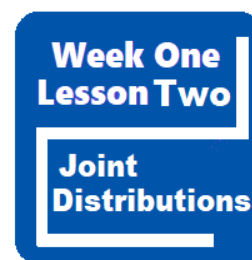


# Statistical Inference for Estimation in Data Science

DTSA 5002 offered on Coursera  
by the University of Colorado, Boulder  
Instructor: J.N. Corcoran



Let  $X$  and  $Y$  be discrete random variables. Often, we want to talk about the probability that  $X = x$  and  $Y = y$  at the same time. This is a *joint probability* which can be written as

$$P(X = x, Y = y).$$

(The comma here is read as “and”.)

In DTSA 5001 (or any other beginner probability course), you would have seen several “event”-based probabilities for events denoted by letters like  $A$ ,  $B$ , and  $C$ . For example, if we roll a fair 6-sided die, we might let  $A$  be the event that the outcome is an even number and  $B$  be the event that the outcome is greater than or equal to 4. We have

$$A = \{2, 4, 6\} \quad \text{and} \quad B = \{4, 5, 6\}.$$

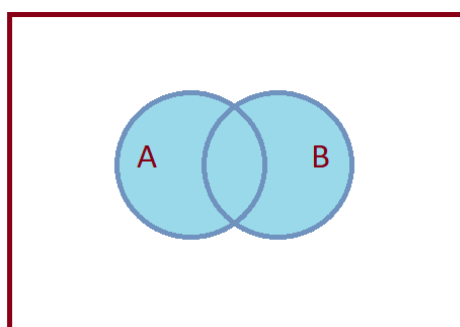
We also have

$$A \cup B = \{2, 4, 5, 6\} \quad \text{and} \quad A \cap B = \{4, 6\}.$$

There are many rules of probability such as

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{1}$$

that can be easily recalled by thinking of probability as area in a Venn diagram containing circles representing  $A$  and  $B$ . In the following figure,  $A \cup B$  is shaded.



Note that the area of  $A \cup B$  is the area of  $A$  plus the area of  $B$  minus the area of the intersection since it got double-counted. This is how we can remember Equation (1).

Getting back to  $P(X = x, Y = y)$ , note that we can apply all of our “event” based probability rules here. We can think of this probability as  $P(A \cap B)$  where

$$\begin{aligned} A &= \text{the event that } X \text{ equals } x \\ B &= \text{the event that } Y \text{ equals } y \end{aligned}$$

## Discrete Joint Distributions

There are not many joint distributions that have names. Most often, they consist of probabilities given in tabular form as in the following example.

		$x$		
		1	2	3
$y$	-5	0.2	0.1	0.05
	6	0.15	0.3	0.2

Here, the entries are probabilities corresponding to row and column heading values for  $X$  and  $Y$ . For example,

$$P(X = 2, Y = 6) = 0.3.$$

Suppose now that we just want to find the probability that  $X = 2$  and are not concerned with the value of  $Y$ ? There are 2 ways this can happen. We can either have  $X = 2$  and  $Y = -5$  or we can have  $X = 2$  and  $Y = 6$ . Since these are disjoint events, we can simply add the two probabilities just as we can add two disjoint pieces of area in a Venn diagram. We have

$$P(X = 2) = P(X = 2, Y = -5) + P(X = 2, Y = 6) = 0.1 + 0.3 = 0.4.$$

Note that this and other probabilities for “ $X$  alone” or “ $Y$  alone” can be found in the margins of the tables if we include row and column sums.

		$x$			
		1	2	3	
$y$	-5	0.2	0.1	0.05	0.35
	6	0.15	0.3	0.2	0.65
		0.35	0.4	0.25	

This is why the separate distributions for  $X$  and  $Y$  alone:

$x$	1	2	3
$P(X = x)$	0.35	0.4	0.25

and

$y$	-5	6
$P(Y = y)$	0.35	0.65

are known as **marginal distributions!**

Analogous to the one-dimensional case, We will typically use  $f(x, y)$  to denote the **joint probability mass function**

$$f(x, y) = P(X = x, Y = y).$$

The marginal pmf for  $X$  is obtained by summing out  $y$ :

$$f_X(x) = \sum_y f(x, y).$$

(We have used this subscript notation on the left-hand side so that we may use the letter  $f$  again for the other marginal pmf.)

Similarly, the marginal pmf for  $Y$  is obtained by summing out  $x$ :

$$f_Y(y) = \sum_x f(x, y).$$

## Continuous Joint Distributions

Suppose now that  $X$  and  $Y$  are continuous random variables. As per our discussion in the univariate case, a joint probability  $P(X = x, Y = y)$  will always be zero.

In the one-dimensional setting, the probability density function (pdf) was a curve under which area represented probability. In the two-dimensional setting, the joint pdf for  $X$  and  $Y$  might be called  $f(x, y)$  and would give a surface under which *volume* represents probability. We must have

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

We can compute probabilities involving  $X$  and  $Y$  by integrating over the appropriate region. For example,

$$P(X \leq 1, 2 < Y < 6) = \int_{-\infty}^1 \int_2^6 f(x, y) dy dx$$

and

$$P(X \leq Y) = \int_{-\infty}^{\infty} \int_x^{\infty} f(x, y) dy dx = \int_{-\infty}^{\infty} \int_{-\infty}^y f(x, y) dx dy.$$

(It may help you to draw the region where  $x \leq y$  in the  $xy$ -plane in order to find the limits of integration needed.)

Analogous to the discrete case, we can get to the pdfs for  $X$  or  $Y$  alone by integrating out the unwanted variable. These *marginal pdfs* are

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

It may be the case that the joint pdf is zero in a lot of places and the integrals from  $-\infty$  to  $\infty$  may be equivalent to integration over a finite region. Examples are given in the video associated with this lesson.