LECTURE 8
CONTINUOUS RANDOM VARIABLES AND PROBABILITY DENSITY FUNCTIONS (PDFs), UNIFORM DISTRIBUTION

The theory of discrete distributions and random variables described in previous lectures is very useful in many situations. However, there are also many situations where it cannot be applied, because it is more appropriate to model variables of interest as continuous, i.e. taking values in $\mathbb{R}$, $\mathbb{R}_+$, or some intervals in $\mathbb{R}$. For example, in economics we typically view prices, quantities, wages, and incomes as continuous.

We previously defined continuous random variables as those with continuous CDFs. Here, we will strengthen this requirement further and assume that the CDF is also *differentiable*.[1]

Assumption 1. *We say that a random variable X is continuously distributed if its CDF F_X is continuous and differentiable.*

Let $f_X(x)$ denote the derivative of F_X at x:

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

According to the Fundamental Theorem of Calculus, for $a \leq b$

$$F_X(b) - F_X(a) = \int_a^b f_X(x)dx.$$

On the other hand,

$$\begin{aligned} F_X(b) - F_X(a) &= P(X \leq b) - P(X \leq a) \\ &= P(a < X \leq b) \\ &= P(a \leq X \leq b) \\ &= P(a < X < b), \end{aligned} \tag{1}$$

where the last two equalities follow from the fact that for a continuously distributed random variable,

$$P(X = x) = 0 \text{ for all } x \in \mathbb{R}.$$

Thus, we can write

$$P(a < X < b) = \int_a^b f_X(x)dx. \tag{2}$$

We can compare the expression for $P(a < X < b)$ in (2) with $P(a < Y < b)$ for a discrete

[1] In mathematics, such functions are called absolutely continuous.

random variable Y. Let S_Y and p_Y be its support and PMF respectively. Then,

$$P(a < Y < b) = \sum_{y \in S_Y : a < y < b} p_Y(y). \tag{3}$$

The comparison of (2) and (3) shows that in the case of continuous random variables, we have to use integration over $[a, b]$ instead of summation as in the discrete case. This is because there is a continuum of values that X can take. However, the comparison also shows that in the continuous case, f_X plays a role similar to that of PMF p_Y in the discrete case. At the same time, it is important to remember that $f_X(x)$ is not the probability of $X = x$ as the latter is zero everywhere. The function f_X is called the probability density function (PDF).

Definition 2. (PDF) The probability density function of a continuous random variable with differentiable CDF F_X is defined as the derivative of F_X:

$$f_X(x) = \left. \frac{dF_X(u)}{du} \right|_{u=x}.$$

The PDF gives as an alternative way to describe continuous distributions. Note that one can always recover a CDF from its PDF. Using (1) and (2), sending a to $-\infty$, and using the fact that $\lim_{a \to -\infty} F_X(a) = 0$, we obtain that

$$\begin{aligned} F_X(b) &= P(X < b) \\ &= \int_{-\infty}^{b} f_X(x)dx. \end{aligned}$$

Example. (Uniform distribution) Let $a < b$, and define a function

$$F_X(x) = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x < b, \\ 1, & x \geq b. \end{cases}$$

The function F_X is obviously a CDF as it has all the required properties: bounded between zero and one, non-decreasing, and etc. It is also continuous everywhere, see Figure 1.

Figure 1: The CDF of the $Uniform(a, b)$ distribution: $F_X(x) = \frac{x-a}{b-a} \cdot 1(a < x < b) + 1(x \geq b)$

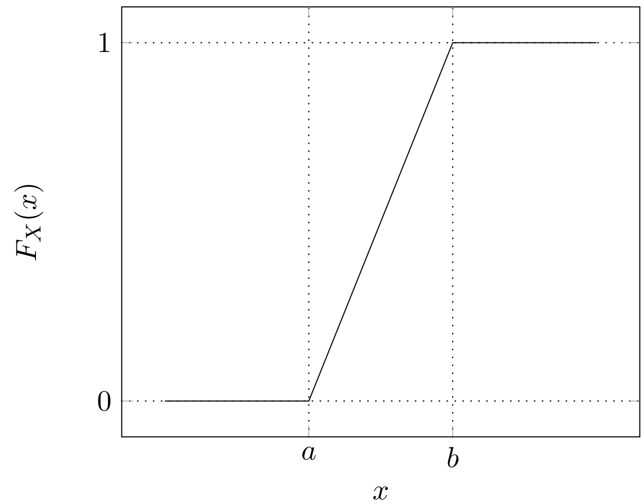
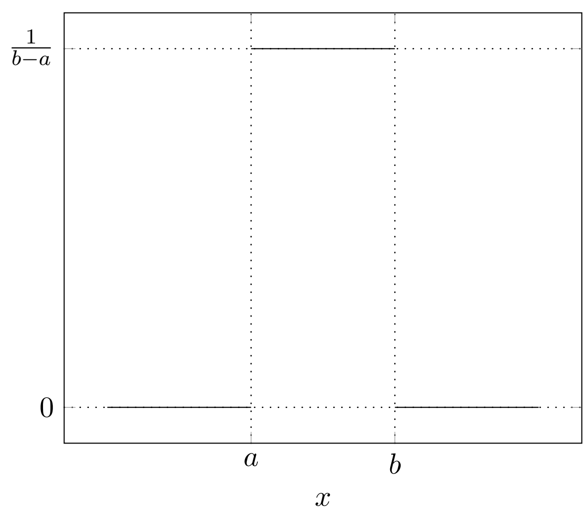


Figure 2: The PDF of the $Uniform(a, b)$ distribution: $f_X(x) = \frac{1}{b-a} \cdot 1(a < x < b)$



This CDF function is differentiable everywhere except for $x = a, b$. For $x < a$ or $x > b$, the derivative is zero, and it is given by

$$f_X(x) = \frac{1}{b-a} \text{ for } x \in (a, b).$$

However, since the integral over a point (integral with the lower bound equal to the upper bound) is zero, it does not matter how we define the derivative of F_X at a and b. Hence, we can define the PDF as

$$f_X(x) = \begin{cases} 0, & x \leq a, \\ \frac{1}{b-a}, & a < x < b, \\ 0, & x \geq b. \end{cases}$$

The graph of the PDF can be seen in Figure 2. Since the PDF is zero for $x \leq a$ and $x \geq b$, the probability of drawing X in those regions is zero. For example,

$$P(X \leq a) = \int_{-\infty}^{a} f_X(x)dx = \int_{-\infty}^{a} 0 \cdot dx = 0.$$

Thus, the support of the distribution of X is the interval (a, b). Note that the PDF is constant over the support, i.e. the probability mass is uniformly distributed over the support, which explains the name of the distribution. Let $a \leq x_1 < x_2 \leq b$. Then,

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} \frac{dx}{b-a} = \frac{x_2 - x_1}{b-a}.$$

We can see that the probability of $X \in (x_1, x_2)$ depends only on the length of the interval.

Next, we will discuss the properties of PDFs.

Theorem 3. *A PDF function f must satisfy the following properties.*
(a) $f(x) \geq 0$ for all $x \in \mathbb{R}$.
(b) $\int_{-\infty}^{\infty} f(x)dx = 1$.

Proof. Let F be the CDF corresponding to the PDF f. Part (a) of the theorem follows from the fact that F is non-decreasing:

$$f(x) = \frac{dF(x)}{dx} \geq 0.$$

Part (b) also follows from the properties of the CDF. Recall that

$$\begin{aligned} \lim_{b \to \infty} F(b) &= 1, \\ \lim_{a \to -\infty} F(a) &= 0. \end{aligned}$$

Next,

$$\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \lim_{a\to-\infty, b\to\infty} \int_a^b f(x)dx \\
&= \lim_{a\to-\infty, b\to\infty} (F(b) - F(a)) \\
&= \lim_{b\to\infty} F(b) - \lim_{a\to-\infty} F(a) \\
&= 1 - 0 \\
&= 1.
\end{aligned}$$

□

Since the PDF is always non-negative, and because a random variable cannot take values from the regions where the PDF is exactly zero, we can define the support of the distribution of a continuous random variable as the set of points where the PDF is strictly positive.

Definition 4. Let X be a continuously distributed random variable with a PDF $f_X(x)$. The support of the distribution of X is $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$.

The PDF gives us a convenient way to generate continuous distributions. Let h be a non-negative function with a finite integral over $\mathbb{R}$:

$$c = \int_{-\infty}^{\infty} h(x)dx > 0.$$

Define

$$f(x) = h(x)/c.$$

The function $f(x)$ is non-negative and integrates to one. Hence, it is a PDF. The corresponding CDF is

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(u)du \\
&= \frac{1}{c}\int_{-\infty}^{x} h(u)du.
\end{aligned}$$

Example. Consider the following function:

$$h(x) = \begin{cases} 0, & x \le 0, \\ x, & 0 < x < 1, \\ 0, & x \ge 1. \end{cases}$$

We have

$$\begin{aligned}
\int_{-\infty}^{\infty} h(x)dx &= \int_{-\infty}^{0} h(x)dx + \int_{0}^{1} h(x)dx + \int_{1}^{\infty} h(x)dx \\
&= \int_{-\infty}^{0} 0 \cdot dx + \int_{0}^{1} xdx + \int_{1}^{\infty} 0 \cdot dx \\
&= \int_{0}^{1} xdx \\
&= \left.\frac{x^2}{2}\right|_{0}^{1} \\
&= \frac{1}{2}.
\end{aligned}$$

Hence, $c = 1/2$ and

$$f(x) = 2x \cdot 1(0 < x < 1)$$

is a PDF. For $0 < x < 1$, the CDF F is given by

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} 2u \cdot 1(0 < u < 1) \\
&= 2\int_{0}^{x} udu \\
&= x^2.
\end{aligned}$$

Also,

$$\begin{aligned}
F(x) &= 0 \text{ for } x \leq 0, \\
F(x) &= 1 \text{ for } x \geq 1.
\end{aligned}$$

The functions $f(x)$ and $F(x)$ are plotted in Figures 3 and 4 respectively.

Figure 3: The plot of PDF $f(x) = 2x \cdot 1(0 < x < 1)$

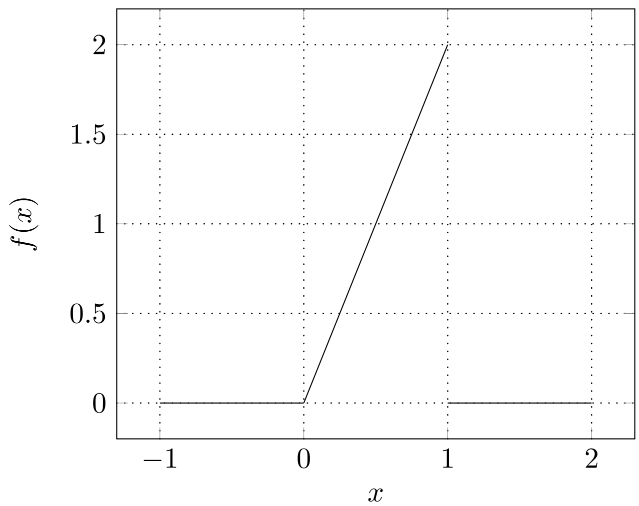


Figure 4: The plot of CDF $F(x) = x^2 \cdot 1(0 < x < 1) + 1(x \geq 1)$