

# Multivariate Probability Distributions

March 20, 2020

## 1 Multivariate Discrete Random Variables

So far we have mostly focused on just one random variable at a time, but in many problems there might be several random variables of interest. On the homework, we have already seen several examples where these ideas are in the background, but now we will formalize it. We begin with the discrete random variables:

**DEFINITION 1.1.** *Let  $Y_1$  and  $Y_2$  be discrete random variables. The joint probability mass function (jpmf or pmf) for  $Y_1$  and  $Y_2$  is given by*

$$p_{Y_1, Y_2}(y_1, y_2) = \mathbb{P}(\{Y_1 = y_1\} \cap \{Y_2 = y_2\})$$

In other words for any two numbers  $y_1, y_2$ ,  $p_{Y_1, Y_2}(y_1, y_2)$  gives that probability that  $Y_1$  equals  $y_1$  and  $Y_2$  equals  $y_2$ . Let's look at some examples

**EXAMPLE 1.2.** *Roll a fair 4-sided dice twice. Let  $Y_1$  be the result of the first roll and  $Y_2$  be the result of the second. What is the joint pmf of  $Y_1$  and  $Y_2$ ?*

**Solution:** Since the results of the two rolls are independent, and the dice are fair, each of the  $4^2 = 16$  possibilities are equally likely. Since the sum of all probabilities is 1, we conclude each possibility has probability  $1/16$ . Formally we can write:

$$p_{Y_1, Y_2}(y_1, y_2) = \frac{1}{16} \text{ for } y_1 = 1, 2, 3, 4; y_2 = 1, 2, 3, 4 \text{ and } 0 \text{ otherwise}$$

**EXAMPLE 1.3.** *There are 20 balls in a box, labeled  $1, 2, \dots, 20$ . You draw 2 balls out of the box without replacement. Let  $X_1$  be the number of the first ball you draw and  $X_2$  be the number on the second ball. What is the joint pmf of  $X_1$  and  $X_2$ ?*

**Solution:** There are  $20 \times 19$  ways to choose 2 balls out 20, keeping track of order (note: it's not  $\binom{20}{2}$  because we're keeping track of order). All of choices are equally likely so

$$p_{X_1, X_2}(x_1, x_2) = \frac{1}{20 \times 19} \text{ for } x_1 = 1, 2, \dots, 20; x_2 = 1, 2, \dots, 20 \text{ with } x_1 \neq x_2 \text{ and } 0 \text{ otherwise .}$$

The random variables don't need to come from some given experiment sometimes they can just be given, for example:

EXAMPLE 1.4. *Verify that*

$$p_{X_1, X_2}(1, 2) = 1/3, p_{X_1, X_2}(5, -2) = 1/2, p_{X_1, X_2}(6, 4) = 1/6 \text{ and } 0 \text{ otherwise}$$

*is a valid joint pmf.*

**Solution:** Since clearly  $0 < 1/3, 1/6, 1/2$ , we just need to note the  $1/2 + 1/3 + 1/6 = 1$ .

This last example means there are two random variables  $X_1$  and  $X_2$ , the random variable  $X_1$  can be 1, 5, or 6 and the random variable  $X_2$  can be 2, -2, or 4. If the  $X_1$  is 5 then  $X_2$  must be -2. Note that this situation is very different from the first example where knowing  $Y_1$  told you nothing about  $Y_2$ . We'll expand on this idea more in the next section.

## 2 Multivariate Continuous Random Variables

Just like in the case of one random variable we can define a Cumulative Distribution function for pairs of random variables.

DEFINITION 2.1. *Let  $X_1$  and  $X_2$  be random variables, the Joint Cumulative Distribution Function (JCDF or CDF) of  $X_1$  and  $X_2$  is*

$$F_{X_1, X_2}(x_1, x_2) = \mathbb{P}(\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$$

We'll first discuss some general properties of the CDF, then we'll introduce the joint pdf and see some examples.

Properties of the CDF:

1. For any  $x_1, x_2$ ,

$$\lim_{t_1 \rightarrow -\infty, t_2 \rightarrow -\infty} F(t_1, t_2) = \lim_{t_1 \rightarrow -\infty} F(t_1, x_2) = \lim_{t_2 \rightarrow -\infty} F(x_1, t_2) = 0$$

The first limit is the probability that both random variables are more negative than any real number. In the next two limits, it's the probability just one of two random variables is more negative than any real number, while the other just needs to be less than some fixed  $x_1$  or  $x_2$ , respectively. In all three cases the probability goes to 0.

- 2.

$$\lim_{t_1 \rightarrow \infty, t_2 \rightarrow \infty} F(t_1, t_2) = 1$$

This is giving the probability the random variables are both less than some arbitrarily large number, which goes to 1. Note that we need both limits.

3. For any  $a_1 \leq b_1, a_2 \leq b_2$ ,

$$F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geq 0$$

The left side is just a complicated way of writing  $\mathbb{P}(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2)$ , which must be non-negative.

The CDF is always an increase function if it additionally continuous, then in turns out it can be differentiated almost everywhere. We call this derivative the pdf.

DEFINITION 2.2. *The joint probability density function (jpdf or pdf) is given by*

$$f_{X_1, X_2}(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1, X_2}(x_1, x_2)$$

From the fundamental theorem of calculus we have the inverse relationship

$$F_{X_1, X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1, X_2}(t_1, t_2) dt_1 dt_2$$

Note, once again, the use of dummy variables  $t_1, t_2$ .

From the properties of the CDF, we can deduce that

$$f_{X_1, X_2}(x_1, x_2) \geq 0 \text{ for all } x_1, x_2$$

and that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = 1.$$

It turns out that any function with these properties is the jpdf of some pair of random variables.

Just like in the 1 variables case our formula for computing probabilities is similar to the discrete case, but we replace pmf with the pdf and sums with integrals:

$$\mathbb{P}(a_1 \leq X_1 \leq b_1, a_2 \leq X_2 \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{X_1, X_2}(x_1, x_2) dx_2 dx_1$$

EXAMPLE 2.3. *Let  $X_1$  and  $X_2$  be continuous random variables with jpdf:*

$$f_{X_1, X_2}(x_1, x_2) = \begin{cases} C, & \text{if } -1 \leq x_1 \leq 3 \text{ and } 2 \leq x_2 \leq 4 \\ 0, & \text{otherwise.} \end{cases}$$

*What is the value of  $C$ ?*

**Solution:** We need

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = \int_{-1}^3 \int_2^4 C dx_2 dx_1 = 1.$$

Note that we restrict the bounds of integration, because the pdf, and hence the integral of the pdf, is zero outside this range. You can then compute this integral, or note that this double integral gives the area of a rectangle with side lengths 4 and 2, so the integral is 8. The height  $C$  must then chosen so that the total volume is 1, so

$$C = 1/8$$

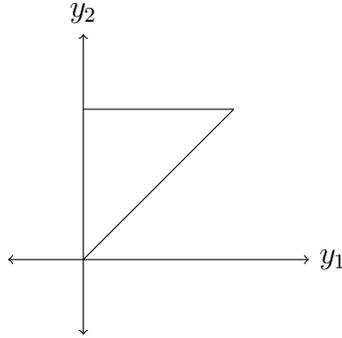
EXAMPLE 2.4. *Let  $Y_1$  and  $Y_2$  be continuous random variables with jpdf:*

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} \frac{3}{4}y_1^2, & \text{if } 0 \leq y_1 \leq y_2 \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

- (a) Verify that  $f$  is a pdf.
- (b) What is the  $\mathbb{P}(Y_2 > 1, Y_1 < 1)$ ?
- (c) What is the  $\mathbb{P}(Y_2 > 1, Y_1 > 1)$ ?

**Solution:**

(a) The first thing we need to do is plot the support of the jpdf, namely the region bounded by the lines  $0 = y_1$ ,  $y_1 = y_2$  and  $y_2 = 2$ . We do this plot the three lines, finding their intersection points and then determining which side of line the inequality holds. In this problem we get



the triangle:

Now we want to compute the integral of  $\frac{3}{4}y_1^2$  over this region. Recall from Multivariable calculus we want to compute an iterated integral. We can either integrate over the  $y_1$  or  $y_2$  first. Let's do both ways to see the difference.

If we integrate over  $y_1$  first, we fix a  $y_2$  in between 0 and 2 (this is all possible  $y_2$  values in the region), then for each fixed  $y_2$  see how  $y_1$  varies. In this case, it starts at 0 and continues until it hits the line  $y_1 = y_2$ , so our integral is:

$$\int_0^2 \int_0^{y_2} \frac{3}{4}y_1^2 dy_1 dy_2.$$

We now just need to do this integral and see that the answer is 1.

$$\begin{aligned} \int_0^2 \int_0^{y_2} \frac{3}{4}y_1^2 dy_1 dy_2 &= \frac{3}{4} \int_0^2 \left. \frac{y_1^3}{3} \right|_0^{y_2} dy_2 \\ &= \frac{3}{4} \int_0^2 \left( \frac{y_2^3}{3} - 0 \right) dy_2 \\ &= \frac{3}{4} \left. \frac{y_2^4}{3 \cdot 4} \right|_0^2 \\ &= 1 \end{aligned}$$

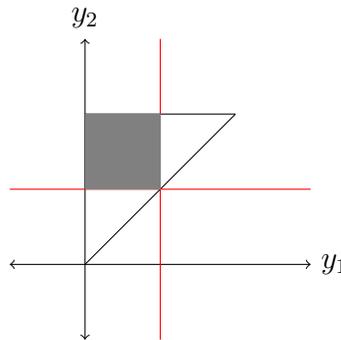
Let's try to integrate over  $y_2$  first. We fix  $y_1$  between 0 and 2 (this is all possible  $y_1$  values in the region), then for each fixed  $y_1$  see how  $y_2$  varies. In this case, it starts at the line  $y_2 = y_1$  and continues until it hits the line  $y_2 = 2$ , so our integral is:

$$\int_0^2 \int_{y_1}^2 \frac{3}{4}y_1^2 dy_2 dy_1.$$

We now just need to do this integral and see that the answer is 1.

$$\begin{aligned} \int_0^2 \int_{y_1}^2 \frac{3}{4} y_1^2 dy_2 dy_1 &= \frac{3}{4} \int_0^2 y_1^2 y_2 \Big|_{y_1}^2 dy_1 \\ &= \frac{3}{4} \int_0^2 y_1^2 (2 - y_1) dy_1 \\ &= \frac{3}{4} \left( 2 \frac{y_1^3}{3} - \frac{y_1^4}{4} \right) \Big|_0^2 \\ &= 1 \end{aligned}$$

(b) Inside of this triangle the pdf is positive and outside it is negative. To determine  $\mathbb{P}(Y_2 > 1, Y_1 < 1)$ ,



we draw the lines  $y_2 = 1$  and  $y_1 = 1$ .

Then  $\mathbb{P}(Y_2 > 1, Y_1 < 1)$  will correspond to integrating over the intersection of the the region above the line  $y_2 = 1$  (because we want  $Y_2 > 1$ ), the region to the left of  $y_1 = 1$  (because we want  $Y_1 < 1$ ), and the triangle (because that is where the pdf is positive).

If we integrate over  $y_1$  first, we fix a  $y_2$  in between 1 and 2, then for each fixed  $y_2$  see how  $y_1$  varies. In this case, it's just between 0 and 1. So our integral is:

$$\mathbb{P}(Y_2 > 1, Y_1 < 1) = \int_1^2 \int_0^1 \frac{3}{4} y_1^2 dy_1 dy_2$$

Note the the value of the pdf  $y_1^2$  didn't appear until this point, we can forget about while setting up the region of integration. To compute this probability we simply do the inner integral and then do the outer integral:

$$\begin{aligned} \mathbb{P}(Y_2 > 1, Y_1 < 1) &= \int_1^2 \int_0^1 \frac{3}{4} y_1^2 dy_1 dy_2 \\ &= \frac{3}{4} \int_1^2 \frac{y_1^3}{3} \Big|_0^1 dy_2 \\ &= \frac{3}{4} \int_1^2 \left( \frac{1}{3} - 0 \right) dy_2 \\ &= \frac{3}{4} \frac{1}{3} (2 - 1) \\ &= \frac{1}{4} \end{aligned}$$

(c) Exercise for you. The answer is  $\frac{11}{16}$ .