DP IB Maths: AA HL

4.7 Further Probability Distributions

Contents

4.7.1 Probability Density Function

Calculating Probabilities using PDF

A **continuous random variable** can take *any* value in an interval so is typically used when continuous quantities are involved (time, distance, weight, etc)

What is a probability density function (p.d.f.)?

- For a continuous random variable, a function can be used to model probabilities
  - This function is called a **probability density function** (p.d.f.), denoted by f(x)
- For f(x) to represent a p.d.f. the following conditions must apply
  - $f(x) \geq 0$ for **all** values of x
  - The **area** under the graph of y = f(x) must **total** 1
- In most problems, the **domain** of x is restricted to an interval, $a \leq X \leq b$ say, with all values of x outside of the interval having f(x)=0

How do I find probabilities using a probability density function (p.d.f.)?

- The probability that the continuous random variable X lies in the interval $a \leq X \leq b$, where X has the probability density function f(x), is given by

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

  - $P(a \leq X \leq b) = P(a < X < b)$
    - For **any** continuous random variable (including the normal distribution) P(X = n) = 0
    - One way to think of this is that a = b in the integral above
- For **linear** functions it can be easier to find the probability using the area of geometric shapes
  - Rectangles: A = bh
  - Triangles: A = ½(bh)
  - Trapezoids: A = ½(a+b)h

How do I determine whether a function is a pdf?

- Some questions may ask for justification of the use of a given function for a probability density function
  - In such cases check that the function meets the two conditions
    - $f(x) \geq 0$ for **all** values of x
    - **total area** under the graph is 1

How do I use a pdf to find probabilities?

STEP 1
Identify the **probability density function,** f(x) - this may be given as a **graph,** an **equation** or as a **piecewise function**

e.g. $f(x) = \begin{cases} 0.02x & 0 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$

Identify the **limits** of X for a particular problem

Remember that P(a ≤ X ≤ b) = P(a < X < b)

STEP 2

Sketch, or use your GDC to draw, the graph of y = f(x)

Look for basic shapes (rectangles, triangles and trapezoids) as finding these areas is easier without using integration

Look for symmetry in the graph that may make the problem easier

Break the area required into two or more parts if it makes the problem easier

STEP 3

Find the area(s) required using basic shapes or integration and answer the question

- Trickier problems may involve finding a limit of the integral given its value
  - i.e. Find one of the boundaries in the domain of X, given the probability
    - e.g. Find the value of a given that P(0 ≤ X ≤ a) = 0.09

Worked example

The continuous random variable, X, has probability density function.

$$f(x) = \begin{cases} 0.08x & 0 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$$

a) Show that f(x) can represent a probability density function.

There are <u>two</u> conditions for a function to be a p.d.f.

$$\int_0^5 0.08x \, dx = \left[\frac{0.08x^2}{2} \right]_0^5$$

$$= \left[0.04x^2 \right]_0^5$$

$$= 0.04 \times 5^2 - 0 = 1$$

$$\therefore \int_{-\infty}^{\infty} f(x) \, dx = 1$$

Also, $f(x) \geqslant 0$ for all values of x
so $f(x)$ meets both conditions to represent
a probability density function.

b) Find, both geometrically and using integration, P(0 ≤ X ≤ 2).

STEP 1: Identify the p.d.f. - given piecewise in the question

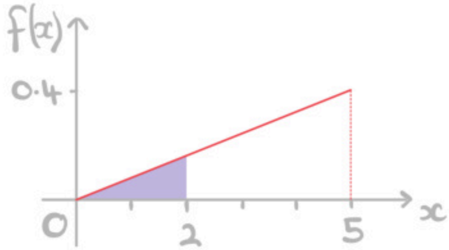
Identify the limits of X

$f(x) = 0.08x$

$P(0 \leq X \leq 2)$: Lower limit 0

Upper limit 2

STEP 2: Sketch, or use GDC to draw, $y = f(x)$



The probability required is the area of a triangle

STEP 3: Find the required area(s) and answer the question

Geometrically

$A = \frac{1}{2} \times 2 \times (\underbrace{0.08 \times 2}_{f(2)}) = 0.16$

Using integration

$A = \int_0^2 0.08x \, dx$

$A = \left[0.04x^2\right]_0^2 = 0.16$

$\boxed{\therefore P(0 \leq X \leq 2) = 0.16}$

c) Write down $P(X = 3.2)$.

$\boxed{P(X = 3.2) = 0}$ $P(X = n) = 0$ for all values of n

Median & Mode of a CRV

What is meant by the median of a continuous random variable?

- The **median,** *m*, of a **continuous random variable,** *X*, with **probability density function** *f(x)* is defined as the value of *X* such that

$$\mathrm{P}(X < m) = \mathrm{P}(X > m) = 0.5$$

- Since $\mathrm{P}(X = m) = 0$ this can also be written as $\mathrm{P}(X \le m) = \mathrm{P}(X \ge m) = 0.5$
- **If** the p.d.f. is **symmetrical** (i.e. if the graph of $y = f(x)$ is symmetrical) then the median will be equal to the *x*-coordinate of the line of symmetry
  - In such cases the graph of $y = f(x)$ has an **axis** of **symmetry** in the line $x = m$

How do I find the median of a continuous random variable?

- The **median,** *m*, of a continuous random variable, *X*, with probability density function *f(x)* is defined as the value of *X* such that

$$\int_{-\infty}^{m} f(x)\ \mathrm{d}x = \frac{1}{2}$$

or

$$\int_{m}^{\infty} f(x)\ \mathrm{d}x = \frac{1}{2}$$

- The **equation** that should be used will depend on the **information** in the **question**
  - If the **graph** of $y = f(x)$ is **symmetrical,** symmetry may be used to **deduce** the **median**
  - This may often be the case if *f(x)* is **linear** and the **area under the graph** is a basic **shape** such as a **rectangle**

How do I find the median of a continuous random variable with a piecewise p.d.f.?

- For **piecewise functions,** the **location** of the **median** will determine **which equation** to use in order to find it
  - For example
    - if $f(x) = \begin{cases} \frac{1}{5}x & 0 \le x \le 2 \\ \frac{2}{15}(5-x) & 2 \le x \le 5 \\ 0 & \text{otherwise} \end{cases}$
    - then $\int_{0}^{2} \frac{1}{5}x\ \mathrm{d}x = 0.4$ so the median must lie in the interval $2 \le x \le 5$

- so to find the median, m, solve $\int_2^m \frac{2}{15}(5-x)\,\mathrm{d}x = 0.1$
  ('0.4 of the area' already used for $0 \le x \le 2$)
- Use a GDC to plot the function and evalutae integral(s)

What is meant by the mode of a continuous random variable?

- The **mode** of a **continuous random variable,** *X*, with **probability density function** *f(x)* is the **value** of *x* that produces the **greatest value** of *f(x)*

How do I find the mode of a continuous random variable?

- This will depend on the **type** of **function** *f(x)*; the easiest way to find the **mode** is by considering the **shape** of the **graph** of *y = f(x)*
- If the **graph** is a **curve** with a **maximum point,** the **mode** can be **found** by **differentiating** and **solving** *f'(x) = 0*
  - If there is **more than one solution** to *f'(x) = 0* then **further work** may be needed in deducing the mode
    - There could be **more than one** mode
    - Look for **valid values** of *x* from the **domain** of the p.d.f.
    - Use the **second derivative** (*f''(x)*) to **deduce** the **nature** of each **stationary point**
    - **Check** the **values** of *f(x)* at the **lower** and **upper limits** of *x*, one of these could be the **maximum value** *f(x)* reaches

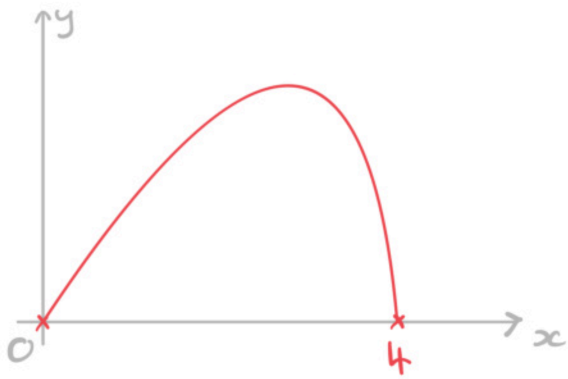
Worked example

The continuous random variable *X* has probability function *f(x)* defined as

$$f(x) = \frac{1}{64}(16x - x^3) \quad 0 \le x \le 4$$

a) Find the median of X, giving your answer to three significant figures.

Sketch the graph of $y = f(x)$ using your GDC to help



For the median, solve "$\int_{-\infty}^{m} f(x)\,dx = \frac{1}{2}$"

$$\frac{1}{64}\int_{0}^{m} (16x - x^3)\,dx = \frac{1}{2}$$

$f(x) = 0$ for $x < 0$

$$\left[8x^2 - \frac{1}{4}x^4\right]_0^m = 32$$

$$8m^2 - \frac{1}{4}m^4 = 32$$

$$m^4 - 32m^2 + 128 = 0$$

This is a 'hidden quadratic' in m^2.

Using a GDC,

$m = \pm 5.226\,251\ldots$

or $m = \pm 2.164\,784\ldots$

Only one of these four values lies in the range $0 \leq x \leq 4$

$$\boxed{\therefore \text{ Median, } m = 2.16 \quad (3 \text{ s.f.})}$$

b) Find the *exact* value of the mode of X.

Differentiate, solving $f'(x)=0$ to find the mode

$f'(x) = \frac{1}{64}(16-3x^2)$

$16-3x^2=0$

Using a GDC (ensure you get exact answers)

$x = \pm\frac{4\sqrt{3}}{3}$

Clearly from sketch of graph, $x=\frac{4}{3}\sqrt{3}$ is a (local) maximum

Also, $x = -\frac{4\sqrt{3}}{3}$ does not lie in the interval $0 \le x \le 4$

$\boxed{\therefore \text{Mode} = \frac{4}{3}\sqrt{3}}$

Mean & Variance of a CRV

What are the mean and variance of a continuous random variable?

- E(*X*) is the **expected value,** or **mean,** of the **continuous random variable** *X*
  - E(*X*) can also be denoted by μ
- Var(*X*) is the **variance** of the continuous random variable *X*
  - Var(*X*) can also be denoted by σ^2
  - The **standard deviation,** σ, is the **square root** of the **variance**

How do I find the mean and variance of a continuous random variable?

- The mean is given by

$$\mu = \mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

  - This is given in the **formula booklet**
  - If the graph of $y = f(x)$ has **axis** of **symmetry,** $x = a$, then $\mathrm{E}(X) = a$
- The **variance** is given by

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$

where $\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$

  - This is given in the **formula booklet**
  - Another version of the variance is given in the **formula booklet**

$$\mathrm{Var}(x) = \int_{-\infty}^{\infty} (x-\mu)^2\ f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x - \mu^2$$

  - but the first version above is usually more practical for solving problems
- Be careful about confusing E(X^2) and [E(*X*)]^2
  - $\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$ "mean of the squares"
  - $[\mathrm{E}(X)]^2 = \left[\int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x\right]^2$ "square of the mean"

How do I find the mean and variance of a linear transformation of a continuous random variable?

- For the **continuous random variable,** *X*, with **mean** E(*X*) and **variance** Var(*X*) then

$$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$$

and

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

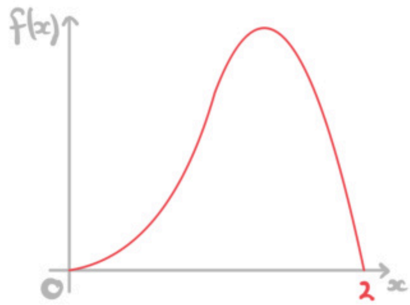
Worked example

A continuous random variable, X, is modelled by the probability distribution function, $f(x)$, such that

$$f(x) = \begin{cases} 1.5x^2(1-0.5x) & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

a) Find the mean of X.

Use your GDC to sketch the graph of $y = f(x)$



No symmetry!

$\therefore \mu = \int_0^2 x\left[1.5x^2(1-0.5x)\right] dx$ Evaluate using your GDC

$\boxed{\mu = 1.2}$

To do without GDC.

$\mu = \int_0^2 (1.5x^3 - 0.75x^4)\, dx$

$\mu = \left[0.375x^4 - 0.15x^5\right]_0^2$

$\mu = 6 - 4.8 = 1.2$

b) Find standard deviation of X.

$\sigma = \sqrt{Var(X)}$ $\quad Var(X) = E(X^2) - [E(X)]^2$

$$E(X^2) = \int_0^2 x^2 [1.5x^2(1-0.5x)]\,dx$$

Using GDC, $E(X^2) = 1.6$

$$\therefore \sigma = \sqrt{\underbrace{1.6}_{E(X^2)} - (\underbrace{1.2}_{E(X)})^2} = \sqrt{0.16} = 0.4$$

$$\boxed{\sigma = 0.4}$$