \left(\frac{4}{3} - \frac{2}{3}x\right)\mathrm{d}x = \frac{7}{8}\ (= 0.875)$ ***(M1)A1***

[4 marks]

c. (i) $\mathrm{E}(Y - 2X) = 2\mathrm{E}(X) - 2\mathrm{E}(X) = 0$ ***A1***

(ii) $\mathrm{Var}\,(X) = \left(\mathrm{E}(X^2) - \mathrm{E}(X)^2\right) = \frac{13}{72}$ ***A1***
$Y = X_1 + X_2 \Rightarrow \mathrm{Var}\,(Y) = 2\mathrm{Var}\,(X)$ ***(M1)***
$\mathrm{Var}\,(Y - 2X) = 2\mathrm{Var}\,(X) + 4\mathrm{Var}\,(X) = \frac{13}{12}$ ***M1A1***
[5 marks]

d. (i) attempt to use $cf(x) = \int f(u)\mathrm{d}u$ ***M1***

obtain $cf(x) = \begin{cases} x^2, & 0 \leqslant x \leqslant 0.5, \quad \boldsymbol{A1} \\ \frac{4x}{3} - \frac{1}{3}x^2 - \frac{1}{3}, & 0.5 \leqslant x \leqslant 2, \quad \boldsymbol{A2} \end{cases}$

(ii) attempt to solve $cf(x) = 0.5$ ***M1***
$\frac{4x}{3} - \frac{1}{3}x^2 - \frac{1}{3} = 0.5$ ***(A1)***
obtain 0.775 ***A1***

Note: Accept attempts in the form of an integral with upper limit the unknown median.

Note: Accept exact answer $2 - \sqrt{1.5}$.

[7 marks]

Examiners report

a. There was a curious issue about the lower quartile in part (a): The LQ coincides with a quarter of the range of the distribution $\frac{2}{4} = 0.5$. Sadly this is wrong reasoning – the correct reasoning involves a consideration of areas.

b. There was a curious issue about the lower quartile in part (a): The LQ coincides with a quarter of the range of the distribution $\frac{2}{4} = 0.5$. Sadly this is wrong reasoning – the correct reasoning involves a consideration of areas.

In part (b) many candidates used hand calculation rather than their GDC.
The random variable Y was not well understood, and that followed into incorrect calculations involving Y – 2X.

c. There was a curious issue about the lower quartile in part (a): The LQ coincides with a quarter of the range of the distribution $\frac{2}{4} = 0.5$. Sadly this is wrong reasoning – the correct reasoning involves a consideration of areas.

In part (b) many candidates used hand calculation rather than their GDC.
The random variable Y was not well understood, and that followed into incorrect calculations involving Y – 2X.

d. There was a curious issue about the lower quartile in part (a): The LQ coincides with a quarter of the range of the distribution $\frac{2}{4} = 0.5$. Sadly this is wrong reasoning – the correct reasoning involves a consideration of areas.

In part (b) many candidates used hand calculation rather than their GDC.
The random variable Y was not well understood, and that followed into incorrect calculations involving Y – 2X.

A random variable X has probability density function

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \le x < 1 \\ \frac{1}{4} & 1 \le x < 3 \\ 0 & x \ge 3 \end{cases}$$

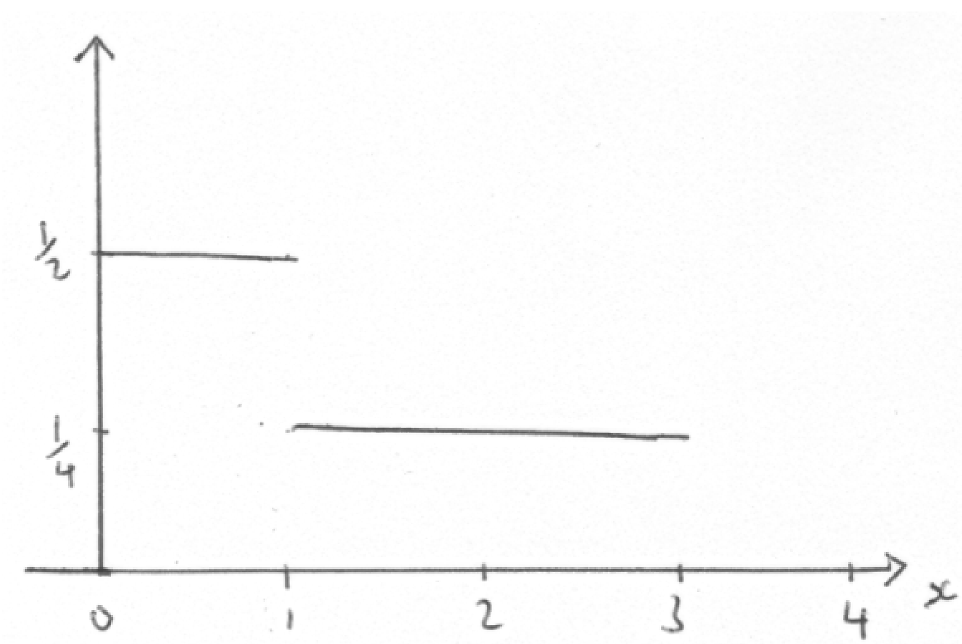
a. Sketch the graph of $y = f(x)$. [1]

b. Find the cumulative distribution function for X. [5]

c. Find the interquartile range for X. [3]

Markscheme

a.



A1

Note: Ignore open / closed endpoints and vertical lines.

Note: Award ***A1*** for a correct graph with scales on both axes and a clear indication of the relevant values.

[1 mark]

b.
$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{2} & 0 \le x < 1 \\ \frac{x}{4} + \frac{1}{4} & 1 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

considering the areas in their sketch or using integration ***(M1)***

$F(x) = 0,\ x < 0,\ F(x) = 1,\ x \ge 3$ ***A1***

$F(x) = \frac{x}{2},\ 0 \le x < 1$ ***A1***

$F(x) = \frac{x}{4} + \frac{1}{4},\ 1 \le x < 3$ ***A1A1***

Note: Accept $<$ for $\le$ in all places and also $>$ for $\ge$ first ***A1***.

[5 marks]

c. $Q_3 = 2,\ Q_1 = 0.5$ ***A1A1***

$\text{IQR is } 2 - 0.5 = 1.5$ ***A1***

[3 marks]

Total [9 marks]

Examiners report

a. Part (a) was correctly answered by most candidates. Some graphs were difficult to mark because candidates drew their lines on top of the ruled lines in the answer book. Candidates should be advised not to do this. Candidates should also be aware that the command term 'sketch' requires relevant values to be indicated.

b. In (b), most candidates realised that the cumulative distribution function had to be found by integration but the limits were sometimes incorrect.

c. In (c), candidates who found the upper and lower quartiles correctly sometimes gave the interquartile range as $[0.5,\ 2]$. It is important for candidates to realise that that the word range has a different meaning in statistics compared with other branches of mathematics.