Lecture 7: Properties of pmf, distribution functions, and continuous random variables with density

1 Characterising properties of the pmf

Let X be a discrete random variable with pmf $p_X(x)$. Recall that, for any $A \subseteq \mathbb{R}$, we have

$$\mathbb{P}(X \in A) = \sum_{x \in S_X \cap A} p_X(x). \tag{1.1}$$

There are two basic properties of the pmf.

Property 1. We have

$$p_X(x) \geqslant 0 \quad \text{for any } x \in S_X.$$

This is obvious since $p_X(x)$ is a probability.

Property 2. We have

$$\sum_{x \in S_X} p_X(x) = 1.$$

This follows directly from the more general formula (1.1), by taking $A = \mathbb{R}$. The fact that $\mathbb{P}(X \in \mathbb{R}) = 1$ is trivially since X is assumed to be real-valued.

Remark 1.1. Any function p_X defined on a given countable subset $S \subseteq \mathbb{R}$ is the pmf of a discrete random variable whose set of possible values is S. However, let us not bother with this theoretical point.

Example 1.1. Let X be a discrete random variable whose pmf is given by the following table.

x	1	2	3	4	5
$p_X(x)$	α	2α	3α	4α	5α

What is α and $\mathbb{P}(2 \leqslant X \leqslant 4)$?

Solution. According to Property 1, we must have

$$\alpha + 2\alpha + 3\alpha + 4\alpha + 5\alpha = 1.$$

This implies $\alpha = \frac{1}{15}$. According to the formula (1.1), we have

$$\mathbb{P}(2 \leqslant X \leqslant 4) = p_X(2) + p_X(3) + p_X(4) = \frac{9}{15}.$$

2 The cumulative distribution function and its basic properties

We have mentioned the "distribution" of a random variable only in vague terms (which is uniquely determined by the pmf in the case of discrete random variables). It is important to have a general mathematical notion that fully captures the information about distribution. This leads us to the concept of the cumulative distribution function.

Definition 2.1. Let $X : \Omega \to \mathbb{R}$ be a given random variable. The *cumulative distribution function* (*cdf*) of X is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = \mathbb{P}(X \leqslant x), \quad x \in \mathbb{R}.$$

Note that the cdf is defined for an arbitrary random variable (not necessarily discrete). The cdf is a function defined on the entire real line, while the pmf (in the case of discrete random variables) is a function defined on the set S_X of possible values. Of course we can also think of the pmf as a function defined on the entire real line by realising the simple fact that

$$p_X(x) = \mathbb{P}(X = x) = 0 \quad \text{if } x \notin S_X.$$

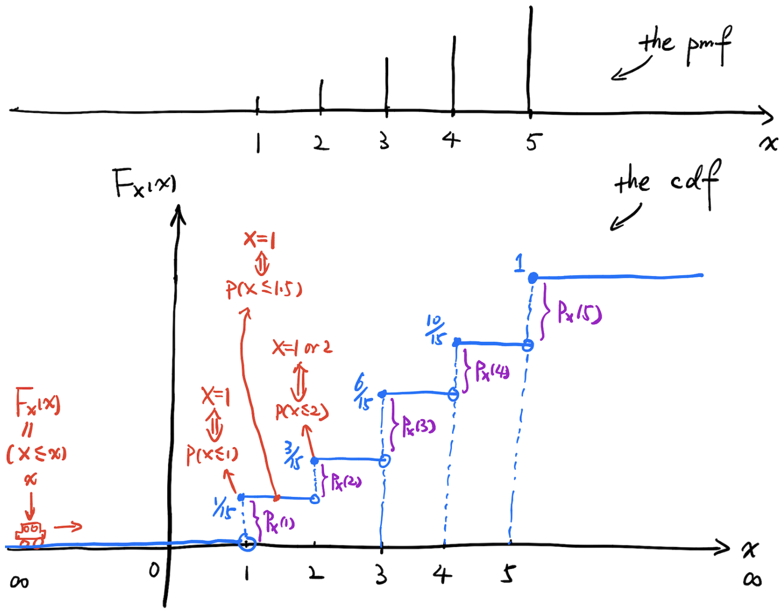


Figure 1: Sketching the cdf of a discrete random variable.

In the discrete case, the pmf uniquely determines the cdf as seen from

$$F_X(x) = \sum_{y \in S_X, y \leqslant x} p_X(y),$$

which is a direct consequence of the formula (1.1).

We use the following example to illustrate how we can write down the cdf precisely and sketch its graph in the case of discrete random variables.

Example 2.1. We consider the pmf given by Example 1.1 (with $\alpha = \frac{1}{15}$). Recall that, for each real number x, the value $F_X(x)$ is the probability $\mathbb{P}(X \leqslant x)$. The intuitive procedure of writing down the cdf can be described by the following fun dynamical procedure. In Figure 1 below, imagine that you are sitting in a car starting from the $-\infty$ position and starting to drive towards the positive direction. At every location x, you ask yourself what $\mathbb{P}(X \leqslant x)$ is, which gives you the value of $F_X(x)$ at the location. In the current example, at any location strictly before $x = 1$, we do not encounter any probability mass and thus

$$F_X(x) = \mathbb{P}(X \leqslant x) = 0.$$

This is true all the way up to (but strictly before) $x = 1$ no matter how close to 1 the car is. Now suppose that our car is right at the position $x = 1$. We immediately pick up a mass of $\frac{1}{15}$ since

$$F_X(1) = \mathbb{P}(X \leqslant 1) = \mathbb{P}(X = 1) = p_X(1) = \frac{1}{15}.$$

Therefore, as we passes through the point $x = 1$, the value of $F_X(x)$ jumps from 0 to $\frac{1}{15}$. After passing $x = 1$ let us keep driving towards the next threshold $x = 2$. As long as we have not reached $x = 2$, the graph of $F_X(x)$ stays flat, because there is no new mass to pick up between the points 1 and 2. Mathematically,

$$F_X(x) = \mathbb{P}(X \leqslant x) = p_X(1)$$

for all $1 \leqslant x < 2$. Next, in a similar way as before, when the car reaches the position $x = 2$, we pick up a new pass $p_X(2) = \frac{2}{15}$, and the value of F_X immediately jumps up to

$$F_X(2) = \mathbb{P}(X \leqslant 2) = p_X(1) + p_X(2) = \frac{1}{15} + \frac{2}{15} = \frac{3}{15}.$$

As we keep driving onwards, the graph remains flat until we reach the next threshold $x = 3$. By picking up the new mass $p_X(3)$ at the position $x = 3$, the function $F_X(x)$ jumps up to

$$F_X(3) = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} = \frac{6}{15}.$$

By continuing this procedure, we obtain a full description of the function $F_X(x)$ on the entire real line. Note that after passing the last threshold $x = 5$, we have picked up all the masses and thus the function $F_X(x)$ jumps up to 1 and stays there forever. Having this intuition in mind, we can easily write down the entire definition of F_X as follows:

$$F_X(x) = \begin{cases} 0, & x < 1; \\ 1/15, & 1 \leqslant x < 2; \\ 3/15, & 2 \leqslant x < 3; \\ 6/15, & 3 \leqslant x < 4; \\ 10/15, & 4 \leqslant x < 5; \\ 1, & x \geqslant 5, \end{cases}$$

and sketch its graph as shown in Figure 1.

The cdf of a random variable has the following basic properties. The intuition is mostly clear with the aid of Figure 1 in the previous example.

Property 1. We have

$$0 \leqslant F_X(x) \leqslant 1 \quad \text{for all } x \in \mathbb{R}.$$

This is apparent as $F_X(x)$ is a probability.

Property 2. For any $a < b$, we have

$$\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a).$$

Proof. Consider the following decomposition:

$$\{X \leqslant b\} = \{X \leqslant a\} \cup \{a < X \leqslant b\}.$$

The two events on the right hand side are clearly disjoint. Therefore,

$$F_X(b) = \mathbb{P}(X \leqslant b) = \mathbb{P}(X \leqslant a) + \mathbb{P}(a < X \leqslant b) = F_X(a) + \mathbb{P}(a < X \leqslant b).$$

$\square$

Property 3. $F_X(x)$ is an non-decreasing function. This is clear from Property 2 and the fact that $\mathbb{P}(a < X \leqslant b) \geqslant 0$.

Property 4. We have

$$F_X(-\infty) := \lim_{x \to -\infty} F_X(x) = 0, \ \ F_X(\infty) := \lim_{x \to +\infty} F_X(x) = 1.$$

Proof. We only verify the first part as the second part is shown in a similar way. The main idea is to construct a monotone sequence of events and use the continuity property of probability functions. For each $n \geqslant 1$, let us define $A_n = \{X \leqslant -n\}$. Then we have

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots .$$

In additon, we also have $\cap_{n=1}^{\infty} A_n = \emptyset$, as it is impossible to have $X(\omega) \leqslant -n$ for all n. According to the continuity property of probability functions, we have

$$0 = \mathbb{P}(\emptyset) = \lim_{n \to \infty} \mathbb{P}(A_n) = \lim_{n \to \infty} \mathbb{P}(X \leqslant -n) = \lim_{n \to -\infty} F_X(n).$$

$\square$

Property 5. $F_X(x)$ is a right continuous function, namely

$$\lim_{h\downarrow 0} F_X(x+h) = F_X(x) \quad \text{for every } x \in \mathbb{R}.$$

Proof. The idea is similar to the proof of Property 4. Let us show that

$$\lim_{n\to\infty} F_X\big(x+\frac{1}{n}\big) = F_X(x).$$

For this purpose, we define $A_n = \{X \leqslant x + 1/n\}$. Then we have

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$$

and

$$\cap_{n=1}^{\infty} A_n = \{X \leqslant x\}.$$

The latter property can be easily seen by the element-wise approach. According to the continuity property of probability functions, we have

$$F_X(x) = \mathbb{P}(X \leqslant x) = \lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} F_X\big(x+\frac{1}{n}\big).$$

$\square$

Note that a cdf may fail to be left continuous (see Example 2.1). Indeed, for each x, the possible jump size of F_X at x is precisely given by the probability mass $\mathbb{P}(X = x)$. This is the content of the next property.

Property 6. For every $x \in \mathbb{R}$, we have

$$F_X(x) - F_X(x-) = \mathbb{P}(X = x),$$

where

$$F_X(x-) = \lim_{h\downarrow 0} F_X(x-h)$$

is the left limit of F_X at x.

Proof. Let us define $A_n = \{x - 1/n < X \leqslant x\}$. Then

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$$

and

$$\cap_{n=1}^{\infty} A_n = \{X = x\}.$$

It follows that

$$
\begin{aligned}
\mathbb{P}(X=x) &= \lim_{n\to\infty}\mathbb{P}(A_n) = \lim_{n\to\infty}\mathbb{P}\big(x-\frac{1}{n}<X\leqslant x\big) \\
&= \lim_{n\to\infty}\big(F_X(x)-F_X\big(x-\frac{1}{n}\big)\big) \\
&= F_X(x)-F_X(x-).
\end{aligned}
$$

□

From Property 6, we see that F_X is continuous at x if and only if $\mathbb{P}(X=x)=0$. For a discrete random variable, F_X has a jump (with jump size $p_X(x)$) at every $x\in S_X$ that has a positive probability mass.

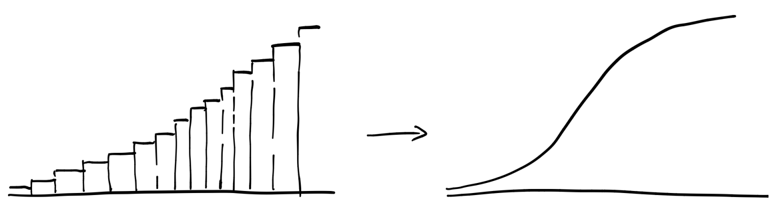
Remark 2.1. Any function $F:\mathbb{R}\to[0,1]$ that is non-decreasing, right continuous and satisfies

$$F(-\infty)=0,\ F(\infty)=1$$

is the cdf of a random variable. We do not prove this theoretical fact.

3 Continuous random variables with density

Recall that, the cdf for a discrete random variable looks like a step function, where the jumps occur precisely at the discrete values that X takes. It is not hard to imagine, as the set of possible values becomes more and more dense in $\mathbb{R}$, the cdf will look like a continuous curve:



Definition 3.1. A random variable $X:\Omega\to\mathbb{R}$ is said to be a *continuous random variable* if the cdf of X is continuous.

The behaviour of a continuous random variable is drastically different from the discrete case at least from the following two aspects.

First of all, since the cdf is continuous in this case, from Property 6 we know that $\mathbb{P}(X = x) = 0$ for every x. In other words, there is no probability masses assigned to any value x. As a result, for the study of continuous random variables, we have to rely on events like $\{a < X \leqslant b\}$ to extract meaningful information about the distribution.

Another notable difference is that, the set S_X of possible values must be uncountable in this case. Indeed, if S_X were countable, say $S_X = \{x_1, x_2, x_3, \cdots\}$. Then we would have

$$1 = \mathbb{P}(X \in S_X) = \sum_{n=1}^{\infty} \mathbb{P}(X = x_n) = \sum_{n=1}^{\infty} 0 = 0,$$

which is absurd. Therefore, S_X cannot be countable.

As a result of the above two points, the way of studying continuous random variables will also be very different from the discrete case. To motivate the suitable method, recall that in the discrete case we can compute $\mathbb{P}(a < X \leqslant b)$ via the formula

$$\mathbb{P}(a < X \leqslant b) = \sum_{x \in S_X,\ a < x \leqslant b} p_X(x). \tag{3.1}$$

It is reasonable to believe that, in the continuous case, an effective way of understanding a similar probability would be via integration. Namely, we should look for a notion of "density function" $f_X(x)$ that plays the analogous role of the pmf $p_X(x)$, and replace the summation in (3.1) by the integral $\int_a^b f_X(x)dx$. This idea leads us to the next fundamental type of random variables: continuous random variables with density. The study of this type of random variables relies on calculus in a crucial way.

Definition 3.2. Let X be a continuous random variable whose cdf is $F_X(x)$. We say that X *has a density*, if there is a function $f_X : \mathbb{R} \to [0, \infty)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy \quad \text{for all } x \in \mathbb{R}. \tag{3.2}$$

In this case, we say that $f_X(x)$ is the *probability density function* (*pdf*) of X.

Let X be a continuous random variable with pdf $f_X(x)$. Since $F_X(x)$ is the integral of $f_X(x)$, from calculus we know that

$$f_X(x) = \frac{dF_X(x)}{dx}. \tag{3.3}$$

As a result, we see that if a pdf exists it must be unique (given by the derivative of F_X).

Remark 3.1. It is a deep theoretical point that, in general the formula (3.3) is only true for "almost every" x rather than every x. Nonetheless, if $f_X(x)$ is continuous (it does not need to be in general), then it is true for all $x \in \mathbb{R}$.

Remark 3.2. It is an even deeper theoretical point that, not every continuous random variable admits a pdf. To put it in abstract terms, there exists a continuous cdf $F(x)$ whose derivative is zero "almost everywhere". In this case, the equation (3.2) can never hold and the corresponding random variable is continuous but without a density.