3.4 The Probability Density Function

In contrast to discrete random variables we can no longer look at the distribution function $f_X(i) = \mathbb{P}(X = i)$. This follows from the previous discussion where we mentioned that $\mathbb{P}(X = i) = 0$ for every continuous random variable. Instead we study the probability density function (pdf).

Definition 3.4.1. Let X be a real-valued continuous random variable, then we define the probability density function $f_X : \mathbb{R} \to \mathbb{R}^+$ as follows, for any $a, b \in \mathbb{R}$:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)\mathrm{d}x.$$

The main idea from the above definition is that if we want to find probability that a random variable X lies within some interval $[a, b]$, then all we have to do is find the area under the probability density function from $x = a$ to $x = b$. When we looked at discrete variables the probability was given by the height of the function, for continuous variables the probability is given by the area. This leads to an immediate consequence that the area under the entire curve must equal one. I.e for any real-valued continuous random variable:

$$\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1.$$

Example 3.4.1. Suppose X is a continuous random variable, with probability density function given as follows:

$$f_X(x) = \begin{cases} 1 - x & \text{for } x \in [0, 1]; \\ x - 1 & \text{for } x \in [1, 2]; \\ 0 & \text{otherwise.} \end{cases}$$

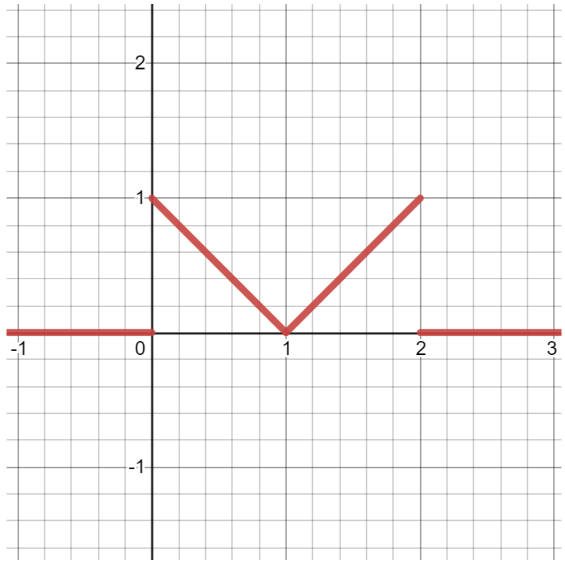
(i) Sketch the pdf of X.

(ii) Is $f_X(x)$ a probability distribution function?

(iii) Find $\mathbb{P}\left(X \leq \frac{1}{2}\right)$.

(iv) Find $\mathbb{P}\left(\frac{1}{2} \leq X \leq 2\right)$.

For part (i) we are interested in sketching the function $f_X(x)$ for all $x \in \mathbb{R}$. Firstly we note that for $x < 0$, and $x > 2$ we have that $f_X(x) = 0$. For the reaming cases we have two parts. For $x \in [0, 1]$ we have that $f_X(x) = 1 - x$, hence this is a line of gradient negative one and y-intercept one. While for $x \in [1, 2]$ we have a a line of gradient one, which intersects the x-axis at one. Hence we sketch it as follows:



For part (ii), we note that $f_X(x)$ is clearly non-negative for all $x \in \mathbb{R}$ hence we only need to show that the area under this curve is equal to one. Again for $x \in (-\infty, 0)$ and $(2, \infty)$ the area under the curve is 0. Hence we must show that,

$$\int_0^2 f_X(x)\ dx = 1.$$

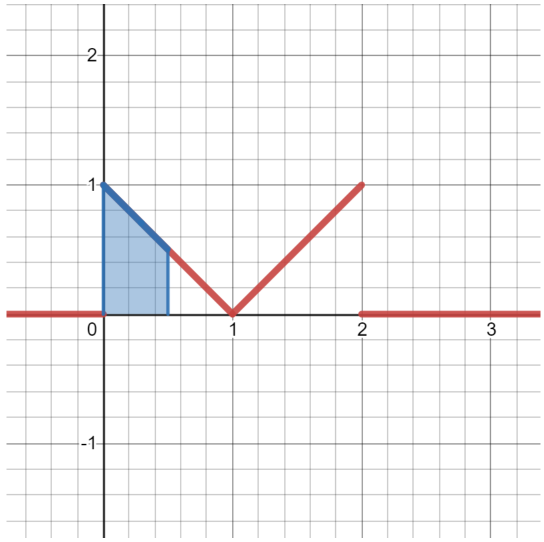
Now we note this region is made up of two identical triangles, both having a height of one, and a base of one. Hence we have that:

$$\int_0^2 f_X(x) = \frac{1}{2} \times 1 \times 1 + \frac{1}{2} \times 1 \times 1 = 1.$$

Therefore $f_X(x)$ is a distribution function. For part (iii) we need to calculate the following:

$$\mathbb{P}\left(X \leq \frac{1}{2}\right) = f_X\left(\frac{1}{2}\right) = \int_{-\infty}^{1/2} f_X(x)\ dx.$$

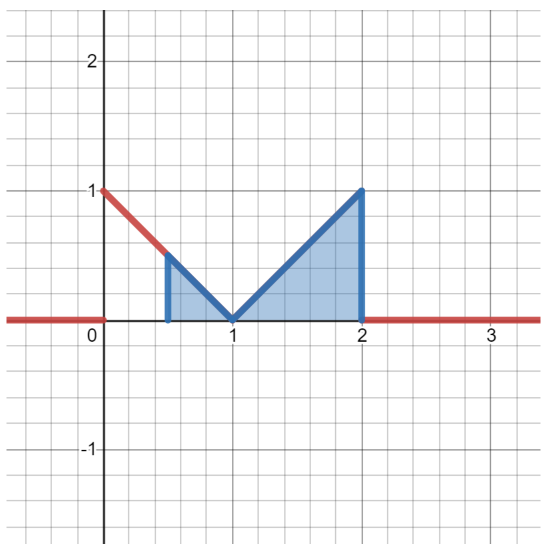
Again as $f_X(x) = 0$ for all $x < 0$ we only need to consider the area between 0 and 1/2. Again we can shade this region,



Thus this region is a trapezium, with parallel sides of length 1 and 0.5. While it has a base of length 0.5. Therefore it follows that:

$$\mathbb{P}\left(X \leq \frac{1}{2}\right) = \frac{1}{2} \times 0.5 \times (1 + 0.5) = 0.375.$$

For part (iv) we proceed in a similar fashion, again we shade the relevant region:



This region is made of precisely two triangles, thus is can be briefly checked that:

$$\mathbb{P}\left(\frac{1}{2} \leq X \leq 2\right) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} = 0.625.$$

Alternatively, you may observe that $\mathbb{P}(X \leq 0.5) = 1 - \mathbb{P}(0.5 \leq X \leq 2)$, this can be seen by comparing the two figures above.