

Continuous random variables A2

Read online at pastopic.com/cambridge/further-mathematics-9231/notes/continuous-random-variables

Last checked 1 October 2026

© 2026 Pastopic. Free for personal study. Please share the link, not the file.

What you need to know

By the end of this topic you should be able to:

1. **Work with a probability density function (pdf) $f(x)$** , including one that is **defined piecewise** (a different formula on different intervals). Probabilities are areas under its graph, the total area is 1, and you can use this to find unknown constants.
2. **Find the expectation of any function of X** with the general result $E(g(X)) = \int g(x) f(x) dx$. This covers $E(X)$ and $\text{Var}(X)$, but also things like $E(\sqrt{X})$, $E(\frac{1}{X})$ or $\text{Var}(X^2)$.
3. **Move between the pdf and the cumulative distribution function (CDF) $F(x) = P(X \leq x)$** : integrate f to get F , differentiate F to get f . Use either one to find probabilities and **percentiles**: the median, the quartiles, the interquartile range, or any " p th percentile".
4. **Find the distribution of a related variable Y** in simple cases, such as $Y = X^2$, $Y = \sqrt{X}$ or $Y = X^3$: first find its CDF $G(y)$ from F , then differentiate to get its pdf $g(y)$.

The normal distribution is used in other topics; this topic is about working from a given f or F .

Understand it

Probability as area

A **continuous** random variable can take any value in an interval, such as a waiting time or a length. The chance of hitting one exact value, like $X = 2.000 \dots$, is 0, so we only ever ask for the probability that X lies in an interval. That probability is the **area under the pdf** over the interval:

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Because single points have probability 0, $P(X < b)$ and $P(X \leq b)$ are the same.

A function can only be a pdf if

- $f(x) \geq 0$ everywhere (a negative area would be a negative probability), and
- the **total area is 1**: $\int_{\text{all } x} f(x) dx = 1$.

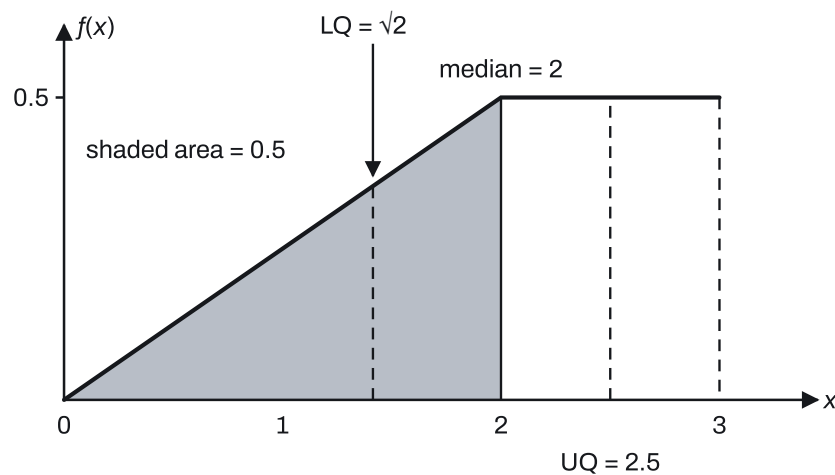
Outside the interval where the formula is given, $f(x) = 0$; always write "0 otherwise" as part of the definition.

A piecewise pdf

Here is a pdf in two pieces:

$$f(x) = \begin{cases} \frac{1}{4}x & 0 \leq x < 2, \\ \frac{1}{2} & 2 \leq x \leq 3, \\ 0 & \text{otherwise.} \end{cases}$$

The first piece is a triangle with area $\frac{1}{2} \times 2 \times \frac{1}{2} = \frac{1}{2}$; the second is a rectangle with area $1 \times \frac{1}{2} = \frac{1}{2}$. The total is 1, so this is a valid pdf. The graph does not have to join up where the pieces meet (this one happens to), but it may never go below the axis.



When you integrate a piecewise pdf, split the integral at the joins and add the pieces. If the formula has an unknown constant k , setting the total area equal to 1 gives an equation for k . Areas of triangles and rectangles are a quick check when a piece is a straight line.

The cumulative distribution function

The **CDF** gives the area from the far left up to x :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt.$$

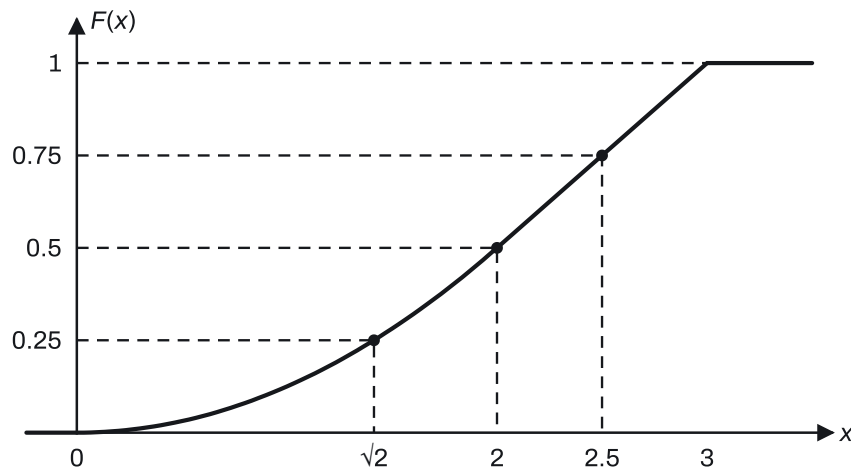
So F climbs from 0 to 1 and never goes down, and $f(x) = F'(x)$: the pdf is the gradient of the CDF.

For the pdf above:

- for $0 \leq x < 2$: $F(x) = \int_0^x \frac{1}{4}t dt = \frac{1}{8}x^2$, so $F(2) = \frac{1}{2}$;
- for $2 \leq x \leq 3$: all of the first piece ($\frac{1}{2}$) plus the part of the second piece up to x : $F(x) = \frac{1}{2} + \int_2^x \frac{1}{2} dt = \frac{1}{2} + \frac{1}{2}(x - 2) = \frac{1}{2}(x - 1)$.

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{8}x^2 & 0 \leq x < 2, \\ \frac{1}{2}(x-1) & 2 \leq x \leq 3, \\ 1 & x > 3. \end{cases}$$

The second piece must give the same value as the first at the join ($F(2) = \frac{1}{2}$ from both), because a CDF has no jumps. Another way to get the second piece: integrate $\frac{1}{2}$ to get $\frac{1}{2}x + c$, then choose c so that $F(2) = \frac{1}{2}$ (or so that $F(3) = 1$). Forgetting this constant is the most common error in the topic.



With F you can read probabilities straight off: $P(1 < X < 2.5) = F(2.5) - F(1) = \frac{3}{4} - \frac{1}{8} = \frac{5}{8}$.

Median, quartiles and percentiles

The **median** m splits the area in half: $F(m) = \frac{1}{2}$. The **lower quartile** solves $F(q_1) = \frac{1}{4}$, the **upper quartile** $F(q_3) = \frac{3}{4}$, and the p **th percentile** solves $F(x) = \frac{p}{100}$. The **interquartile range** is $q_3 - q_1$.

With a piecewise F , first decide **which piece** the answer is in by checking the value of F at the join. Here $F(2) = \frac{1}{2}$:

- median: $F(2) = \frac{1}{2}$ exactly, so $m = 2$;
- lower quartile: $\frac{1}{4} < \frac{1}{2}$, so use the first piece: $\frac{1}{8}q_1^2 = \frac{1}{4}$, $q_1 = \sqrt{2}$;
- upper quartile: $\frac{3}{4} > \frac{1}{2}$, so use the second piece: $\frac{1}{2}(q_3 - 1) = \frac{3}{4}$, $q_3 = \frac{5}{2}$.

The interquartile range is $\frac{5}{2} - \sqrt{2} = 1.09$ (3 s.f.). An equation such as $\frac{1}{8}q^2 = \frac{1}{4}$ has two roots; keep only the one inside the piece's interval.

Expectation of a function of X

For a discrete variable, $E(X) = \sum x p$. For a continuous one the sum becomes an integral:

$$E(X) = \int x f(x) dx, \quad E(g(X)) = \int g(x) f(x) dx.$$

Multiply the pdf by whatever you want the expectation of, and integrate over every piece. Then

$$\text{Var}(X) = E(X^2) - (E(X))^2.$$

For the pdf above:

- $E(X) = \int_0^2 \frac{1}{4}x^2 dx + \int_2^3 \frac{1}{2}x dx = \frac{2}{3} + \frac{5}{4} = \frac{23}{12};$
- $E(X^2) = \int_0^2 \frac{1}{4}x^3 dx + \int_2^3 \frac{1}{2}x^2 dx = 1 + \frac{19}{6} = \frac{25}{6};$
- $\text{Var}(X) = \frac{25}{6} - \left(\frac{23}{12}\right)^2 = \frac{71}{144} = 0.493 \text{ (3 s.f.)};$
- $E\left(\frac{1}{X}\right) = \int_0^2 \frac{1}{x} \cdot \frac{1}{4}x dx + \int_2^3 \frac{1}{x} \cdot \frac{1}{2} dx = \frac{1}{2} + \frac{1}{2} \ln \frac{3}{2} = 0.703 \text{ (3 s.f.)}.$

The same idea gives the variance of a function: $\text{Var}(g(X)) = E(g(X)^2) - (E(g(X)))^2$, so $\text{Var}(\sqrt{X}) = E(X) - (E(\sqrt{X}))^2$ and $\text{Var}(X^2) = E(X^4) - (E(X^2))^2$.

A related variable $Y = g(X)$

Suppose $Y = \sqrt{X}$ for the X above. You cannot just substitute into the pdf. Go through the **CDF**, because a CDF is a probability, and probabilities can be rewritten:

$$G(y) = P(Y \leq y) = P(\sqrt{X} \leq y) = P(X \leq y^2) = F(y^2).$$

This works because $\sqrt{}$ is an increasing function on $x \geq 0$, so squaring both sides keeps the inequality the right way round. So:

1. **Replace x by y^2 in every piece of F :** $G(y) = \frac{1}{8}y^4$ and $G(y) = \frac{1}{2}(y^2 - 1)$.
2. **Change the intervals too.** X runs from 0 to 3, so $Y = \sqrt{X}$ runs from 0 to $\sqrt{3}$, and the join at $x = 2$ moves to $y = \sqrt{2}$:

$$G(y) = \begin{cases} 0 & y < 0, \\ \frac{1}{8}y^4 & 0 \leq y < \sqrt{2}, \\ \frac{1}{2}(y^2 - 1) & \sqrt{2} \leq y \leq \sqrt{3}, \\ 1 & y > \sqrt{3}. \end{cases}$$

3. **Differentiate** to get the pdf of Y : $g(y) = \frac{1}{2}y^3$ for $0 \leq y < \sqrt{2}$, $g(y) = y$ for $\sqrt{2} \leq y \leq \sqrt{3}$, and 0 otherwise.

In general, for an increasing g with inverse $x = g^{-1}(y)$, $G(y) = F(g^{-1}(y))$. Two useful facts follow:

- **Percentiles carry across.** Because $G(y) = F(x)$ when $y = g(x)$, the median of Y is $g(\text{median of } X)$. Here the median of Y is $\sqrt{2}$.
- **Means do not.** In general $E(g(X)) \neq g(E(X))$. To find $E(Y)$, either integrate $y g(y)$ over Y 's interval, or integrate $g(x) f(x)$ over X 's interval: both give the same answer.

Key facts and formulas

"Given" means it is in the MF19 list of formulae in the exam. "Learn it" means you must remember it.

Formula	Status
$E(X) = \int x f(x) dx$	Given
$\text{Var}(X) = \int x^2 f(x) dx - (E(X))^2$	Given
$\int_{\text{all } x} f(x) dx = 1 \text{ and } f(x) \geq 0$	Learn it
$P(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$	Learn it
$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$, and $f(x) = F'(x)$	Learn it
$F = 0$ below the range and $F = 1$ above it; F has no jumps	Learn it
$E(g(X)) = \int g(x) f(x) dx$	Learn it
$\text{Var}(g(X)) = E(g(X)^2) - (E(g(X)))^2$	Learn it
Median m : $F(m) = \frac{1}{2}$; quartiles: $F(q_1) = \frac{1}{4}$, $F(q_3) = \frac{3}{4}$; p th percentile: $F(x) = \frac{p}{100}$	Learn it
$Y = g(X)$ with g increasing: $G(y) = F(g^{-1}(y))$, then $g_Y(y) = G'(y)$	Learn it

For a chi-squared goodness-of-fit test, the expected frequency of a class is $n \times P(X \text{ in the class})$, found from the pdf.

How to do it

In the Paper 4 questions we have tagged for this topic (34 questions, 2021–2025; 11 repeat a question from another paper of the same session, so there are 23 different questions), the types came up this often. Almost every question is a chain of parts: 32 of the 34 mix two or more types, so the counts add up to 103.

Question type	Questions
Median, quartiles and percentiles	27
Expectation and variance, including $E(g(X))$	22
The distribution of $Y = g(X)$	18
Finding unknown constants	16
Finding the CDF from the pdf, or the pdf from the CDF	13
Probabilities and expected frequencies	4
Sketching the pdf	3

1. Median, quartiles and percentiles (27 questions)

1. **Get** $F(x)$ for every piece (see type 5), unless you already have it.

2. **Find F at each join** and see which interval the target probability ($\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{p}{100}$) falls in.
3. **Solve $F(x) = \text{target}$** using that piece only. If you would rather use areas, remember to include the whole of the earlier piece: $F(\text{join}) + \int_{\text{join}}^m f(x) dx = \frac{1}{2}$.
4. **Keep the root inside the interval** and say why you reject the other one. Give it exactly if asked ("exact value").

For the pdf in "Understand it", $F(2) = \frac{1}{2}$, so any quartile below the median is in the first piece and any above it is in the second.

Variations you will meet:

- **Interquartile range:** find both quartiles (often in different pieces) and subtract. "Exact" means keep surds, for example $\frac{5}{2} - \sqrt{2}$.
- **Other percentiles:** the 10th percentile solves $F(x) = 0.1$, the 90th $F(x) = 0.9$, the 40th $F(x) = 0.4$. In Example 2, $F(1) = \frac{1}{7} \approx 0.143$, so the 10th percentile is in the first piece.
- **"Find a such that $P(X \leq a) = \dots$ "** is the same job with a different target.
- **Area-based shortcuts.** When the right-hand piece is a straight line down to 0, the area to the right of m is a triangle, and $\frac{1}{2}(\text{base})(\text{height}) = \frac{1}{2}$ can be quicker than integrating.
- **The equation is a quadratic in x^2 or in $\sqrt{x}$.** For example $x^4 - 13x^2 + 36 = 0$ is a quadratic in x^2 : $x^2 = 4$ or 9. Solve for the square, then take the root that lies in the interval.
- **A percentile of $Y = g(X)$:** either solve $G(y) = \text{target}$, or find the percentile of X and apply g to it (valid because g is increasing).
- **Comparing medians** ("is the median of Y bigger than the median of X ?"): work out both, or work out $G(m_X)$ and compare it with $\frac{1}{2}$. If $G(m_X) < \frac{1}{2}$, the median of Y is bigger than m_X . For $Y = \sqrt{X}$, the median of Y is $\sqrt{m_X}$, which is bigger than m_X when $0 < m_X < 1$ and smaller when $m_X > 1$.

2. Expectation and variance, including $E(g(X))$ (22 questions)

1. **Write the integrand** as (the thing you want the expectation of) $\times f(x)$, for **each** piece, with that piece's limits.
2. **Integrate and add** the pieces.
3. For a variance, find $E(X^2)$ the same way, then $\text{Var}(X) = E(X^2) - (E(X))^2$.

Variations you will meet:

- **Given $E(X)$, find $\text{Var}(X)$:** you still have to find $E(X^2)$; use the given mean in the formula (keep it as a fraction).
- **E of a function**, such as $E(\frac{1}{X})$, $E(\sqrt{X})$ or $E(X^2)$: multiply f by $\frac{1}{x}$, $\sqrt{x}$ or x^2 and simplify the powers before integrating, for example $x^2 \times \frac{1}{x^3} = x^{-1}$, which integrates to a logarithm.
- **Variance of a function:** $\text{Var}(\sqrt{X}) = E(X) - (E(\sqrt{X}))^2$; $\text{Var}(X^2) = E(X^4) - (E(X^2))^2$.
- **Only F is given:** differentiate first to get f , then integrate $xf(x)$. Never integrate $x F(x)$.
- **$E(Y)$ after finding $g(y)$:** integrate $y g(y)$ over Y 's interval, or integrate $g(x)f(x)$ over X 's interval; the second is often easier and checks the first.

- **Exponentials and logarithms** can appear, for example f built from e^x or e^{-x} terms; the method is the same, and $\int \frac{1}{x} dx = \ln x$ gives exact log answers.

3. The distribution of $Y = g(X)$ (18 questions)

1. **Get $F(x)$** , every piece with its interval (0 below, 1 above).
2. **Write $G(y) = P(Y \leq y) = P(X \leq g^{-1}(y)) = F(g^{-1}(y))$** . For $Y = X^2$ substitute $x = \sqrt{y}$; for $Y = \sqrt{X}$ substitute $x = y^2$; for $Y = X^3$ substitute $x = y^{1/3}$; for $Y = X^{1/3}$ substitute $x = y^3$.
3. **Transform every interval end**, including the joins: if X goes from a to b with a join at c , then Y goes from $g(a)$ to $g(b)$ with a join at $g(c)$.
4. **Differentiate G** to get $g(y)$ if the pdf is asked for. Write "0 otherwise".

Variations you will meet:

- **Only the CDF of Y is asked:** stop after step 3, but give the full definition, with 0 and 1 outside the range.
- **The pdf of Y is asked:** the CDF route above is the safest. Some mark schemes also accept the chain-rule shortcut $g_Y(y) = f(x(y)) \times \frac{dx}{dy}$, applied to each piece.
- **Other increasing functions**, such as $Y = (X + 1)^2$ on $X \geq 0$ or $Y = e^{2X}$: the same steps, with $x = \sqrt{y} - 1$ or $x = \frac{1}{2} \ln y$.
- **Follow-on parts:** the median or a percentile of Y (type 1), or $E(Y)$ (type 2).

4. Finding unknown constants (16 questions)

Every condition you are given becomes one equation. Collect as many equations as unknowns and solve.

1. **Total area is 1:** integrate every piece, with the constants left in, and set the sum equal to 1. For straight-line pieces, triangle and trapezium areas are quicker.
2. **Continuity**, when you are told (or shown in a diagram) that the graph of f has no break: the two formulas are equal at the join.
3. **A given probability or percentile:** for example " $P(X < 1) = 0.3$ " or "the lower quartile is 1.5" gives $\int f dx =$ that probability over the right interval.
4. **A given mean:** $\int x f(x) dx =$ the given value (see Quick check 5).
5. **Constants in a CDF:** use $F = 0$ at the bottom of the range, $F = 1$ at the top, and matching values at each join, because F has no jumps.

"Show that $k = \dots$ " is a given answer: write the integral with its limits, the integrated expression, the equation and its solution. A missing step loses the mark.

5. Finding the CDF from the pdf, or the pdf from the CDF (13 questions)

pdf to CDF:

1. Integrate the first piece from its lower limit to x .

- For each later piece, either add $F(\text{join})$ to the integral from the join to x , or integrate with $+c$ and choose c so that the pieces match at the join (or so that $F = 1$ at the top).
- Write the full definition with **four lines** for a two-piece pdf: 0 below the range, the two pieces with their intervals, and 1 above. Check $F(\text{top}) = 1$.

CDF to pdf: differentiate each piece, keep its interval, and add "0 otherwise". Check the result is never negative.

6. Probabilities and expected frequencies (4 questions)

- A probability:** $P(a < X < b) = F(b) - F(a)$ is the quickest route once you have F ; otherwise integrate f from a to b , splitting at any join.
- Expected frequencies for a goodness-of-fit test:** for each class, expected frequency $= n \times P(X \text{ in the class})$, with the probability from F or from integrating f . Show the integral with its limits when you are asked to show a given value. Then run the chi-squared test as in that topic, combining classes whose expected frequency is below 5.

7. Sketching the pdf (3 questions)

- If you are given F , differentiate each piece first.
- Draw each piece over its own interval; a straight line needs only its end values.
- Label the x -values of the ends and the joins on the axis, and the height at any peak or corner.
- Show $f = 0$ outside the range. Do not add parts of the formula outside its interval.

Worked examples

The bold codes show where marks are earned, as in Cambridge mark schemes. The number is how many marks.

- M1** method mark: for a correct method, even if the numbers slip.
 - A1** accuracy mark: for a correct answer; it needs the M mark before it.
 - B1** mark for a correct result on its own, with no method needed.
-

Example 1

The continuous random variable X has pdf $f(x) = k(x^2 + 1)$ for $0 \leq x \leq 3$, and $f(x) = 0$ otherwise.

- Show that $k = \frac{1}{12}$. **[2]**
- Find $E(X)$ and $\text{Var}(X)$. **[4]**
- Find the cumulative distribution function of X , and hence find $P(X > 2)$. **[4]**

(a) $\int_0^3 k(x^2 + 1) dx = k \left[\frac{1}{3}x^3 + x \right]_0^3 = k(9 + 3) = 12k = 1$, so $k = \frac{1}{12}$. **M1** for integrating and setting equal to 1, **A1** for the given answer with the working shown.

$$(b) E(X) = \frac{1}{12} \int_0^3 (x^3 + x) dx = \frac{1}{12} \left[\frac{1}{4}x^4 + \frac{1}{2}x^2 \right]_0^3 = \frac{1}{12} \left(\frac{81}{4} + \frac{9}{2} \right) = \frac{33}{16}. \text{ **M1 A1**}$$

$$E(X^2) = \frac{1}{12} \int_0^3 (x^4 + x^2) dx = \frac{1}{12} \left(\frac{243}{5} + 9 \right) = \frac{24}{5}.$$

$$\text{Var}(X) = \frac{24}{5} - \left(\frac{33}{16} \right)^2 = \frac{699}{1280} = 0.546 \text{ (3 s.f.)}. \text{ **M1** for the variance formula with their values, **A1**}$$

$$(c) \text{ For } 0 \leq x \leq 3: F(x) = \frac{1}{12} \int_0^x (t^2 + 1) dt = \frac{x^3 + 3x}{36}. \text{ **M1 A1**}$$

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{x^3 + 3x}{36} & 0 \leq x \leq 3, \\ 1 & x > 3. \end{cases}$$

A1 for the 0 and 1 with correct intervals.

$$P(X > 2) = 1 - F(2) = 1 - \frac{8 + 6}{36} = \frac{11}{18}. \text{ **B1**}$$

Example 2

The continuous random variable X has pdf

$$f(x) = \begin{cases} kx^2 & 0 \leq x < 1, \\ k & 1 \leq x \leq 3, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{3}{7}$. **[2]**

(b) Find the cumulative distribution function of X . **[4]**

(c) Find the median and the 10th percentile of X . **[4]**

$$(a) \int_0^1 kx^2 dx + \int_1^3 k dx = \frac{1}{3}k + 2k = \frac{7}{3}k = 1, \text{ so } k = \frac{3}{7}. \text{ **M1 A1**}$$

$$(b) \text{ For } 0 \leq x < 1: F(x) = \int_0^x \frac{3}{7}t^2 dt = \frac{1}{7}x^3. \text{ **B1**}$$

At the join, $F(1) = \frac{1}{7}$. For $1 \leq x \leq 3$: $F(x) = \frac{1}{7} + \int_1^x \frac{3}{7} dt = \frac{1}{7} + \frac{3}{7}(x - 1) = \frac{3x - 2}{7}$. **M1** for including the first piece's total (or finding the constant), **A1**

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{7}x^3 & 0 \leq x < 1, \\ \frac{1}{7}(3x - 2) & 1 \leq x \leq 3, \\ 1 & x > 3. \end{cases}$$

A1 for the full definition. Check: $F(3) = \frac{7}{7} = 1$.

(c) $F(1) = \frac{1}{7} < \frac{1}{2}$, so the median is in the second piece: $\frac{1}{7}(3m - 2) = \frac{1}{2}$, $3m = \frac{11}{2}$, $m = \frac{11}{6}$. **M1 A1**

$F(1) = \frac{1}{7} \approx 0.143 > 0.1$, so the 10th percentile is in the first piece: $\frac{1}{7}x^3 = 0.1$, $x^3 = 0.7$, $x = 0.888$ (3 s.f.). **M1 A1**

Example 3

The continuous random variable X has pdf $f(x) = \frac{2}{x^2}$ for $1 \leq x \leq 2$, and $f(x) = 0$ otherwise. The random variable Y is defined by $Y = X^3$.

(a) Find the cumulative distribution function of X . **[2]**

(b) Find the probability density function of Y . **[4]**

(c) Find the exact median of Y , and find $E(Y)$. **[4]**

(a) For $1 \leq x \leq 2$: $F(x) = \int_1^x 2t^{-2} dt = [-2t^{-1}]_1^x = 2 - \frac{2}{x}$. With $F(x) = 0$ for $x < 1$ and $F(x) = 1$ for $x > 2$. **M1 A1**

(b) $G(y) = P(X^3 \leq y) = P(X \leq y^{1/3}) = F(y^{1/3}) = 2 - 2y^{-1/3}$. **M1** for replacing x by $y^{1/3}$.

x runs from 1 to 2, so y runs from $1^3 = 1$ to $2^3 = 8$. **B1**

$g(y) = G'(y) = \frac{2}{3}y^{-4/3}$ for $1 \leq y \leq 8$, and $g(y) = 0$ otherwise. **M1** for differentiating, **A1**

(c) $G(y) = \frac{1}{2}$: $2 - 2y^{-1/3} = \frac{1}{2}$, so $y^{-1/3} = \frac{3}{4}$, $y^{1/3} = \frac{4}{3}$ and the median is $y = \frac{64}{27}$. **M1 A1**

(Check: the median of X solves $2 - \frac{2}{m} = \frac{1}{2}$, $m = \frac{4}{3}$, and $(\frac{4}{3})^3 = \frac{64}{27}$.)

$$E(Y) = \int_1^8 y \cdot \frac{2}{3}y^{-4/3} dy = \frac{2}{3} \left[\frac{3}{2}y^{2/3} \right]_1^8 = 4 - 1 = 3. \text{ **M1 A1**}$$

(Check the other way: $E(X^3) = \int_1^2 x^3 \cdot \frac{2}{x^2} dx = [x^2]_1^2 = 3$. Note that $E(X) = 2 \ln 2 \approx 1.386$ and $1.386^3 \approx 2.66$, not 3: the mean does not carry across, the median does.)

Example 4

The continuous random variable X has cumulative distribution function

$$F(x) = \begin{cases} 0 & x < 0, \\ ax^3 & 0 \leq x < 2, \\ 1 - b(4 - x)^2 & 2 \leq x \leq 4, \\ 1 & x > 4, \end{cases}$$

where a and b are constants. The probability density function of X is continuous at $x = 2$.

(a) Show that $a = \frac{1}{20}$ and find b . [3]

(b) Sketch the graph of the probability density function of X . [2]

(c) Find $E(X)$. [3]

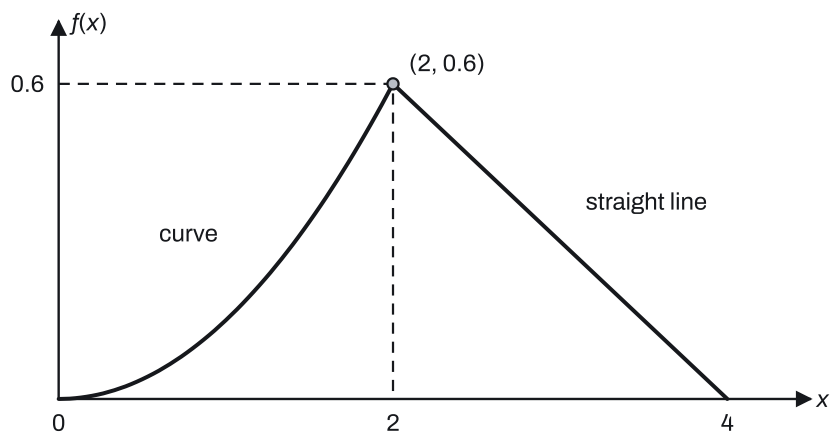
(d) Find the interquartile range of X . [4]

(a) F has no jump at 2: $8a = 1 - 4b$. **B1**

The pdf is $f(x) = 3ax^2$ on the first piece and $2b(4 - x)$ on the second. Continuous at 2: $12a = 4b$, so $b = 3a$. **M1** for differentiating and equating at $x = 2$.

Then $8a = 1 - 12a$, so $a = \frac{1}{20}$ and $b = \frac{3}{20}$. **A1**

(b) $f(x) = \frac{3}{20}x^2$ for $0 \leq x < 2$, $f(x) = \frac{3}{10}(4 - x)$ for $2 \leq x \leq 4$, 0 otherwise. It rises as a curve to $(2, 0.6)$, then falls in a straight line to $(4, 0)$. **B1** for the shape, **B1** for 2, 4 and 0.6 labelled.



(c) $E(X) = \int_0^2 \frac{3}{20}x^3 dx + \int_2^4 \frac{3}{10}(4x - x^2) dx = \frac{3}{20}(4) + \frac{3}{10}\left(\frac{16}{3}\right) = \frac{3}{5} + \frac{8}{5} = \frac{11}{5}$. **M1** for $xf(x)$ on both pieces, **M1** for correct limits, **A1**

(d) $F(2) = \frac{8}{20} = \frac{2}{5}$. **B1**

Lower quartile: $\frac{1}{4} < \frac{2}{5}$, so first piece: $\frac{1}{20}q_1^3 = \frac{1}{4}$, $q_1^3 = 5$, $q_1 = \sqrt[3]{5} = 1.710$. **M1**

Upper quartile: $\frac{3}{4} > \frac{2}{5}$, so second piece: $1 - \frac{3}{20}(4 - q_3)^2 = \frac{3}{4}$, $(4 - q_3)^2 = \frac{5}{3}$, $q_3 = 4 - \sqrt{\frac{5}{3}} = 2.709$ (taking the root below 4). **M1**

Interquartile range = $2.709 - 1.710 = 0.999$ (3 s.f.). **A1**

Common mistakes

- **Leaving out the constant in the second piece of the CDF.** Integrating the second piece of f and stopping gives a CDF that jumps at the join. Examiners see this every year. Add $F(\text{join})$, or find c from $F(\text{join})$ or $F(\text{top}) = 1$, and check that the pieces match.
- **Writing " $F(x) = 0$ otherwise".** A CDF is 0 **below** the range and 1 **above** it. Give them as two separate lines with their intervals.
- **Leaving out "0 otherwise" for a pdf,** or giving the wrong value outside the range.
- **Using the wrong piece for a median or quartile.** A common error is to solve $\int_{\text{join}}^m f \, dx = \frac{1}{2}$ and forget the probability already in the first piece. Find F at the join first.
- **Keeping both roots.** A quadratic for a percentile often has one root outside the range; reject it and say so.
- **Setting the pdf equal to $\frac{1}{2}$ (or 0.3) instead of the CDF.** Percentiles come from F , never from f .
- **Not changing the interval for Y .** If X runs from 1 to 2 and $Y = X^3$, then Y runs from 1 to 8. Transform the joins too.
- **Finding $G(y)$ and stopping** when the pdf of Y was asked for: differentiate. Mixing up G and g when finding $E(Y)$ is also common: integrate $y g(y)$, not $y G(y)$.
- **Integrating the pdf the long way** for a probability when F is already known. $F(b) - F(a)$ is quicker and safer.
- **Rounding too early.** Keep fractions (like $\frac{11}{6}$ or $\frac{33}{16}$) until the end; a rounded mean squared in the variance formula can change the third significant figure. Give exact answers when the question says "exact".

How to get the marks

- **"Show that $k = \dots$ ":** set up the integral with its limits (or the areas), show the integrated expression, form the equation and solve it. The answer is given, so every step must be visible.
- **"Find the CDF":** the method mark is for integrating each piece; the accuracy marks need the right constant and a complete definition, covering all real x , with 0 below and 1 above. Boundaries may overlap (writing $\leq$ on both sides of a join is accepted), but there must be no gaps.
- **"Find the pdf":** differentiate each piece and state "0 otherwise"; the intervals must cover everything.
- **Medians and percentiles:** the method mark is for choosing the right piece and setting $F = \text{target}$. Show F at the join so the examiner sees why you chose it. Give one answer only.
- **Expectation and variance:** show $\int x f(x) \, dx$ (or $x^2 f(x)$, or $g(x) f(x)$) with limits for each piece. For a variance the mark scheme wants to see $E(X^2) - (E(X))^2$ with numbers in.
- **Related variable:** marks go for substituting into F (not f), for the new intervals, and for differentiating. If you use the chain-rule shortcut instead, apply it to every piece.
- **"Determine whether"** (for example comparing two medians): a calculation plus a clear conclusion in words.
- **Sketches:** correct shape for each piece, with the key x -values (ends and joins) labelled.
- **Accuracy:** 3 significant figures unless told otherwise; exact (fractions, surds, logs) when the question says "exact". Method marks follow through from an earlier wrong answer, so keep going.

Quick check

1. The continuous random variable X has pdf $f(x) = k(2x + 3)$ for $0 \leq x \leq 1$, and 0 otherwise. Find k , the exact median of X , and $E(X)$.
2. The continuous random variable X has CDF $F(x) = \frac{1}{4}x^2(3 - x)$ for $0 \leq x \leq 2$, with $F(x) = 0$ for $x < 0$ and $F(x) = 1$ for $x > 2$. Find the pdf of X , $P(0.5 < X < 1.5)$ and $\text{Var}(X)$.
3. The continuous random variable X has CDF $F(x) = \frac{1}{8}x^3$ for $0 \leq x \leq 2$ (0 below, 1 above). Find the pdf of $Y = X^2$ and the median of Y .
4. A model for X has pdf $f(x) = \frac{1}{8}x$ for $0 \leq x \leq 4$, and 0 otherwise. For 64 observations, find the expected frequencies for the classes $0 \leq x < 1$, $1 \leq x < 2$, $2 \leq x < 3$ and $3 \leq x \leq 4$. Which classes should be combined before a chi-squared test?
5. The continuous random variable X has pdf $f(x) = a + bx$ for $0 \leq x \leq 2$, and 0 otherwise. Given that $E(X) = \frac{13}{12}$, find a and b .

Answers

1. $k[x^2 + 3x]_0^1 = 4k = 1$, so $k = \frac{1}{4}$. $F(x) = \frac{1}{4}(x^2 + 3x)$; $F(m) = \frac{1}{2}$ gives $m^2 + 3m - 2 = 0$, so $m = \frac{-3+\sqrt{17}}{2}$ (the other root is negative). $E(X) = \frac{1}{4} \int_0^1 (2x^2 + 3x) dx = \frac{1}{4} \left(\frac{2}{3} + \frac{3}{2} \right) = \frac{13}{24}$.
2. $f(x) = F'(x) = \frac{1}{4}(6x - 3x^2) = \frac{3}{4}x(2 - x)$ for $0 \leq x \leq 2$, 0 otherwise. $P(0.5 < X < 1.5) = F(1.5) - F(0.5) = \frac{27}{32} - \frac{5}{32} = \frac{11}{16}$. The graph of f is symmetrical about $x = 1$, so $E(X) = 1$; $E(X^2) = \frac{3}{4} \int_0^2 (2x^3 - x^4) dx = \frac{3}{4} \left(8 - \frac{32}{5} \right) = \frac{6}{5}$, so $\text{Var}(X) = \frac{6}{5} - 1 = \frac{1}{5}$.
3. $G(y) = F(\sqrt{y}) = \frac{1}{8}y^{3/2}$ for $0 \leq y \leq 4$, so $g(y) = \frac{3}{16}y^{1/2}$ for $0 \leq y \leq 4$, 0 otherwise. Median: $\frac{1}{8}y^{3/2} = \frac{1}{2}$, $y^{3/2} = 4$, $y = 4^{2/3} = 2.52$ (3 s.f.).
4. $F(x) = \frac{x^2}{16}$, so the class probabilities are $\frac{1}{16}, \frac{3}{16}, \frac{5}{16}, \frac{7}{16}$ and the expected frequencies are 4, 12, 20 and 28. The first is below 5, so combine the first two classes (expected 16).
5. Total area: $2a + 2b = 1$. Mean: $\int_0^2 (ax + bx^2) dx = 2a + \frac{8}{3}b = \frac{13}{12}$. Subtracting, $\frac{2}{3}b = \frac{1}{12}$, so $b = \frac{1}{8}$ and $a = \frac{3}{8}$.

Past questions to try

- **9231/42/M/J/22 Q3**: a constant, a variance from a given mean, and a median that needs the right piece.
- **9231/42/O/N/22 Q4**: a constant, an interquartile range across two pieces, and a variance.
- **9231/43/M/J/23 Q1**: CDF, the pdf of a related variable, and an exact median.
- **9231/42/O/N/24 Q4**: CDF, the pdf of $Y = (X - 1)^4$, its median and $E(Y)$.