Lecture 8: Properties of pdf, expectation of discrete and continuous random variables, accounting trick

1 Properties of the pdf

The most important property of a pdf is that it can be used to calculate probabilities. More precisely,

$$\mathbb{P}(a < X \leqslant b) = \int_a^b f_X(x)dx \quad \text{for any } a < b.$$

This follows from:

$$\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a) = \int_{-\infty}^b f_X(x)dx - \int_{-\infty}^a f_X(x)dx = \int_a^b f_X(x)dx.$$

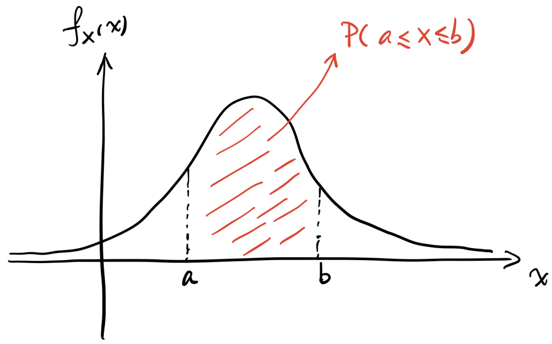
This integral is also equal to any of the following probabilities:

$$\mathbb{P}(a < X < b),\ \mathbb{P}(a \leqslant X < b),\ \mathbb{P}(a \leqslant X \leqslant b).$$

This is clear since

$$\mathbb{P}(X = a) = \mathbb{P}(X = b) = 0$$

for a continuous random variable X, so that whether we include the end points or not is not relevant. In concise words, the *area* beneath the graph of $f_X(x)$ over $[a, b]$ gives $\mathbb{P}(X \in [a, b])$:



Two other basic properties of a pdf are given below. They are analogous to Properties 1 and 2 for the pmf.

Property 1. $f_X(x) \geqslant 0$.

Proof. Since the cdf $F_X(x)$ is a non-decreasing function, we know that its derivative (which is the pdf $f_X(x)$) must be non-negative. $\square$

Property 2. We have

$$\int_{-\infty}^{\infty} f_X(x)dx = 1.$$

Proof. This follows from

$$1 = F_X(\infty) = \lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} \int_{-\infty}^{x} f_X(y)dy = \int_{-\infty}^{\infty} f_X(x)dx.$$

Note that $\int_{-\infty}^{\infty} f_X(x)dx$ is nothing but the probability $\mathbb{P}(-\infty < X < \infty)$ which is clearly equal to one. $\square$

Remark 1.1. Any function that satisfies Properties 1 and 2 is the pdf of some random variable.

The pdf for continuous random variables with density and the pmf for discrete random variables are playing similar roles for studying distributions. The following table gives a brief comparsion between the two objects:

Discrete	Continuous
pmf $p_X(x)$	*pdf* $f_X(x)$
prob. masses $p_X(x)$ at x	no positive masses $p_X(x) = 0 \quad \forall\, x$
$\sum_{x \in S_X} p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(t)dt = 1$
$\mathbb{P}(X \in I) = \sum_{x \in I} p_X(x)$	$\mathbb{P}(X \in I) = \int_a^b f_X(t)dt$
$0 \leq p_X(x) \leq 1$	$f_X(x) \geq 0$

Figure 1: Comparison between pmf and pdf (where $I = [a, b]$).

On the other hand, there is a notable difference between the two objects. Unlike the pmf, the pdf $f_X(x)$ is *not* a probability. The specific value of f_X at a particular point is not meanfingful–it is the *integration* of f_X (the area beneath the graph of f_X) over a region I (say $I = [a, b]$) that gives a meaningful probability (the probability that $X \in [a, b]$). The following simple example of a legal pdf shows that $f_X(x)$ need not satisfy $0 \leqslant f_X(x) \leqslant 1$.

Example 1.1. The following function is a pdf, namely it satisfies Properties 1 and 2:

$$f_X(x) = \begin{cases} 10^6, & 0 < x < 10^{-6}; \\ 0, & \text{otherwise.} \end{cases}$$

Interpretation of the pdf. There is still an infinitesimal interpretation of $f_X(x)$. Let x be fixed and let δ be a small number. Let us define the event

$$\{X \approx x\} := \big\{x - \frac{\delta}{2} < X \leqslant x + \frac{\delta}{2}\big\}.$$

Then we have

$$\mathbb{P}(X \approx x) = \int_{x-\delta/2}^{x+\delta/2} f_X(y)dy.$$

From calculus, we know that when δ is small, the above integral is approximately given by $f_X(x) \times \delta$. Therefore, we have

$$\mathbb{P}(X \approx x) \approx f_X(x) \times \delta \quad \text{if } \delta \text{ is small.}$$

Remark 1.2. There are random variables that are neither discrete nor continuous. The cdf given by the following figure provides one such example.

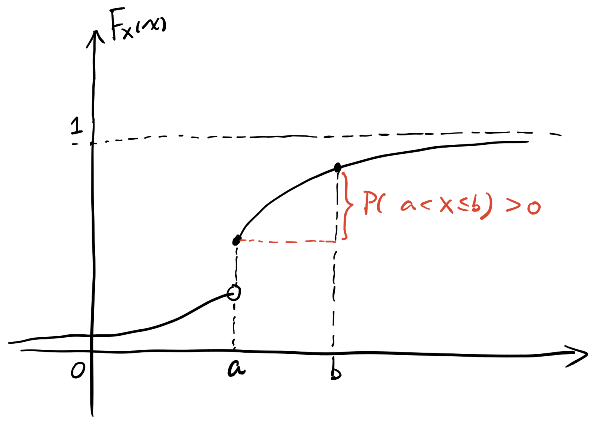


Figure 2: The cdf of a random variable that is neither discrete nor continuous.

Indeed, the underlying random variable cannot be continuous since the cdf is not continuous at $x = a$. It is also not a discrete random variable. Suppose on the contrary that it were, then X takes values in a countable set. In particular, $S_X \cap (a, b]$ is also a countable set, say

$$S_X \cap (a, b] = \{x_1, x_2, x_3, \cdots\}.$$

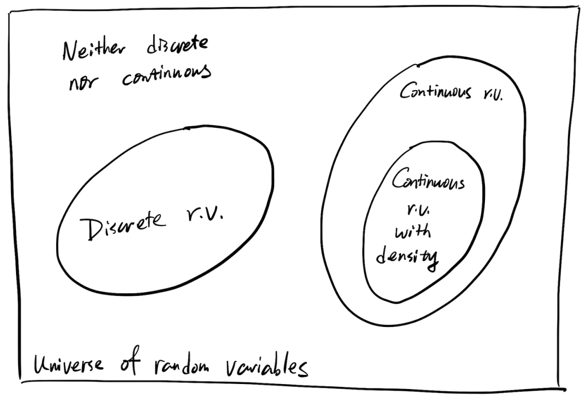
From Figure 2, we see that

$$\mathbb{P}(a < X \leqslant b) = F_X(b) - F_X(a) > 0.$$

On the other hand, since $F_X(x)$ is continuous on $(a, b]$, we know that $\mathbb{P}(X = x_n) = 0$ for all $x_n \in S_X \cap (a, b]$. According to the countable additivity property,

$$\mathbb{P}(a < X \leqslant b) = \sum_{n=1}^{\infty} \mathbb{P}(X = x_n) = 0,$$

which is a contradiction. Therefore, X is not discrete. The argument here is essentially the same as how we see that the set of possible values for a continuous random variable cannot be countable. The universe of random variables can be illustrate by the following picture.



2 The expectation of a random variable

The cdf of a random variable contains all information about the distribution, while sometimes we only need some effective summary / partial features of the distribution rather than the entire cdf. One important feature is the "average value" of a random variable, which leads to the notion of expectation.

We begin with a simple example to illustrate the idea of taking expectation.

Example 2.1. Consider a game where your net winning and the corresponding probabilities are given by the following table.

W	-1	1	10
$P_W(w)$	0.75	0.2	0.05

To understand what your "average winning" should be, suppose that you play the game for 1000 times. The given probabilities suggest that, around 750 times you will lose \$1, around 200 times you will win \$1, and around 50 times you will win \$10. Therefore, your "average winning" should be

$$
\begin{aligned}
E &= \frac{-1 \times 750 + 1 \times 200 + 10 \times 50}{1000} \\
&= (-1) \times 0.75 + 1 \times 0.2 + 10 \times 0.05 \\
&= -0.05.
\end{aligned}
$$

This "average value" gives an indication of whether the game is worth playing.

2.1 The expectation of a discrete random variable

The above simple example clear suggests a general way to define the "average value" of a random variable (at least in the discrete case). We now make it mathematically precise.

Definition 2.1. Let X be a discrete random variable whose set of possible values is S_X. The *expectation* (also known as the *expected value* or *the mean*) of X is defined by

$$\mathbb{E}[X] = \sum_{x \in S_X} x \cdot p_X(x), \tag{2.1}$$

provided that the sum on the right hand side is absolutely convergent.

Intuitively, the expectation is defined by averaging all the possible values of X by the corresponding probability masses. There is an important interpretation of the expectation. Suppose that we perform a random experiment repeatedly for a large number of times (say n). At the i-th experiment let X_i be the observed value of the random variable. Then the "sample average" will converge to the theoretical expectation as $n \to \infty$:

$$\frac{X_1 + \cdots + X_n}{n} \to \mathbb{E}[X] \quad \text{as } n \to \infty.$$

This phenomenon, known as the *law of large numbers*, is one of the most significant results in probability theory. We will come back to this property in a precise mathematical form in later topics.

From the definition, we see that if X is a deterministic constant c (i.e. $\mathbb{P}(X = c) = 1$), then

$$\mathbb{E}[X] = c \times 1 = c.$$

In general, if S_X is a finite set, the expectation $\mathbb{E}[X]$ is always well defined since the sum in (2.1) is a finite sum. However, if S_X is an infinite set (for instance when $S_X = \mathbb{N}$), the sum in (2.1) becomes an infinite series and thus the expectation $\mathbb{E}[X]$ may not always exist. It is important to require that the series is absolutely convergent in order to ensure that $\mathbb{E}[X]$ is well defined. We use the following example to illustrate this.

Example 2.2. Toss a fair coin independently and repeatedly. Let Y be the number of tosses until a Head appears. Let $X = (-2)^Y$. Since $S_Y = \{1, 2, 3, \cdots\}$, we see that

$$S_X = \{(-2)^n : n = 1, 2, 3, \cdots\}.$$

Note that

$$\mathbb{P}(X = (-2)^n) = \mathbb{P}(Y = n) = \left(\frac{1}{2}\right)^{n-1} \times \frac{1}{2} = \frac{1}{2^n}.$$

As a result, the right hand side of (2.1) is given by

$$\sum_{x \in S_X} x \cdot p_X(x) = \sum_{n=1}^{\infty} (-2)^n \times \frac{1}{2^n} = \sum_{n=1}^{\infty} (-1)^n.$$

This series is not absolutely convergent, and the expectation of X is not well defined in this case.

We look a few simple examples of calculating the expectation.

Example 2.3. Toss a fair die. Let X be the resulting value. Then $S_X = \{1, 2, 3, 4, 5, 6\}$, and each outcome has probability $\frac{1}{6}$. Therefore, by definition we have

$$\mathbb{E}[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{7}{2}.$$

This example tells us that in general $\mathbb{E}[X]$ needs not be a possible value of X.

Example 2.4. A factory is producing a particular kind of items. Suppose that 10% of the items are defective, and 90% of the items are non-defective. If an item is defective the factory loses \$1 on it, while if it is non-defective the factory gains a profit \$5. Let X be the profit on a randomly selected item. Then

$$\mathbb{E}[X] = (-1) \times 10\% + 5 \times 90\% = 4.4.$$

2.2 The expectation of a continuous random variable with density

The way of defining the expectation in the discrete case immediately gives a natural idea for the continuous case (with density) by replacing the sum with an integral and replacing the pmf with the pdf.

Definition 2.2. Let X be a continuous random variable with pdf $f_X(x)$. The *expectation* of X is defined by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx, \tag{2.2}$$

provided that the integral on the right hand side is absolutely convergent.

Example 2.5. Let X be a continuous random variable with pdf

$$f_X(x) = \begin{cases} cx^2(1-x), & 0 < x < 1; \\ 0, & \text{otherwse.} \end{cases}$$

Find the value of c and $\mathbb{E}[X]$.

Solution. We use Property 1 of the pdf to calculate c:

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_0^1 cx^2(1-x)dx = c \times \big(\frac{1}{3} - \frac{1}{4}\big) = \frac{c}{12}.$$

Therefore, $c = 12$. By the definition of expectation, we have

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_0^1 x \cdot 12x^2(1-x)dx = \frac{3}{5}.$$

If we recall the infinitesimal interpretation of the pdf, we can describe an intuitive connection between the continuous and discrete cases for the expectation. Let δ be a small number, and let us partition the real line into intervals with length δ:

$$-\infty < \cdots < x_{i-1} < x_i < x_{i+1} < \cdots, \quad x_{i+1} = x_i + \delta.$$

Then we have

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \sum_i \int_{x_i}^{x_i+\delta} x f_X(x)dx.$$

Observe that, since δ is small, we have

$$\int_{x_i}^{x_i+\delta} x f_X(x)dx \approx x_i f_X(x_i)\delta.$$

In addition, according to the interpretation of the pdf we also have

$$\mathbb{P}(X \approx x_i) \approx f_X(x_i)\delta.$$

As a result, we see that

$$\mathbb{E}[X] \approx \sum_i x_i \mathbb{P}(X \approx x_i).$$

The right hand side resembles an expectation in the discrete case.

2.3 The accounting trick

In many situations, it is useful to know how to compute $\mathbb{E}[\psi(X)]$ where X is a random variable and $\psi : \mathbb{R} \to \mathbb{R}$ is a given function.

Before deriving the formula, we first need to introduce an alternative way of calculating the expectation which is sometimes known as the accounting trick. We illustrate this method in the following example.

Example 2.6. A couple is going to have three children. Suppose that a child is equally likely to be a boy or a girl. Let X be the number of girls in the three children. By writing down the pmf and using the definition of $\mathbb{E}[X]$, it is easy to see that $\mathbb{E}[X] = \frac{3}{2}$. To discuss another method, we first write down the sample space:

$$\Omega = \{\text{bbb, bbg, bgb, gbb, bgg, gbg, bbg, ggg}\}$$

where "b" and "g" represent boy and girl respectively. The problem is clearly a classical probability model, and thus each outcome is assigned the probability of $\frac{1}{8}$. As we have mentioned, the expectation $\mathbb{E}[X]$ can be viewed as an average of the values of X by the corresponding probabilities, it is natural to expect that we can also calculate $\mathbb{E}[X]$ via

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}) \\
&= X(\text{bbb})\mathbb{P}(\{\text{bbb}\}) + X(\text{bbg})\mathbb{P}(\{\text{bbg}\}) + X(\text{bgb})\mathbb{P}(\{\text{bgb}\}) \\
&\quad + X(\text{gbb})\mathbb{P}(\{\text{gbb}\}) + X(\text{bgg})\mathbb{P}(\{\text{bgg}\}) + X(\text{gbg})\mathbb{P}(\{\text{gbg}\}) + \\
&\quad + X(\text{ggb})\mathbb{P}(\{\text{ggb}\}) + X(\text{ggg})\mathbb{P}(\{\text{ggg}\}) \\
&= 0 \times \frac{1}{8} + 1 \times \frac{1}{8} + 1 \times \frac{1}{8} + 1 \times \frac{1}{8} + 2 \times \frac{1}{8} + 2 \times \frac{1}{8} + 2 \times \frac{1}{8} + 3 \times \frac{1}{8} \\
&= \frac{3}{2}.
\end{aligned}$$

In general, if the sample space Ω is countable, the alternative formula

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}) \tag{2.3}$$

of computing the expectation is called the *accounting trick*. To see that this is equivalent to the original definition, we can decompose Ω by a partition according to the value of X :

$$\Omega = \cup_{x\in S_X}\{X = x\}.$$

As a result, the summation over $\omega \in \Omega$ can be performed by first summing over each event $\{X = x\}$ and then summing over the different values of x. Namely, we have

$$
\begin{aligned}
&\sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) \\
&= \sum_{x \in S_X} \sum_{\omega \in \{X=x\}} X(\omega)\mathbb{P}(\{\omega\}) \\
&= \sum_{x \in S_X} x \sum_{\omega \in \{X=x\}} \mathbb{P}(\{\omega\}) \quad \text{(since } X(\omega) = x \text{ on the event } \{X = x\}) \\
&= \sum_{x \in S_X} x \cdot \mathbb{P}(X = x) = \sum_{x \in S_X} x \cdot p_X(x) \qquad (2.4)
\end{aligned}
$$

Remark 2.1. This is a deep theoretical point which is not so relevant to our current study. The formula (2.3) is also true for continuous random variables as long as the summation is replaced by a "suitable" notion of integration over the sample space Ω. Symbolically, we always have

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

In fact, in modern probability theory, the above formula is used as the definition of the expectation for arbitrary random variables. It is then a theorem that in the discrete case, the expectation can be computed by the formula (2.1) while in the continuous case it can be computed by the formula (2.2).