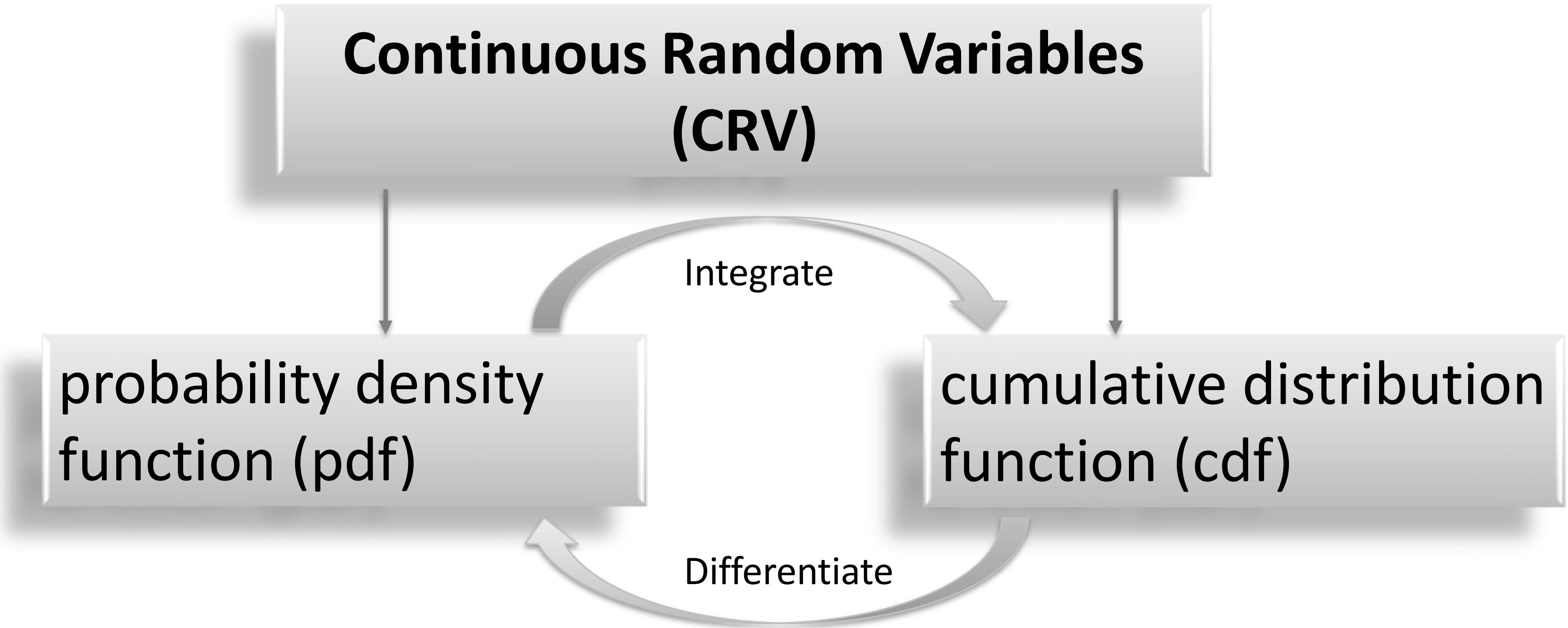
NUFYP Mathematics

7.1 Continuous Random Variables (CRV) – part 1

Catherine Armes, Joohee Hong, Adina Amanbekkyzy

Lecture outline



Textbook Reference

The content of this lecture is from the following textbook:

Chapter 3

Statistics 2 Edexcel AS and A Level Modular Mathematics S2 published by Pearson Education Limited

ISBN 978 0 435519 13 1

Further examples can be found in the textbook.

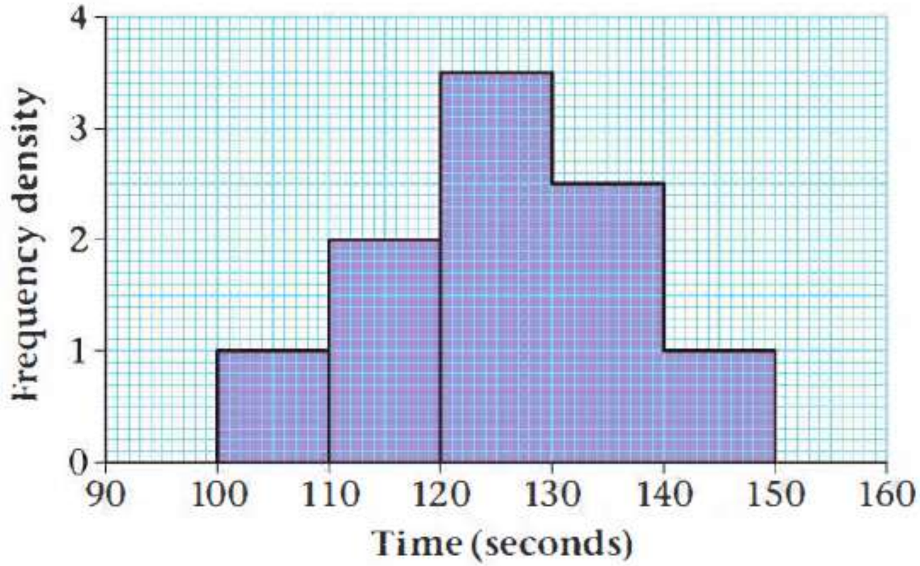
Revisit: Discrete random variables

- We have been studying discrete random variables (drv's) and a binomial distribution.
- We saw that, because the variable X was discrete, X could only take specific values in a range.
- Also, we could express the probabilities using a function, called a probability function.
- We used the summation notation, $\sum P(X = x)$, to sum the probabilities.
- The sum of all of the probabilities equalled one.

7.1.1 Understand the concept of a continuous random variable and its probability density function

- Continuous random variables **can take any value**.
- As with drv's, we can express the probabilities for crv's using a function called a **probability density function (pdf)**.
- Again as with drv's, for crv's **the sum of all of the probabilities is equal to one**.
- We will use a **method of integration to sum all of the probabilities** since the Sigma method only works for discrete values of X.

Consider the histogram which shows the variable t, the time in seconds taken by 100 children to do their homework.



Recall:

Area of a bar $=$ frequency density $\times$ 1 cm$=k\times$ frequency

Total area $=k\times$ total frequency

Total area $= k \times$ total frequency

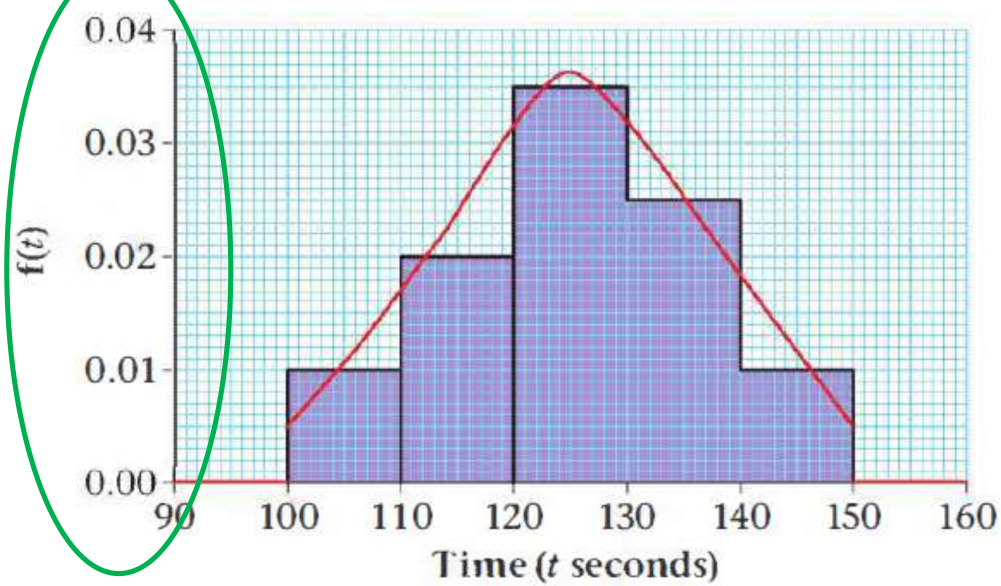
Given that the total frequency is n, to make the total area equal to 1,

$$k = \frac{1}{n}$$

Using $k = \frac{1}{n}$, we can rescale the histogram as follows.

$$n = 100 \rightarrow k = 0.01$$

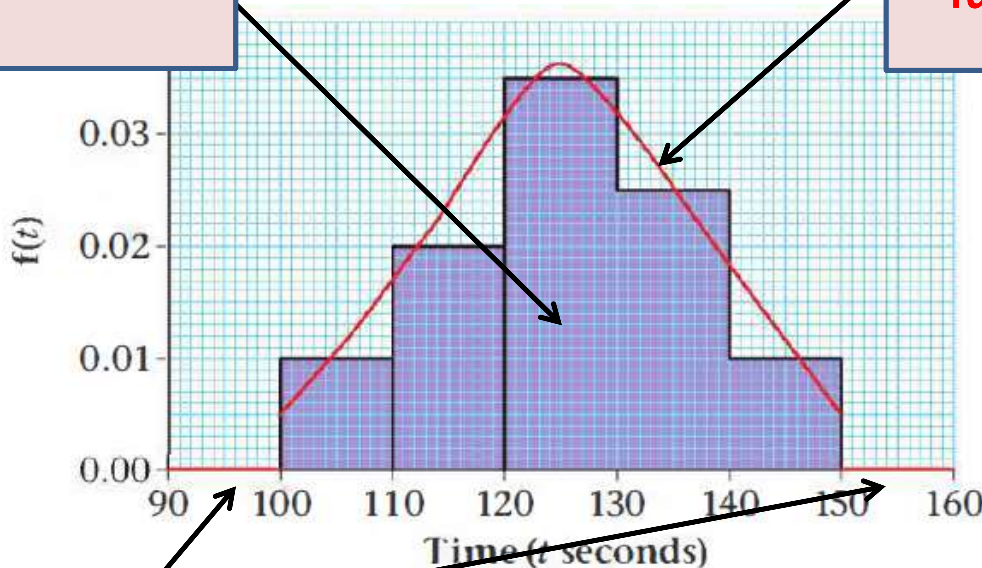
Note that the numbers are changed to make the total area 1



Now imagine drawing a smooth curve instead of the histogram, shown below with a red line:

The total area between the pdf and the x-axis is 1

This red line is the **probability density function** (**pdf** for short)



The probability is zero outside the range

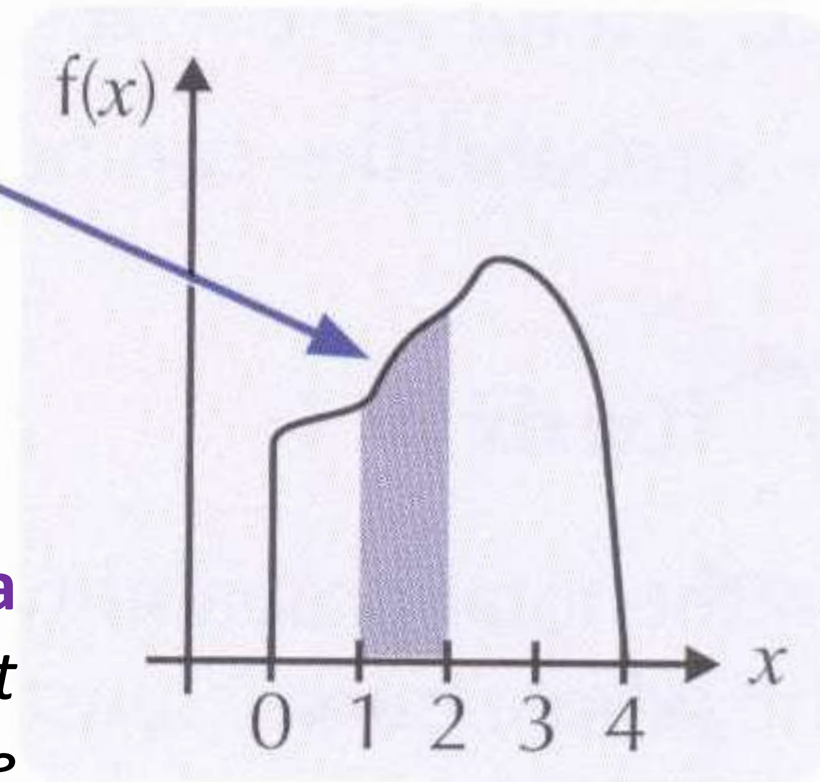
We can use definite integrals to find an area which represents a probability

Interpretation of p.d.f. graphs

- $f(x)$ is a p.d.f.

- The area under a p.d.f. shows the probability that the random variable will take a value in that range.

- For example, the **shaded area** shows *the probability that this CRV will take a value between 1 and 2*.

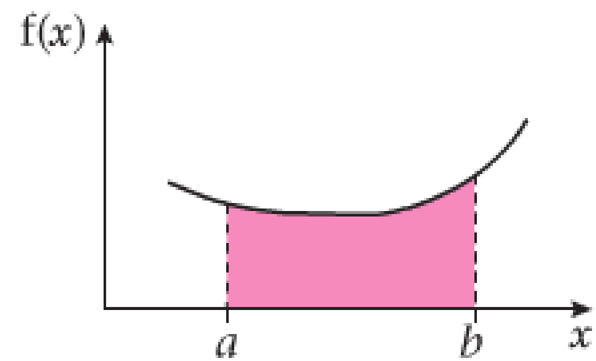


Properties of pdf

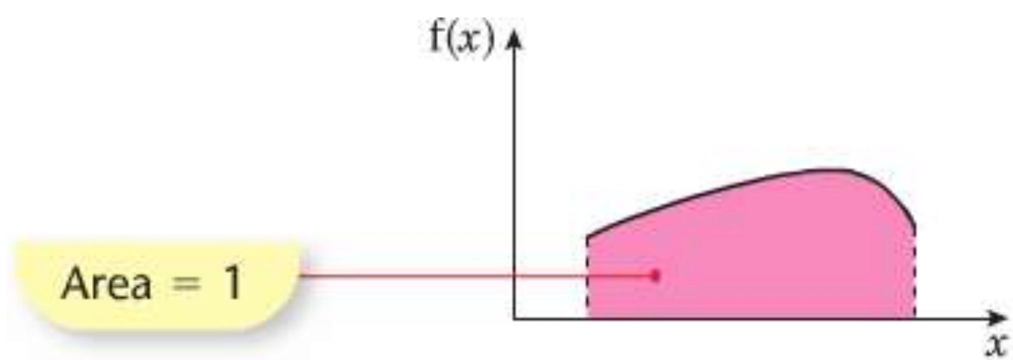
If X is a continuous random variable with p.d.f. $f(x)$ then:

1. $f(x) \geq 0$ since we cannot have negative probabilities.

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



3. $\int_{-\infty}^{\infty} f(x)dx = 1$



Example 1

Which of the following could be a probability density function?

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

b $f(x) = \begin{cases} k(x-2), & 3 \leqslant x \leqslant 5, \\ 0, & \text{otherwise.} \end{cases}$

c $f(x) = \begin{cases} kx(x-2), & 1 \leqslant x \leqslant 3, \\ 0, & \text{otherwise.} \end{cases}$

Need to check:

(1) $f(x) \geq 0$ for the given range

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

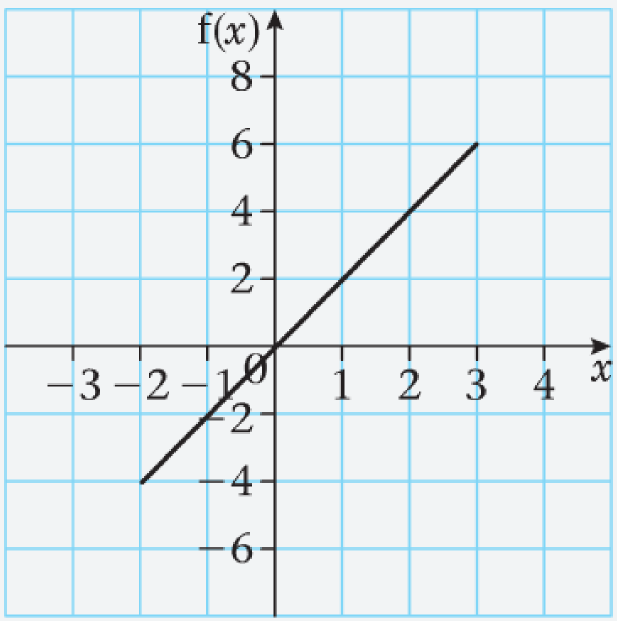
Solution

$$\textbf{a}\quad f(x) = \begin{cases} 2x, & -2 \leqslant x \leqslant 3, \\ 0, & \text{otherwise.} \end{cases}$$

(a)

Start with a sketch of f(x) for the given values of x.

a

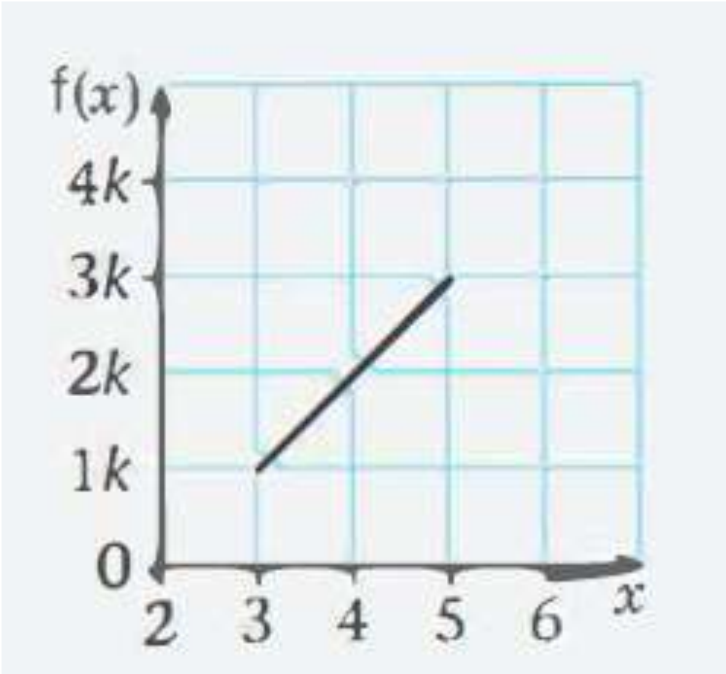


It is not a p.d.f. since $f(x) < 0$ for $x < 0$

This is not a pdf

(To be a pdf, f(x) must be positive for all values of x.)

(b)



b $f(x) = \begin{cases} k(x-2), & 3 \leqslant x \leqslant 5, \\ 0, & \text{otherwise.} \end{cases}$

This is a pdf for $k = \frac{1}{4}$

(1) $f(x)$ is positive between $x = 3$ and $x = 5$

(2) The total area under pdf between $x = 3$ and $x = 5$ should be 1.

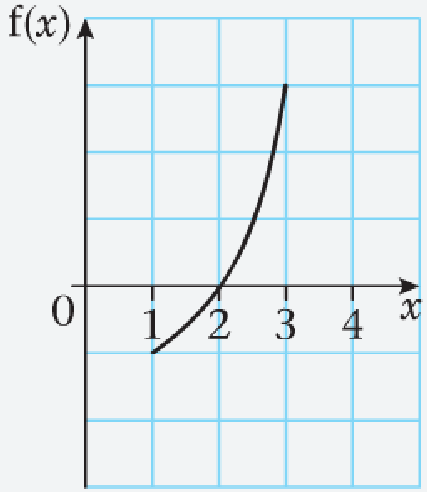
Total area $= \frac{1}{2} \times 2 \times 4k = 4k$

If $k = \frac{1}{4}$, then the total area becomes 1.

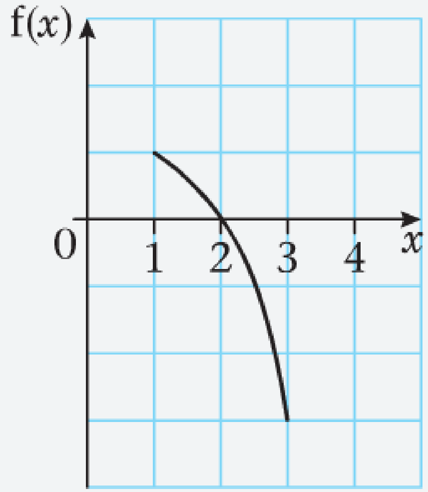
c $f(x) = \begin{cases} kx(x-2), & 1 \leqslant x \leqslant 3, \\ 0, & \text{otherwise.} \end{cases}$

This is not a pdf

c Sketched below are the graphs for when $k > 0$ and $k < 0$.



graph if $k > 0$



graph if $k < 0$

So for any value of k there is some value of x in the given range such that $f(x) < 0$.

Therefore $f(x)$ cannot be a probability density function.

Example 2

The random variable X has probability density function:

$$f(x) = \begin{cases} k, & 1 < x < 2, \\ k(x-1), & 2 \leqslant x \leqslant 4, \\ 0, & \text{otherwise.} \end{cases}$$

a) Sketch $f(x)$.

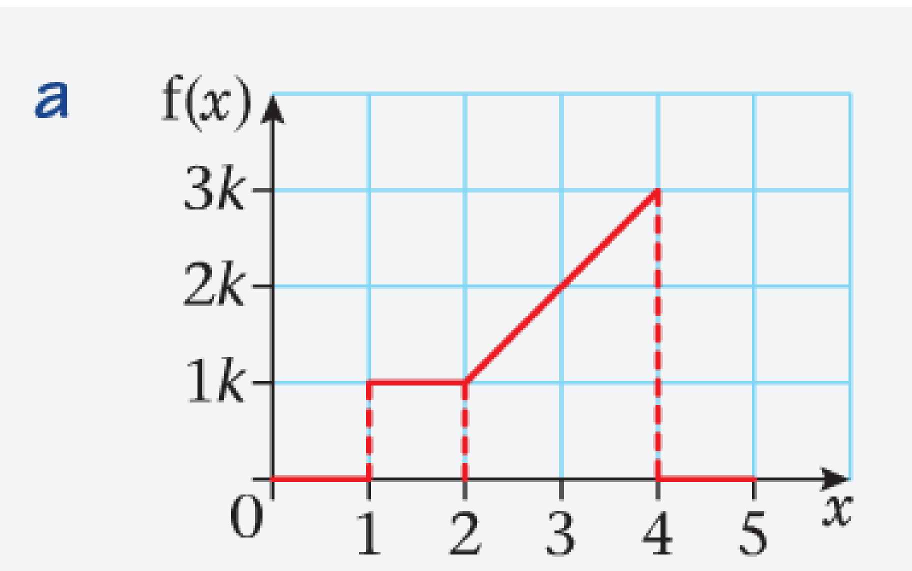
b) Find the value of k.

Solution

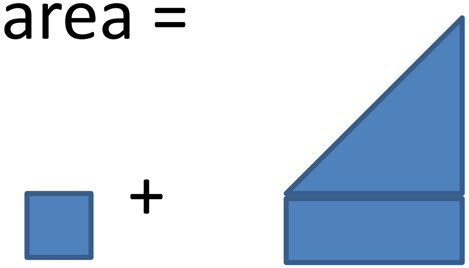
$$f(x) = \begin{cases} k, & 1 < x < 2, \\ k(x-1), & 2 \leq x \leq 4, \\ 0, & \text{otherwise.} \end{cases}$$

Notice that $f(2) = k$ either using the first one or the second one.

This is because a pdf is a continuous function.



Total area =



$= k + \frac{1}{2} \times 2 \times 4k = 5k$

So, $k = \frac{1}{5}$

Alternative solution: You can also use definite integrals to find the total area

$$f(x) = \begin{cases} k, & 1 < x < 2, \\ k(x-1), & 2 \leqslant x \leqslant 4, \\ 0, & \text{otherwise.} \end{cases}$$

b

$$\int_1^2 k\,dx + \int_2^4 k(x-1)\,dx = 1$$

$$[kx]_1^2 + \left[\frac{kx^2}{2} - kx\right]_2^4 = 1$$

$$k + [(8k - 4k) - (2k - 2k)] = 1$$

$$5k = 1$$

$$k = \frac{1}{5}$$

Your turn!

The random variable X has probability density function:

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

Find the value of k and sketch the pdf.

Solution

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

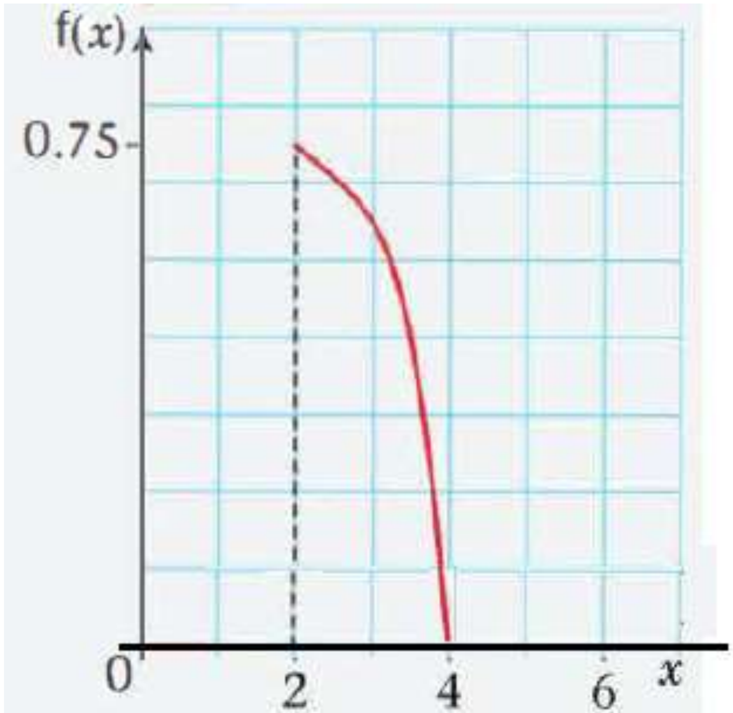
$$\int_2^4 k(4x - x^2)dx = 1$$

$$k\left[2x^2 - \frac{x^3}{3}\right]_2^4 = 1$$

$$k\left[\left(32 - \frac{64}{3}\right) - \left(8 - \frac{8}{3}\right)\right] = 1$$

$$k\left(\frac{16}{3}\right) = 1$$

$$k = \left(\frac{3}{16}\right)$$



Example 3

The random variable X has the probability density function:

$$f(x) = \begin{cases} {}^{3x^2}/_{8}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

Find $P(0.5 \leq X \leq 1.2)$ to 2 decimal places.

Note: Now that X is continuous, it can take any value.

Solution

Remember from earlier that:

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

So, we need to integrate $f(x)$ between 0.5 and 1.2:

$$P(0.5 \leq X \leq 1.2) = \int_{0.5}^{1.2} \left(\frac{3x^2}{8}\right) dx = \left[\frac{x^3}{8}\right]_{0.5}^{1.2}$$

$$= \left(\frac{1.2^3}{8}\right) - \left(\frac{0.5^3}{8}\right) = 0.200375 = 0.20 \text{ (2 d.p.)}$$

Do you notice anything about the inequality signs?

- Our formula used $P(a < X < b)$ (less than signs),

but the question was to find $P(0.5 \le X \le 1.2)$ (less than or equal to).

- Why was this not a problem?

  Can we really do this?

- To answer this, consider how to find $P(X = 0.5)$:

How to find P(X = 0.5)?

P(X = 0.5)=0

- We would try and find the area between the curve and the x axis at the particular value of X = 0.5.
- In other words, the area of a straight vertical line.
- Since in geometry, a line has no width and therefore no area, then the area of a straight line is zero.

- Alternatively, consider the limits of the definite integration $\lim_{a \to 0} \int_{0.5-a}^{0.5+a} f(x)dx;$ the upper and lower limits would both be 0.5, and therefore the value would be zero.

- If X is a continuous random variable then

$P(X = k) = 0$, for any real constant k

- This leads us to the conclusion that, if X is a continuous random variable then,

$P(X \le k) = P(X < k)$, for any real constant k

- There is no difference between $<$ and $\le$

(similarly, $>$ and $\ge$) for continuous random variables!

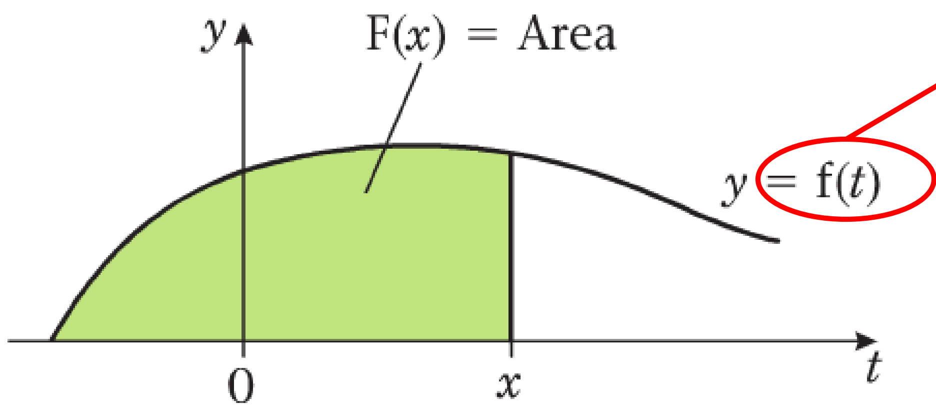
7.1.2 Define a cumulative distribution function

- The **cumulative distribution function (cdf)** for crv's is the same definition as for drv's:

$$\boxed{F(x) = P(X \leq x)}$$

- To calculate $F(x)$ we need to sum the probabilities from $X = -\infty$ to x.

- To sum the probabilities, we must find the area between the pdf and the x-axis from $-\infty$ up to a variable value denoted by x.

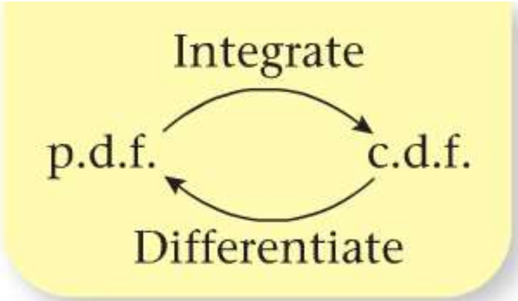


f is a function of t, not x, to avoid confusion.
We call t a dummy variable

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

Relationship between $F(x)$ and $f(x)$

- We can find $F(x)$ by integrating $f(x)$
- We can find $f(x)$ by differentiating $F(x)$.



- If X is a continuous random variable with cdf $F(x)$ and pdf $f(x)$,

$$\boxed{f(x) = \frac{d}{dx}F(x)} \quad \text{and} \quad \boxed{F(x) = \int_{-\infty}^{x} f(t)dt}$$

Example 4

The random variable X has probability density function:

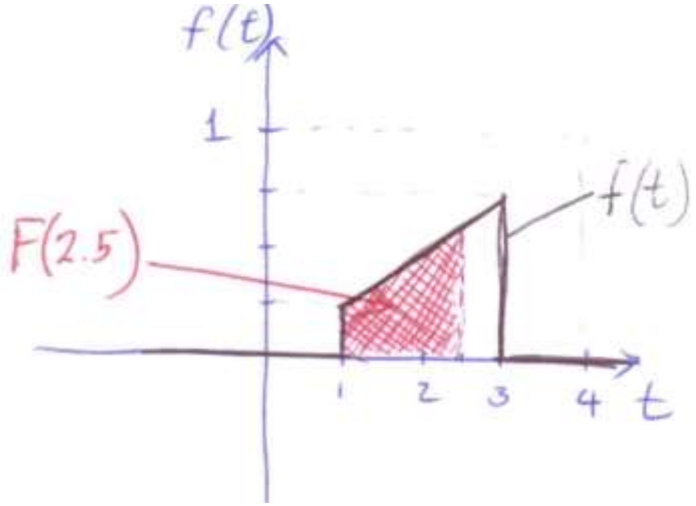
$$f(x) = \begin{cases} \frac{1}{4}x, & 1 \leqslant x \leqslant 3, \\ 0, & \text{otherwise.} \end{cases}$$

Find

a) F(2.5), b) F(x).

(Although the question doesn't say 'continuous random variable', this is implied by the terminology probability density function. Remember, drv's have probability functions.)

Solution



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

a) $F(2.5) = \int_{-\infty}^{2.5} f(t)dt = \int_{-\infty}^{1} f(t)dt + \int_{1}^{2.5} f(t)dt =$

$$0 + \int_{1}^{2.5} \frac{1}{4} t\, dt = \left[\frac{t^2}{8}\right]_1^{2.5} = \frac{1}{8}\left(2.5^2 - 1\right) = \frac{21}{32}$$

Solution

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

b) To find F(x) we need to use the definition:

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

First, find $F(x)$ for $x < 1$ and $x > 3$, outside of the given range.

For $x < 1$,

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} 0\,dt = 0$$

For $x > 3$,

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{3} f(t)dt + \int_{3}^{x} 0dt = 1$$

b) For $1 \leq x \leq 3$,

Method 1

$$F(x) = \int_{1}^{x} \frac{1}{4} t\,dt$$

$$= \left[\frac{t^2}{8}\right]_1^x$$

$$= \frac{x^2}{8} - \frac{1}{8}$$

Method 2

$$F(x) = \int \frac{1}{4} x\,dx$$

$$= \frac{x^2}{8} + C$$

Using the fact that $F(3) = 1$,

$$\frac{9}{8} + C = 1, \qquad C = -\frac{1}{8}$$

So, the cdf function $F(x)$ is

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

Caution!

Don’t forget to define $F(x)$ over the whole range $(-\infty,\infty)$

Example 6

The random variable X has probability density function:

$$f(x) = \begin{cases} \frac{1}{5}, & 1 < x < 2, \\ \frac{1}{5}(x - 1), & 2 \leqslant x \leqslant 4, \\ 0, & \text{otherwise.} \end{cases}$$

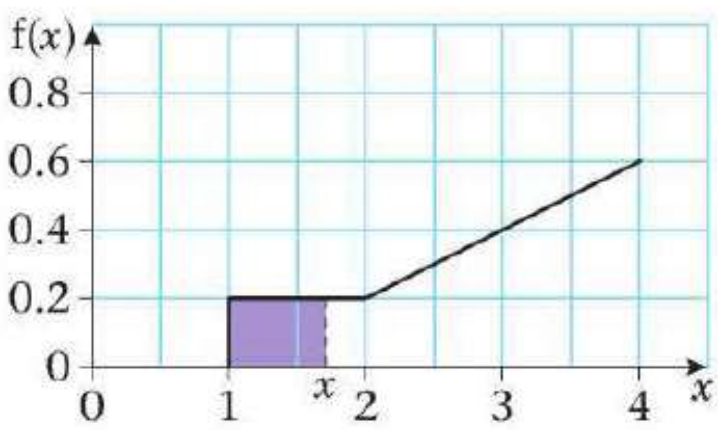
Find F(x).

This time f(x) has two parts defined and so we need to consider the two parts separately. We can use either method 1 or 2 as above; it is personal preference.

Solution

From the range given in pdf, we know that $F(x) = 0$ for $x \leq 1$ and $F(x) = 1$ for $x > 4$.

Let's consider the range $1 < x < 2$

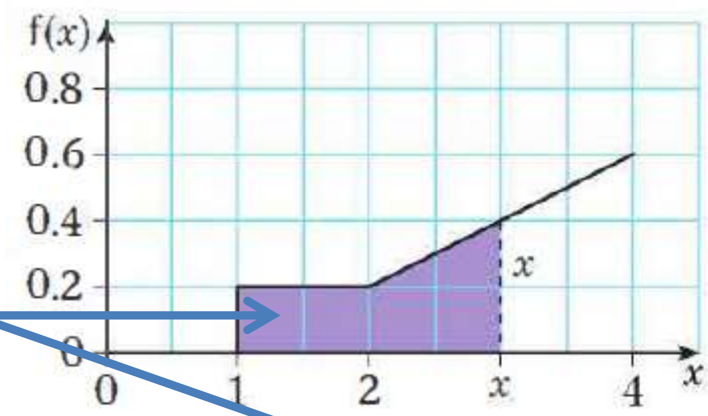


$$F(x) = \int_{-\infty}^{1} f(t)dt + \int_{1}^{x} f(t)dt$$

$$= F(1) + \int_{1}^{x} 0.2dt$$

$$= 0 + 0.2(x-1) = 0.2(x-1)$$

For $2 \leq x \leq 4$,



$0.2(x-1)$

$$F(x) = \int_{-\infty}^{2} f(t)dt + \int_{2}^{x} f(t)dt = F(2) + \int_{2}^{x} f(t)dt$$

$$= 0.2(2-1) + \int_{2}^{x} \tfrac{1}{5}(t-1)dt = 0.2 + \left[\frac{t^2}{10} - \frac{t}{5}\right]_2^x$$

$$= 0.2 + \frac{x^2}{10} - \frac{x}{5} - (0.4 - 0.4) = \frac{x^2}{10} - \frac{x}{5} + \frac{1}{5}$$

Again, we must write out $F(x)$ in full across the whole range $(-\infty, \infty)$:

$$F(x) = \begin{cases} 0, & x \leqslant 1, \\ \frac{1}{5}x - \frac{1}{5}, & 1 < x < 2, \\ \frac{x^2}{10} - \frac{x}{5} + \frac{1}{5} & 2 \leqslant x \leqslant 4, \\ 1. & x > 4. \end{cases}$$

Example 7

The random variable X has cumulative distribution function:

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{5}x + \frac{3}{20}x^2, & 0 \leqslant x \leqslant 2, \\ 1, & x > 2 \end{cases}$$

a) Find $P(X \leq 1.5)$.

b) Find $P(0.5 \leq X \leq 1.5)$.

c) Find $P(X = 1)$.

d) Find the probability density function, $f(x)$.

Solution

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{5}x + \frac{3}{20}x^2, & 0 \leqslant x \leqslant 2, \\ 1, & x > 2 \end{cases}$$

a) $P(X \leq 1.5) = F(1.5) = \frac{1}{5} \times 1.5 + \frac{3}{20} \times 1.5^2 = 0.6375$

b) $P(0.5 \leq X \leq 1.5) = F(1.5) - F(0.5)$

$$= 0.6375 - \left(\frac{1}{5} \times 0.5 + \frac{3}{20} \times 0.5^2\right) = 0.5$$

c) $P(X = 1) = 0$

d) $\frac{d}{dx}F(x) = \frac{1}{5} + \frac{3}{10}x, \quad f(x) = \begin{cases} \frac{1}{5} + \frac{3}{10}x & ,0 \leq x \leq 2 \\ 0 & \text{, othewise} \end{cases}$

Learning outcomes:

- 7.1.1 Understand the concept of a continuous random variable and its probability density function

- 7.1.2 Define a cumulative distribution function

Preview activity: 7.2 CRV 2

- Recall from 6.1:

The expected value of a DRV is defined as

$$E(X) = \sum xp(x)$$

The variance of a DRV is defined as

$$Var(X) = \sum x^2 p(x) - \mu^2$$

- Knowing now the concept of CRV and $f(x)$ (probability density function)

Derive the formulas for mean and variance of X a continuous random variable with p.d.f $f(x)$.