Maths with Matt

Physics and maths resources

OCR MEI A Level Further Mathematics

H645/Y422 - Statistics Major

Thursday 7 October 2021

Questions and Fully Worked Model Solutions

This booklet places the **question and the solution together**. The mathematics is set out using proper fractions, powers, roots and exact notation, with extra written explanation so that a student can see *why* each step is being taken.

Solutions have been checked against the OCR MEI 2021 mark scheme. The aim is a clear **full-mark model response**, not merely a list of final answers.

Question wording is reproduced from the user-provided 2021 paper for study/reference. OCR retains copyright in the original examination material.

Contents

Question 1 2

Question 2 3

Question 3 5

Question 4 7

Question 5 8

Question 6 10

Question 7 12

Question 8 14

Question 9 16

Question 10 17

Question 11 18

www.mathswithmatt.co.uk

Question 1

Question 1

A random sample of 50 newborn female babies has mean head circumference 34.711 cm and sample standard deviation 1.530 cm.
(a) Determine a 95% confidence interval for the population mean. **(4)**
(b) Explain why this interval can be calculated although the population distribution is unknown. **(2)**

Model solution

(a) We are estimating a population mean from a reasonably large sample. Use the Normal critical value 1.96 *for a two-sided 95% confidence interval, with the sample standard deviation as the estimate of the population standard deviation.*

$$\bar{x} \pm 1.96\frac{s}{\sqrt{n}} = 34.711 \pm 1.96\frac{1.530}{\sqrt{50}} = 34.711 \pm 0.424.$$

The 95% confidence interval is $(34.287, 35.135)$ *cm.*

(4)

(b) The sample size is 50*, which is sufficiently large for the Central Limit Theorem to apply. It says that the sampling distribution of* $\overline{X}$ *is approximately Normal for a large sample, even when the population distribution itself is not known to be Normal.*

Therefore the Normal confidence-interval method is justified.

(2)

Question 2

Question 2

Three fair six-sided dice are rolled. X is the difference between the highest and lowest scores. Its distribution is

r	0	1	2	3	4	5
$P(X = r)$	$\frac{1}{36}$	$\frac{5}{36}$	$\frac{2}{9}$	$\frac{1}{4}$	$\frac{2}{9}$	$\frac{5}{36}$

(a) Show that $P(X = 0) = \frac{1}{36}$. **(2)**
(b) Draw a graph illustrating the distribution. **(2)**
(c) Describe its shape. **(1)**
(d) Find $E(X)$ and $Var(X)$, showing detailed reasoning. **(5)**
The player receives 30X pence. (e) Find the variance of the amount received. **(1)**
(f) The player pays k pence; the organiser's mean profit is 12.5 pence. Find k. **(2)**

Model solution

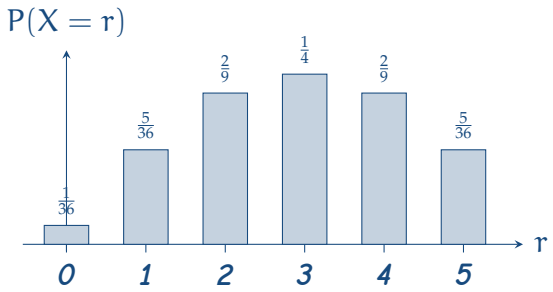
(a) A difference of zero means that all three dice show the same score. There are six such ordered outcomes, one for each common score, out of 6^3 equally likely outcomes.

$$P(X = 0) = \frac{6}{6^3} = \frac{1}{36}.$$

$$P(X = 0) = \frac{1}{36}.$$

(2)

(b) Since X is discrete, use a vertical-bar probability graph.



The bars at $r = 0, 1, 2, 3, 4, 5$ have heights $\frac{1}{36}, \frac{5}{36}, \frac{2}{9}, \frac{1}{4}, \frac{2}{9}, \frac{5}{36}$, respectively.

(2)

(c) The probabilities rise to one peak at 3, then fall; the longer/lower-probability tail is towards smaller values.

The distribution is slightly negatively skewed.

(1)

(d) For a discrete distribution, multiply each possible value by its probability and add. To

find a variance, first find $\mathrm{E}(X^2)$*, then use* $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$*.*

$$\begin{aligned}\mathrm{E}(X) &= 0\left(\frac{1}{36}\right) + 1\left(\frac{5}{36}\right) + 2\left(\frac{2}{9}\right) + 3\left(\frac{1}{4}\right) + 4\left(\frac{2}{9}\right) + 5\left(\frac{5}{36}\right) \\ &= \frac{105}{36} = \frac{35}{12}.\end{aligned}$$

$$\begin{aligned}\mathrm{E}(X^2) &= 0^2\left(\frac{1}{36}\right) + 1^2\left(\frac{5}{36}\right) + 2^2\left(\frac{2}{9}\right) + 3^2\left(\frac{1}{4}\right) + 4^2\left(\frac{2}{9}\right) + 5^2\left(\frac{5}{36}\right) \\ &= \frac{371}{36}.\end{aligned}$$

$$\mathrm{Var}(X) = \frac{371}{36} - \left(\frac{35}{12}\right)^2 = \frac{259}{144} = 1.7986\ldots$$

$\mathrm{E}(X) = \frac{35}{12}$ *and* $\mathrm{Var}(X) = \frac{259}{144} \approx 1.799$*.*

(5)

(e) Multiplying a random variable by 30 multiplies its variance by 30^2*.*

$$\mathrm{Var}(30X) = 30^2\,\mathrm{Var}(X) = 900\left(\frac{259}{144}\right) = 1618.75.$$

The variance is 1619 *pence*^{2} *(to the nearest whole number).*

(1)

(f) The organiser receives k *pence and pays the mean prize* $30\mathrm{E}(X)$*. Set this expected profit equal to 12.5 pence.*

$$k - 30\left(\frac{35}{12}\right) = 12.5 \quad \Rightarrow \quad k - 87.5 = 12.5 \quad \Rightarrow \quad k = 100.$$

$k = 100$ *pence.*

(2)

Question 3

Question 3

Each radio message is misunderstood independently with probability 0.04.
(a) Find the probability of exactly 2 misunderstandings in 50 messages. **(2)**
(b) Find the probability that the 10th message is the first misunderstood message. **(1)**
(c) Find the probability of no misunderstandings in 20 messages. **(1)**
(d) Find the expected number of messages required for 3 misunderstandings. **(3)**
(e) Find the probability that the third misunderstanding is message 60. **(3)**

Model solution

(a) There is a fixed number of independent messages with constant misunderstanding probability, so the count has distribution $X \sim \text{Bin}(50, 0.04)$.

$$P(X = 2) = \binom{50}{2}(0.04)^2(0.96)^{48} = 0.276\ldots$$

0.276.

(2)

(b) The first nine must be understood and message 10 must be misunderstood.

$$(0.96)^9(0.04) = 0.0277\ldots$$

0.0277.

(1)

(c) All 20 messages must be understood.

$$P(\textit{no misunderstandings}) = (0.96)^{20} = 0.442\ldots$$

0.442.

(1)

(d) The number of messages until one misunderstanding is geometric with success probability 0.04*, so its mean is* $1/0.04 = 25$*. Three misunderstandings require three such waiting periods.*

$$E(\textit{messages for 3}) = 3(25) = 75.$$

75 messages.

(3)

(e) For the third misunderstanding to occur at message 60, exactly two of the first 59 messages must be misunderstood, then message 60 must be misunderstood.

$$\binom{59}{2}(0.04)^2(0.96)^{57}(0.04) = \mathrm{P}(\mathrm{Bin}(59, 0.04) = 2)(0.04) = 0.267\ldots(0.04).$$

$0.0107.$

(3)

www.mathswithmatt.co.uk

Question 4

Question 4

A source has 1 000 000 nuclei. Each has probability 1/200 000 of decaying in one second, independently. X is the number decaying in one second.
(a) Explain why either a binomial or a Poisson model may be used. **(3)**
Use a Poisson model for (b) and (c). (b) Find $P(X = 6)$ and $P(X > 6)$. **(3)**
(c) Estimate the probability that at least 60 nuclei decay in 10 seconds. **(2)**

Model solution

(a) We count decays from a fixed number of nuclei. The decays are independent and each has the same probability $p = 1/200\,000$*, so exactly the binomial conditions apply:* $X \sim \text{Bin}(1\,000\,000, 1/200\,000)$*. Also,* n *is very large and* p *is very small, so the Poisson approximation is appropriate, with* $\lambda = np = 5$*.*

Both $\text{Bin}(1\,000\,000, 1/200\,000)$ *and its Poisson approximation* $\text{Po}(5)$ *are suitable.*

(3)

(b) Use $X \sim \text{Po}(5)$*. For the upper-tail probability, use the complement of the cumulative probability through 6.*

$$P(X = 6) = e^{-5}\frac{5^6}{6!} = 0.146\ldots, \qquad P(X > 6) = 1 - P(X \leqslant 6) = 1 - 0.762\ldots = 0.238\ldots$$

$P(X = 6) = 0.146$*,* $P(X > 6) = 0.238$*.*

(3)

(c) Over ten one-second periods the mean count is multiplied by 10, so the estimate is $Y \sim \text{Po}(50)$*.*

$$P(Y \geqslant 60) = 1 - P(Y \leqslant 59) = 1 - 0.9077\ldots = 0.0923\ldots$$

0.0923 *(approximately).*

(2)

Question 5

Question 5

Independent capacitances have distributions $A \sim N(3.9, 0.32^2)$, $B \sim N(7.8, 0.41^2)$, and $C \sim N(30.2, 0.64^2)$.
(a) Find the probability that two Type A and one Type B capacitors total at least 16 units. **(3)**
(b) Find the probability that Type C is within 1 unit of the total of four Type B capacitors. **(3)**
For a new Type C batch, $n = 10$, $\sum x = 299.6$, $\sum x^2 = 8981.0$. Assuming a Normal population but not assuming $\sigma = 0.64$, test at 5% whether its population mean is 30.2 units. **(11)**

Model solution

(a) Sums of independent Normal variables are Normal. Add their means and variances; variances, not standard deviations, add.

$$2A + B \sim N\left(2(3.9) + 7.8,\ 2(0.32)^2 + (0.41)^2\right) = N(15.6, 0.3729).$$

$$P(2A + B \geqslant 16) = P\left(Z \geqslant \frac{16 - 15.6}{\sqrt{0.3729}}\right) = 0.256\ldots$$

0.256.

(3)

(b) "Within 1 unit" is most simply expressed by subtracting Type C from the total of four Type B capacitors: $-1 \leqslant 4B - C \leqslant 1$. This difference is Normal.

$$4B - C \sim N\left(4(7.8) - 30.2,\ 4(0.41)^2 + (0.64)^2\right) = N(1, 1.082).$$

$$P(-1 \leqslant 4B - C \leqslant 1) = P\left(\frac{-2}{\sqrt{1.082}} \leqslant Z \leqslant 0\right) = 0.47274\ldots$$

0.473.

(3)

(c) The population standard deviation is unknown and the sample is small, while the population is stated to be Normal. Therefore use a two-sided one-sample t-test with 9 degrees of freedom.

$$H_0 : \mu = 30.2, \qquad H_1 : \mu \neq 30.2,$$

where μ is the population mean capacitance of capacitors in this new batch.

$$\bar{x} = \frac{299.6}{10} = 29.96, \qquad s^2 = \frac{\sum x^2 - (\sum x)^2/n}{n-1} = \frac{8981.0 - 299.6^2/10}{9} = 0.5538.$$

$$t = \frac{\bar{x} - 30.2}{s/\sqrt{10}} = \frac{29.96 - 30.2}{\sqrt{0.5538}/\sqrt{10}} = -1.020.$$

Under H_0, $t \sim t_9$. For a two-tailed 5% test, the critical values are ± 2.262. Since -1.020 lies between them, the result is not significant.

Do not reject H_0. There is insufficient evidence that the new batch's mean capacitance differs from 30.2 units; assuming the specified mean is reasonable.

(11)

www.mathswithmatt.co.uk

Question 6

Question 6

For 200 one-second periods, the muon frequencies for counts $0, 1, 2, 3, 4, 5, 6, \geqslant 7$ are $34, 65, 55, 24, 14, 6, 2, 0$. The proposed model is Po(1.7).

(a) Use the sample mean and variance to discuss this model. **(3)**

In the spreadsheet, for count 1 the missing cells are C3 (Poisson probability), D3 (expected frequency), and E3 (chi-squared contribution). Find them. **(4)**

(c) Explain why counts 5, 6 and $\geqslant 7$ are combined as $\geqslant 5$. **(1)**

(d) Carry out the 5% goodness-of-fit test. **(6)**

Model solution

(a) A Poisson distribution has equal mean and variance, so calculate both from the frequency table and compare them.

$$\bar{x} = \frac{0(34) + 1(65) + 2(55) + 3(24) + 4(14) + 5(6) + 6(2)}{200} = \frac{345}{200} = 1.725.$$

$$s^2 = \frac{\sum fx^2 - (\sum fx)^2/200}{199} = 1.768\ldots$$

The variance 1.768 *is close to the mean* 1.725*, so the data support the suitability of a Poisson model.*

(3)

(b) For one muon, use the Poisson probability formula with $\lambda = 1.7$*; multiply by 200 for the expected frequency; then use* $(O - E)^2/E$ *for the contribution.*

$$C3 = P(X = 1) = e^{-1.7}\frac{1.7^1}{1!} = 0.3106.$$

$$D3 = 200(0.3106) = 62.12 \quad \textit{(more accurately, } 62.1124\textit{)}.$$

$$E3 = \frac{(65 - 62.1124)^2}{62.1124} = 0.1342.$$

$C3 = 0.3106$, $D3 = 62.1124$, $E3 = 0.1342$.

(4)

(c) The chi-squared approximation requires sufficiently large expected frequencies. Some of the separate high-count categories would have expected frequencies below 5, so they must be pooled.

They are combined to make the expected frequency large enough for a valid chi-squared test.

(1)

(d) The categories used are $0, 1, 2, 3, 4, \geqslant 5$*. Test whether the proposed* $\mathrm{Po}(1.7)$ *probabilities fit the observed frequencies.*

H_0 : *the Poisson* $\mathrm{Po}(1.7)$ *model is a good fit;* H_1 : *it is not a good fit.*

$$\chi^2 = \sum \frac{(\mathrm{O} - \mathrm{E})^2}{\mathrm{E}} = 2.43\ldots$$

There are six categories and no parameter has been estimated from the data, so $\mathrm{df} = 6 - 1 = 5$*. At the 5% level, the critical value is* $\chi^2_{5,0.95} = 11.07$*. Since* $2.43 < 11.07$*, the result is not significant.*

Do not reject H_0*: there is insufficient evidence that* $\mathrm{Po}(1.7)$ *is not a good fit.*

(6)

www.mathswithmatt.co.uk

Question 7

Question 7

For 12 adult women, dominant minus non-dominant hand grip strength has mean 2.79 kg, standard deviation 3.92 kg and 95% confidence interval $(0.29935, 5.28065)$.
(a) Explain using the same women for both measurements. **(2)**
(b) State assumptions needed for this confidence interval. **(2)**
(c) Does the interval support a belief that the mean difference is 2 kg? **(1)**
A separate sample of 100 women gives a 95% Normal confidence interval $(1.94, 2.84)$. (d)(i) Find its mean difference and standard deviation. **(3)**
(d)(ii) State what is needed about this sample to infer about adult women generally. **(2)**

Model solution

(a) This is a paired design: each woman supplies both measurements. Taking a difference within each woman removes variation caused by people naturally having different overall grip strengths, leaving the hand effect being investigated.

Pairing controls for between-person differences and directly compares dominant with non-dominant strength.

(2)

(b) The interval is a small-sample t*-interval for the paired differences. The population of differences (dominant minus non-dominant) must be Normally distributed.*

Assume the parent population of paired differences is Normal.

(2)

(c) A plausible population mean at this confidence level is any value inside the interval.

Yes. The value 2 *kg lies in* $(0.29935, 5.28065)$.

(1)

(d)(i) The centre of a symmetric confidence interval is the sample mean. Its half-width is $1.96s/\sqrt{n}$*, because this interval used a Normal distribution.*

$$\bar{d} = \frac{1.94 + 2.84}{2} = 2.39, \qquad 0.45 = 1.96\frac{s}{\sqrt{100}}.$$

$$s = \frac{0.45\sqrt{100}}{1.96} = 2.2959\ldots$$

Mean difference $= 2.39$ *kg; standard deviation* ≈ 2.30 *kg.*

(3)

(d)(ii) Inference from a sample to a population needs a random sample from that population; otherwise the sampling distribution need not represent adult women.

It must be a random sample of adult women, so that valid inference about that population can be made.

(2)

www.mathswithmatt.co.uk

Question 8

Question 8

(a) For 15 athletes, regression of predicted $VO_{2\,max}$, y, on accurately measured $VO_{2\,max}$, x, is $y = 0.6978x + 15.656$, with $r^2 = 0.5279$.
(i) Estimate y when $x = 50$. **(1)**
(ii) Comment on reliability. **(2)**
(iii) The regression of x on y is $x = 0.7565y + 10.493$. Find where the two regression lines meet. **(1)**
(iv) State what that point represents. **(1)**
(b) For 20 runners with marathon times t between 3.5 and 4.5 hours and measured $VO_{2\,max}$ values v, the scatter is roughly elliptical. Given $\sum t = 80.37$, $\sum v = 970.86$, $\sum t^2 = 324.71$, $\sum v^2 = 47829.24$, $\sum tv = 3886.53$:
(i) Explain why a Pearson correlation test may be valid. **(2)**
(ii) Find the PMCC. **(4)**
(iii) Test at 5% whether there is negative population correlation. **(5)**

Model solution

(a)(i) Substitute the supplied accurate value into the regression line of y on x, because it predicts y from x.

$$\widehat{y} = 0.6978(50) + 15.656 = 50.546.$$

Predicted $VO_{2\,max} = 50.5$ units.

(1)

(a)(ii) The value $x = 50$ is within the observed range, so this is interpolation. However, the points are not close to the line and $r^2 = 0.5279$ is not close to 1, so only about 53% of the variation in y is explained by the linear model.

It is only moderately reliable.

(2)

(a)(iii) At their common point both equations hold. Substitute $y = 0.6978x + 15.656$ into the equation for x.

$$x = 0.7565(0.6978x + 15.656) + 10.493,$$

$$0.4721x = 22.336\ldots \quad \Rightarrow \quad x = 47.3\ldots, \qquad y = 0.6978(47.3\ldots) + 15.656 = 48.7\ldots$$

The lines meet at $(47.3, 48.7)$, approximately.

(1)

(a)(iv) Both regression lines pass through the point formed by the two sample means.

It is $(\overline{x}, \overline{y})$, the centroid of the data.

(1)

(b)(i) Pearson's PMCC test is appropriate when the paired variables plausibly come from a bivariate Normal distribution. The roughly elliptical scatter is evidence consistent with this.

The scatter is very roughly elliptical, so a bivariate Normal model, and hence a PMCC test, may be valid.

(2)

(b)(ii) Calculate the corrected sums of squares and products, then standardise the covariance.

$$S_{tv} = 3886.53 - \frac{80.37(970.86)}{20} = -14.87\ldots$$

$$S_{tt} = 324.71 - \frac{80.37^2}{20} = 1.743\ldots, \qquad S_{vv} = 47829.24 - \frac{970.86^2}{20} = 700.78\ldots$$

$$r = \frac{S_{tv}}{\sqrt{S_{tt}S_{vv}}} = \frac{-14.87\ldots}{\sqrt{(1.743\ldots)(700.78\ldots)}} = -0.4255\ldots$$

$r = -0.4255$ *(approximately).*

(4)

(b)(iii) This is a one-tailed PMCC test because the stated direction is negative.

$$H_0 : \rho = 0, \qquad H_1 : \rho < 0,$$

where ρ *is the population PMCC between marathon time and measured* $VO_{2\,max}$ *for runners in the stated time range. For* $n = 20$*, the 5% one-tailed critical magnitude is* 0.3783*.*

$$|-0.4255| > 0.3783.$$

Reject H_0*. At the 5% level there is sufficient evidence of negative correlation between marathon time and* $VO_{2\,max}$ *for these runners.*

(5)

Question 9

Question 9

X is discrete uniform on all integers from $-n$ to n, inclusive, where n is a positive integer.
(a) Given that n is odd, determine $P(X > n/2)$. **(3)**
(b) Find the variance of the sum of 10 independent values of X, in the form $an^2 + bn$. **(3)**

Model solution

(a) There are $2n+1$ equally likely integer values. Since n is odd, the integers strictly greater than $n/2$ are $(n+1)/2, (n+3)/2, \ldots, n$, giving $(n+1)/2$ favourable values.

$$P\left(X > \frac{n}{2}\right) = \frac{(n+1)/2}{2n+1} = \frac{n+1}{2(2n+1)}.$$

$$\frac{n+1}{2(2n+1)}.$$

(3)

(b) A discrete uniform distribution on $2n+1$ consecutive integers has variance $((2n+1)^2-1)/12$. Independent variances add.

$$\mathrm{Var}(X) = \frac{(2n+1)^2 - 1}{12} = \frac{n^2+n}{3},$$

$$\mathrm{Var}(X_1 + \cdots + X_{10}) = 10\left(\frac{n^2+n}{3}\right).$$

$$\frac{10}{3}n^2 + \frac{10}{3}n.$$

(3)

Question 10

Question 10

Morning and evening waits are uniform on $[0, 10]$ and $[0, 6]$; journey times are $N(25, 4)$ and $N(28, 16)$. All are independent and total return time is T. The spreadsheet simulation contains 500 return journeys.

(a) Estimate $P(T \leqslant 56)$ and $P(T > 61)$. **(2)**

(b) W is Normal with the same mean and variance as T. Find $P(W \leqslant 56)$ and $P(W > 61)$. **(5)**

(c) Explain why a much larger simulation would give $P(T > 61)$ extremely close to $P(W > 61)$. **(2)**

Model solution

(a) Relative frequency in the simulation estimates probability. The spreadsheet gives 104 of the 500 values at most 56. It gives 218 values at most 60 and 288 values at most 62, so interpolate at the midpoint to obtain $(218 + 288)/2 = 253$ values at most 61; then take the complement for the upper-tail probability.

$$P(T \leqslant 56) \approx \frac{104}{500} = 0.208, \qquad P(T > 61) \approx 1 - \frac{253}{500} = \frac{247}{500} = 0.494.$$

$$P(T \leqslant 56) \approx 0.208,\ P(T > 61) \approx 0.494.$$

(2)

(b) Add means and variances of independent components. A uniform distribution on $[a, b]$ has mean $(a + b)/2$ and variance $(b - a)^2/12$.

$$E(T) = 5 + 25 + 3 + 28 = 61,$$

$$\mathrm{Var}(T) = \frac{10^2}{12} + 4 + \frac{6^2}{12} + 16 = \frac{94}{3} = 31.333\ldots$$

$$W \sim N\left(61, \frac{94}{3}\right).$$

$$P(W \leqslant 56) = P\left(Z \leqslant \frac{56 - 61}{\sqrt{94/3}}\right) = 0.186\ldots$$

The Normal distribution is symmetric about its mean 61, so exactly half of its probability is above 61.

$$P(W \leqslant 56) = 0.186,\ P(W > 61) = 0.5.$$

(5)

(c) The mean of T is 61. Each component distribution is symmetric, and the sum of independent symmetric variables is symmetric; therefore T is symmetric about 61 and has $P(T > 61) = 0.5$. With many simulations, the relative frequency converges to this probability.

The simulated $P(T > 61)$ will be extremely close to $0.5 = P(W > 61)$.

(2)

Question 11

Question 11

The continuous random variable X has density

$$f(x) = \begin{cases} ax^2, & 0 \leqslant x \leqslant 2, \\ b(3-x)^2, & 2 \leqslant x \leqslant 3, \\ 0, & \text{otherwise,} \end{cases}$$

where $a, b > 0$.

(a) Given $E(X) = 2$, find a and b. **(5)**

(b) Find the median of X. **(3)**

(c) A random sample of 50 values is taken. Given $\text{Var}(X) = 0.2$, estimate $P(\overline{X} < 1.9)$. **(3)**

Model solution

(a) A density integrates to 1. Together with the given mean, this supplies two equations for a and b.

$$\int_0^2 ax^2 \, dx + \int_2^3 b(3-x)^2 \, dx = 1 \quad \Rightarrow \quad \frac{8a}{3} + \frac{b}{3} = 1,$$

$$8a + b = 3.$$

For the mean, the first moment of the first piece is $\int_0^2 ax^3 \, dx = 4a$, while that of the second piece is $\int_2^3 bx(3-x)^2 \, dx = \frac{3b}{4}$. Thus

$$4a + \frac{3b}{4} = 2.$$

$$8a + b = 3, \qquad 16a + 3b = 8 \quad \Rightarrow \quad a = \frac{1}{8}, \quad b = 2.$$

$$a = \tfrac{1}{8},\ b = 2.$$

(5)

(b) First find the probability below 2. This tells us whether the median lies in the first or second part of the density.

$$F(2) = \int_0^2 \frac{1}{8}x^2 \, dx = \frac{1}{3} < \frac{1}{2}.$$

So the median m is between 2 and 3. Set the area up to m equal to one half.

$$\frac{1}{3} + \int_2^m 2(3-x)^2 \, dx = \frac{1}{2},$$

$$\int_2^m 2(3-x)^2 \, dx = \frac{1}{6} \quad \Rightarrow \quad \frac{2}{3} - \frac{2}{3}(3-m)^3 = \frac{1}{6}.$$

$$(3-m)^3 = \frac{3}{4} \quad \Rightarrow \quad m = 3 - \sqrt[3]{\frac{3}{4}} = 2.0914\ldots$$

$$\textit{Median} = 3 - \sqrt[3]{\tfrac{3}{4}} \approx 2.09.$$

(3)

(c) The sample is large, so the Central Limit Theorem gives an approximately Normal distribution for the sample mean. Its mean remains 2 and its variance is divided by 50.

$$\overline{X} \sim N\left(2, \frac{0.2}{50}\right) = N(2, 0.004).$$

$$P(\overline{X} < 1.9) = P\left(Z < \frac{1.9-2}{\sqrt{0.004}}\right) = P(Z < -1.5811\ldots) = 0.0569\ldots$$

The estimated probability is 0.0569.

(3)

www.mathswithmatt.co.uk