Lecture 4

Outline

Cumulative distribution function

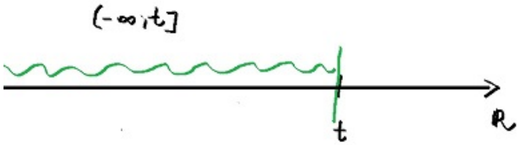
Can we describe the law of a random variable in a different way?

Definition

Let X be a random variable. The function $F_X : \mathbb{R} \to [0,1]$, defined by

$$F_X(t) := \mathbb{P}(X \le t) \quad \text{for} \quad t \in \mathbb{R},$$

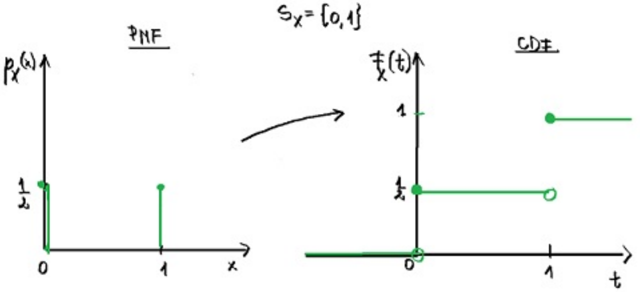
is called the **cumulative distribution function** of X.



Example

Toss a fair coin. Let X be a random variable defined by $X(\{H\}) = 1$ and $X(\{T\}) = 0$. Then

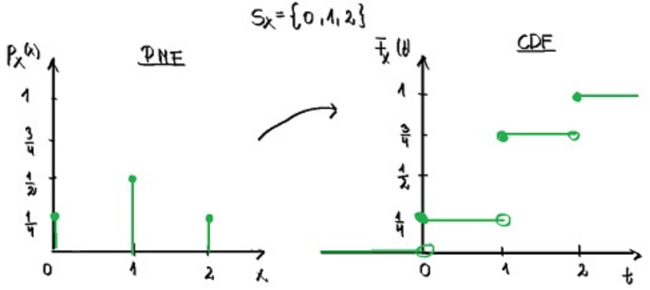
$$F_X(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{2}, & 0 \leq t < 1, \\ 1, & t \geq 1. \end{cases}$$



Example

Toss a fair coin twice. Let X be the number of heads.

$$F_X(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{4}, & 0 \le t < 1, \\ \frac{3}{4}, & 1 \le t < 2, \\ 1, & t \ge 2. \end{cases}$$

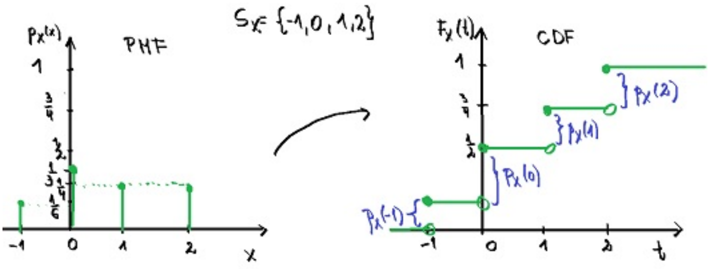


Example

Let X be a random variable with the following probability mass function:

x	-1	0	1	2
$p_X(x)$	1/6	1/3	1/4	1/4.

Find the corresponding cumulative distribution function F_X.



The above examples show that for discrete random variables we can easily switch from the PMF to the CDF and vice versa.

- CDF of a discrete random variable consists of jumps and flat regions,
- the height of a jump in the CDF at x is equal to the value of the PMF at x,
- the flat regions of the CDF correspond to the values outside the support of X (S_X), so the PMF is equal to 0 (in those regions)

Continuous distribution

Continuous random variables (on a contrary to discrete ones) can take on uncountable number of possible values, for example any real value in a given range like $(0, 1)$, $[0, \infty)$,$(-\infty, \infty)$.

How can we decribe probability distribution of a continuous random variable?

Definition (Continuous random variable)

A random variable X is called **continuous** if for some function $f : \mathbb{R} \to \mathbb{R}$ and for any $a, b \in \mathbb{R}$, $a \leq b$:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x)dx.$$

The function f need to satisfy:

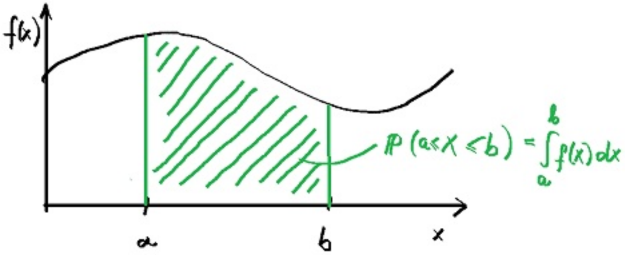
- $f(x) \geq 0, \; \forall x$,
- $\int_{-\infty}^{\infty} f(x)dx = 1$,

and is called **the probability density function** of X.

The **support** of X:

$$S_X = \{x \in \mathbb{R} : f(x) > 0\}.$$

The formula tells you that the probability of the random variable X falling inside the interval $[a, b]$ is **the area under the density curve**:



For any single value $a \in \mathbb{R}$:

$$\mathbb{P}(X = a) = \int_a^a f(x)dx = 0,$$

as a consequence, including or excluding endpoints of an interval doesn't alter its probability:

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b)$$

Cumulative distribution function

In particular, for a continuous random variable X with the probability density f_X, **the cumulative distribution function** (CDF) F_X is given by

$$F_X(t) = \int_{-\infty}^{t} f_X(x)dx, \ \forall t \in \mathbb{R}.$$

- $F_X(t)$ accumulates probability up to the value t,
- $$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a),$$
- $f_X(x)$ is not the probability of any particular event (it is not restricted to be ≤ 1)

Example (f_X can be arbitrarily large)

X - random variable with the following density:

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}}, & x \in (0,1), \\ 0, & x \notin (0,1). \end{cases}$$

Compute $\mathbb{P}\left(X \in (10^{-4}, 10^{-2})\right)$.

- Even though $f_X(x)$ becomes infinitely large as $x \to 0$, this is still the valid pdf:
$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 \frac{1}{2\sqrt{x}}dx = \sqrt{x}|_0^1 = 1,$$
and $f_X(x) \geq 0$, $\forall x \in \mathbb{R}$.
- $$\mathbb{P}\left(X \in (10^{-4}, 10^{-2})\right) = \int_{10^{-4}}^{10^{-2}} \frac{1}{2\sqrt{x}}dx = \sqrt{x}|_{10^{-4}}^{10^{-2}} = \frac{9}{100}.$$

Example

Let X be a random variable taking on its values in the range $[1, 4]$ with the following density function:

$$f_X(x) = \begin{cases} c, & 1 \le x \le 4, \\ 0, & \text{otherwise.} \end{cases}$$

Compute c. Find CDF of X.

Solution

The density function need to satisfy the normalization property:

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_1^4 cdx = c(4-1) = 3c,$$

so $c = \frac{1}{3}$. *Having* f_X *fully described, we can determine CDF of* X*:*

$$F_X(t) = \mathbb{P}(X \le t) = \int_{-\infty}^{t} f_X(dx) = \begin{cases} 0, & t < 1, \\ \int_1^t \frac{1}{3}dx = \frac{t-1}{3}, & 1 \le t < 4, \\ 1, & t \ge 4. \end{cases}$$

Example

Let X be a random variable with the following density function:

$$f_X(x) = \begin{cases} \frac{1}{4}, & x \in (-1, 0), \\ \frac{1}{2}x, & x \in (1, 2), \\ 0, & \text{otherwise.} \end{cases}$$

Determine the cumulative distribution function of the random variable X.
Compute $\mathbb{P}\left(X \in \left(-\frac{1}{2}, \frac{3}{2}\right)\right)$.

Solution

- $$F_X(t) = \begin{cases} 0, & t < -1, \\ \int_{-1}^{t} \frac{1}{4} = \frac{1}{4}(t+1), & -1 \leq t < 0, \\ \frac{1}{4}, & 0 \leq t < 1, \\ \frac{1}{4} + \int_{1}^{t} \frac{1}{2} x dx = \frac{1}{4} + \frac{1}{4}(t^2 - 1), & 1 \leq t < 2, \\ 1, & t \geq 2, \end{cases}$$
- $\mathbb{P}\left(X \in \left(-\frac{1}{2}, \frac{3}{2}\right)\right) = \int_{-\frac{1}{2}}^{\frac{3}{2}} f_X(x) dx = \int_{-\frac{1}{2}}^{0} \frac{1}{4} dx + \int_{1}^{\frac{3}{2}} \frac{1}{2} x dx = \frac{7}{16}$.

We can also determine the pdf from the given cdf:

$$f_X(x) = F'_X(x),$$

for all x for which the derivative exists.

Example

Let X be a random variable with the cdf given by

$$F_X(t) = \begin{cases} 0, & t < 0, \\ 2t^3 - t^4, & 0 \le t < 1, \\ 1, & t \ge 1. \end{cases}$$

Determine the pdf of X.

Solution

F_X is continuous, therefore

$$f_X(x) = F'_X(x) = \begin{cases} 0, & x \notin (0,1), \\ 6x^2 - 4x^3, & x \in (0,1). \end{cases}$$

Properties of the cumulative distribution function

- F_X is nondecreasing: $x \leq y \implies F_X(x) \leq F_X(y)$,
- $\lim_{t \to -\infty} F_X(t) = 0$ and $\lim_{t \to \infty} F_X(t) = 1$,

X - discrete r.v.:

- F_X is a step function that increases only by jumps,
- $$F_X(t) = \sum_{x \in S_X : x \leq t} p_X(x)$$

X- continuous r.v.:

- F_X is a continuous function,
- $$F_X(t) = \int_{-\infty}^{t} f_X(x) dx$$

  and

  $$f_X(x) = F'_X(x),$$

  for all x for which $F'_X(x)$ exists.