



4.1.1 Probability Density Function (PDF)

To determine the distribution of a discrete random variable we can either provide its PMF or CDF. For continuous random variables, the CDF is well-defined so we can provide the CDF. However, the PMF does not work for continuous random variables, because for a continuous random variable $P(X = x) = 0$ for all $x \in \mathbb{R}$. Instead, we can usually define the **probability density function (PDF)**. The PDF is the **density** of probability rather than the probability mass. The concept is very similar to mass density in physics: its unit is probability per unit length. To get a feeling for PDF, consider a continuous random variable X and define the function $f_X(x)$ as follows (wherever the limit exists):

$$f_X(x) = \lim_{\Delta \rightarrow 0^+} \frac{P(x < X \leq x + \Delta)}{\Delta}.$$

The function $f_X(x)$ gives us the probability density at point x . It is the limit of the probability of the interval $(x, x + \Delta]$ divided by the length of the interval as the length of the interval goes to 0. Remember that

$$P(x < X \leq x + \Delta) = F_X(x + \Delta) - F_X(x).$$

So, we conclude that

$$f_X(x) = \lim_{\Delta \rightarrow 0} \frac{F_X(x + \Delta) - F_X(x)}{\Delta}$$

$$= \frac{dF_X(x)}{dx} = F'_X(x), \quad \text{if } F_X(x) \text{ is differentiable at } x.$$

Thus, we have the following definition for the PDF of continuous random variables:

Definition 4.2

Consider a continuous random variable X with an absolutely continuous CDF $F_X(x)$. The function $f_X(x)$ defined by

$$f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x), \quad \text{if } F_X(x) \text{ is differentiable at } x$$

is called the probability density function (PDF) of X .

Let us find the PDF of the uniform random variable X discussed in [Example 4.1](#). This random variable is said to have $Uniform(a, b)$ distribution. The CDF of X is given in Equation 4.1. By taking the derivative, we obtain

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x < a \text{ or } x > b \end{cases}$$

Note that the CDF is not differentiable at points a and b . Nevertheless, as we will discuss later on, this is not important. Figure 4.2 shows the PDF of X . As we see, the value of the PDF is constant in the interval from a to b . That is why we say X is uniformly distributed over $[a, b]$.

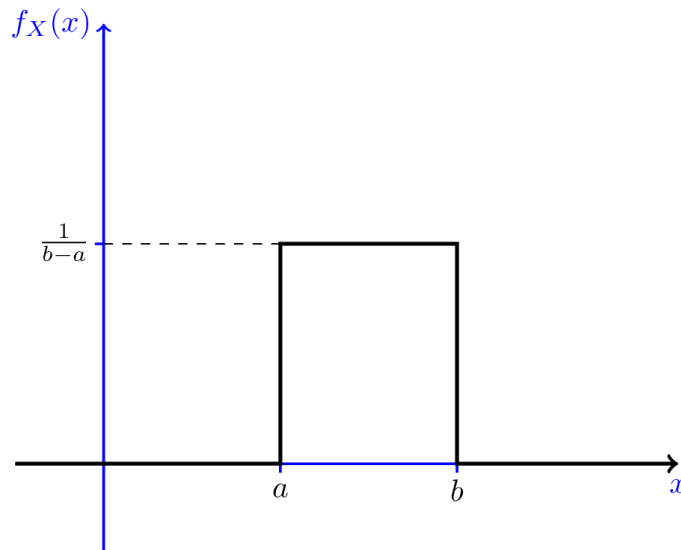


Fig.4.2 - PDF for a continuous random variable uniformly distributed over $[a, b]$.

The uniform distribution is the simplest continuous random variable you can imagine. For other types of continuous random variables the PDF is non-uniform. Note that for small values of δ we can write

$$P(x < X \leq x + \delta) \approx f_X(x)\delta.$$

Thus, if $f_X(x_1) > f_X(x_2)$, we can say $P(x_1 < X \leq x_1 + \delta) > P(x_2 < X \leq x_2 + \delta)$, i.e., the value of X is more likely to be around x_1 than x_2 . This is another way of interpreting the PDF.

Since the PDF is the derivative of the CDF, the CDF can be obtained from PDF by integration (assuming absolute continuity):

$$F_X(x) = \int_{-\infty}^x f_X(u)du.$$

Also, we have

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(u)du.$$

In particular, if we integrate over the entire real line, we must get 1, i.e.,

$$\int_{-\infty}^{\infty} f_X(u)du = 1.$$

That is, the area under the PDF curve must be equal to one. We can see that this holds for the uniform distribution since the area under the curve in Figure 4.2 is one. Note that $f_X(x)$ is density of probability, so it must be larger than or equal to zero, but it can be larger than 1. Let us summarize the properties of the PDF.

Consider a continuous random variable X with PDF $f_X(x)$. We have

1. $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.
2. $\int_{-\infty}^{\infty} f_X(u)du = 1$.
3. $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(u)du$.
4. More generally, for a set A , $P(X \in A) = \int_A f_X(u)du$.

In the last item above, the set A must satisfy some mild conditions which are almost always satisfied in practice. An example of set A could be a union of some disjoint

intervals. For example, if you want to find $P(X \in [0, 1] \cup [3, 4])$, you can write

$$P(X \in [0, 1] \cup [3, 4]) = \int_0^1 f_X(u) du + \int_3^4 f_X(u) du.$$

Let us look at an example to practice the above concepts.

Example 4.2

Let X be a continuous random variable with the following PDF

$$f_X(x) = \begin{cases} ce^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where c is a positive constant.

- Find c .
- Find the CDF of X , $F_X(x)$.
- Find $P(1 < X < 3)$.

Solution

- a. To find c , we can use Property 2 above, in particular

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_X(u) du \\ &= \int_0^{\infty} ce^{-u} du \\ &= c \left[-e^{-u} \right]_0^{\infty} \\ &= c. \end{aligned}$$

Thus, we must have $c = 1$.

- b. To find the CDF of X , we use $F_X(x) = \int_{-\infty}^x f_X(u) du$, so for $x < 0$, we obtain $F_X(x) = 0$. For $x \geq 0$, we have

$$F_X(x) = \int_0^x e^{-u} du = 1 - e^{-x}.$$

Thus,

$$F_X(x) = \begin{cases} 1 - e^{-x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

c. We can find $P(1 < X < 3)$ using either the CDF or the PDF. If we use the CDF, we have

$$P(1 < X < 3) = F_X(3) - F_X(1) = [1 - e^{-3}] - [1 - e^{-1}] = e^{-1} - e^{-3}.$$

Equivalently, we can use the PDF. We have

$$\begin{aligned} P(1 < X < 3) &= \int_1^3 f_X(t) dt \\ &= \int_1^3 e^{-t} dt \\ &= e^{-1} - e^{-3}. \end{aligned}$$

Range

The range of a random variable X is the set of possible values of the random variable. If X is a continuous random variable, we can define the range of X as the set of real numbers x for which the PDF is larger than zero, i.e.,

$$R_X = \{x | f_X(x) > 0\}.$$

The set R_X defined here might not exactly show all possible values of X , but the difference is practically unimportant.
