

# Math 525: Lecture 24

April 12, 2018

## 1 Conditional distribution and expectations

Let  $X$  and  $Y$  be random variables. If the event  $\{X = x\}$  occurs with positive probability, we can define the *conditional distribution function* by

$$F_{Y|X}(y | x) = \mathbb{P}\{Y \leq y | X = x\}.$$

If  $Y$  is also integrable, we can also define the *conditional expectation* of  $Y$  given  $X = x$ :

$$\mathbb{E}[Y | X = x] = \frac{\mathbb{E}[Y I_{\{X=x\}}]}{\mathbb{P}\{X = x\}}.$$

We saw in a previous class that if in addition to being integrable,  $Y$  is discrete, the above simplifies:

$$\mathbb{E}[Y | X = x] = \frac{\sum_n y_n \mathbb{P}\{Y = y_n, X = x\}}{\mathbb{P}\{X = x\}} = \sum_n y_n \mathbb{P}\{Y = y_n | X = x\}.$$

However, when  $X$  has a continuous distribution, the event  $\{X = x\}$  has probability zero. While it is possible to define  $\mathbb{E}[Y | \cdot]$  in a general way that avoids this issue, doing so requires knowledge of some concepts from measure theory (e.g., Radon-Nikodym derivatives). Instead of working in the most general setting, we will instead tackle the case in which  $X$  and  $Y$  admit a joint density  $f_{XY}$ .

To motivate our definition of conditional distribution in this setting, suppose events of the form  $\{x \leq X \leq x + h\}$  have positive probability (where  $h > 0$ ) and consider

$$\mathbb{P}(Y \leq y | x \leq X \leq x + h).$$

Since  $X$  and  $Y$  admit a joint density,

$$\mathbb{P}(Y \leq y | x \leq X \leq x + h) = \frac{\mathbb{P}\{Y \leq y, x \leq X \leq x + h\}}{\mathbb{P}\{x \leq X \leq x + h\}} = \frac{\int_{-\infty}^y \int_x^{x+h} f_{XY}(u, v) du dv}{\int_x^{x+h} f_X(u) du}.$$

If  $h$  is small, we expect the above to be approximately

$$\frac{h \int_{-\infty}^y f_{XY}(x, v) dv}{h f_X(x)} = \frac{\int_{-\infty}^y f_{XY}(x, v) dv}{f_X(x)} = \int_{-\infty}^y \left( \frac{f_{XY}(x, v)}{f_X(x)} \right) dv.$$

This motivates the definition below.

**Definition 1.1.** Let  $X$  and  $Y$  be random variables which admit a joint density  $f_{XY}$ . We can define the *conditional density* of  $Y$  given  $X = x$  by

$$f_{Y|X}(y | x) = \begin{cases} \frac{f_{XY}(x, y)}{f_X(x)} & \text{if } f_X(x) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

The *conditional distribution function* of  $Y$  given  $X = x$  is

$$F_{Y|X}(y | x) = \int_{-\infty}^y f_{Y|X}(v | x) dv.$$

The *conditional expectation* of  $Y$  given  $X = x$  is

$$\mathbb{E}[Y | X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y | x) dy.$$

*Remark 1.2.* If  $X$  and  $Y$  are independent in the above, the conditional density of  $Y$  given  $X = x$  should be independent of  $X$ . Indeed, then

$$f_{Y|X}(y | x) = \frac{f_{XY}(x, y)}{f_X(x)} = f_Y(y).$$

Let's also verify that the conditional density works as desired. Indeed,

$$\begin{aligned} \int_a^b \overbrace{\left( \int_c^d f_{Y|X}(y | x) dy \right)}^{\mathbb{P}(c < Y \leq d | X = x)} f_X(x) dx &= \int_a^b \int_c^d \frac{f_{XY}(x, y)}{f_X(x)} f_X(x) dy dx \\ &= \int_a^b \int_c^d f_{XY}(x, y) dy dx \\ &= \mathbb{P}(a < X \leq b, c < Y \leq d). \end{aligned}$$

**Definition 1.3.** Suppose  $X$  is a discrete random variable or that  $X$  and  $Y$  admit a joint density. Suppose  $Y$  is integrable, so that the function  $\psi$  given by

$$\psi(x) = \mathbb{E}[Y | X = x]$$

is well-defined (if  $X$  is discrete, we can safely ignore points at which  $\{X = x\}$ ). Then, the conditional expectation of  $Y$  given  $X$  is

$$\mathbb{E}[Y | X] \equiv \psi \circ X.$$

Unlike the previous definition, we have not specified the value of  $X$  in the conditional expectation. Therefore, unlike  $\mathbb{E}[Y | X = x]$  (which is a scalar),  $\mathbb{E}[Y | X]$  is itself a random variable. We now prove what is known as the *tower property of expectations*:<sup>1</sup>

---

<sup>1</sup>Actually, the tower property is more general, applying to the more general notion of conditional expectation we hinted at before.

**Proposition 1.4.** *Let  $X$  and  $Y$  be random variables which admit a joint density  $f_{XY}$ . Suppose  $f_X \neq 0$ . If  $Y$  is integrable, then  $\mathbb{E}[Y | X]$  is also integrable and*

$$\mathbb{E}[\mathbb{E}[Y | X]] = \mathbb{E}Y.$$

*Proof.* First, suppose  $Y$  is positive. Then,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[Y | X]] &= \mathbb{E}[\psi(X)] \\ &= \int_{-\infty}^{\infty} \psi(x) f_X(x) dx \\ &= \int_{-\infty}^{\infty} \mathbb{E}[Y | X = x] f_X(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y|X}(y | x) dy f_X(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \frac{f_{XY}(x, y)}{f_X(x)} dy f_X(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{XY}(x, y) dy dx \\ &= \mathbb{E}[Y]. \end{aligned}$$

If  $Y$  is not positive, we can split it into its positive and negative parts  $Y = Y^+ - Y^-$ .  $\square$

The tower property also holds for discrete random variables, in which case the intuition is very easy to understand:

**Example 1.5.** Consider two bags, labelled  $a$  and  $b$ . We pick bag  $a$  with probability  $P$  and bag  $b$  with probability  $1 - p$ . We use the random variable  $X$  to indicate which bag we picked (i.e.,  $X$  takes values in  $\{a, b\}$ ). Both bags are filled with green and red balls. Define

$$p_a = \frac{\# \text{ of green balls in bag } a}{\text{total } \# \text{ of balls in bag } a} \quad \text{and} \quad p_b = \frac{\# \text{ of green balls in bag } b}{\text{total } \# \text{ of balls in bag } b}.$$

From the bag we have chosen, we pick a single ball. Let

$$Y = \begin{cases} 1 & \text{if we picked a green ball} \\ 0 & \text{if we picked a red ball.} \end{cases}$$

Then,

$$\mathbb{E}[Y | X = a] = p_a \quad \text{and} \quad \mathbb{E}[Y | X = b] = p_b.$$

Therefore,

$$\mathbb{E}[\mathbb{E}[Y | X]] = P(\mathbb{E}[Y | X = a]) + (1 - P)\mathbb{E}[Y | X = b] = Pp_a + (1 - P)p_b = \mathbb{E}[Y].$$

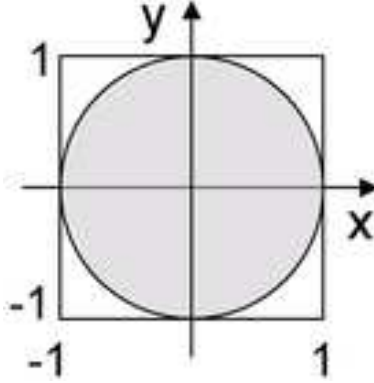


Figure 1: Circle with area  $\pi$

## 2 Introduction to Monte Carlo methods

A Monte Carlo method is any algorithm that exploits the law of large numbers to compute a (deterministic) quantity. This is best motivated by example:

**Example 2.1.** Suppose we want to approximate the value of  $\pi$ . Recall that for a (Borel) function  $f$ ,

$$\mathbb{E}[f(X)] = \frac{1}{b-a} \int_a^b f(x)dx \quad \text{if } X \sim U[a, b].$$

The same is true in higher dimensions:

$$\mathbb{E}[f(X, Y)] = \frac{1}{d-c} \frac{1}{b-a} \int_c^d \int_a^b f(x, y) dx dy \quad \text{if } X \sim U[a, b] \text{ and } Y \sim U[c, d].$$

Recall that  $\pi$  is the area of a unit circle. That is, letting  $C = \{(x, y) : x^2 + y^2 \leq 1\}$ ,

$$\pi = \int_{-1}^1 \int_{-1}^1 I_C(x, y) dx dy.$$

In other words,

$$\mathbb{E}[I_C(X, Y)] = \frac{\pi}{4} \quad \text{where } X, Y \sim U[-1, 1].$$

Let  $X_n, Y_n \sim U[-1, 1]$ . Then, by the strong law of large numbers,

$$\lim_{N \rightarrow \infty} \frac{I_C(X_1, Y_1) + \dots + I_C(X_N, Y_N)}{N} = \frac{\pi}{4} \text{ a.s.}$$

This gives us a very simple algorithm to approximate  $\pi$ :

1. Pick a positive integer  $N$  and set  $S_N$  to zero.
2. Generate two (independent) random variables  $X$  and  $Y$  from the uniform distribution  $U[-1, 1]$ . If  $X^2 + Y^2 \leq 1$ , increment  $S_N$  by one.
3. Repeat step (2)  $n$  times. Then,  $\pi/4 \approx S_N/N$ .

Some MATLAB code for this procedure is given below:

---

**Algorithm 1** Monte Carlo method to approximate  $\pi$ 

---

```
% Generate (X, Y) pairs
pairs = 2 * rand(2, N) - 1;

% Count which pairs are in the circle
S_N = sum(sum(pairs.^2) <= 1);

pi_approximate = 4 * (S_N / N);
```

---

With  $N$  equal to 10 million, below are some results of running the code:

Run	Result
1	3.1414432
2	3.1416316
3	3.1418980
4	3.1416388
5	3.1424888

Even though we used 10 million samples in the above, our approximation is only accurate up to three significant digits! The slow convergence<sup>2</sup> is explained by the CLT, which says

$$\sqrt{N} \left( \frac{S_N}{N} - \frac{\pi}{4} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma)$$

where  $\sigma^2 = \text{Var}(I_C(X, Y))$ . With a slight abuse of notation, we may write

$$|\pi_{\text{approximate}} - \pi| = \left| 4 \frac{S_N}{N} - \pi \right| = O \left( \frac{1}{\sqrt{N}} \right)$$

to represent this fact.

---

<sup>2</sup>There are much more effective ways to compute  $\pi$  (e.g., Chudnovsky algorithm), but they do not have to do with Monte Carlo.