

# Math 130B

Liam Hardiman

August 1, 2022

## Abstract

I'm writing these lecture notes for UC Irvine's Math 130B course, taught in the summer of 2022. This is a five-ish week course where I plan to get through chapters 6-8 of Ross' book [1]. The class structure consists of a two hour lecture followed by a one hour discussion section three days a week. I'm aiming to get through one or two sections of the book per lecture with a midterm soon after chapter 6, maybe partway into chapter 7.

## Contents

<b>1</b>	<b>Jointly Distributed Random Variables</b>	<b>1</b>
1.1	Joint Distribution Functions . . . . .	1
1.2	Independent Random Variables . . . . .	7

## 1 Jointly Distributed Random Variables

Many of life's more interesting problems are multifaceted. For example, in a clinical trial for a cholesterol drug, we might be interested in a patient's cholesterol levels *and* how many hours they exercise each week. Or if we're interested in California's gas consumption, we'd be interested in how much gas each station sells *and* its price of gas.

In this section, we address how to look at more than one random variable at the same time.

### 1.1 Joint Distribution Functions

Remember that we can define the *probability mass function* of a discrete random variable  $X$  to be the function that takes in a value and returns the probability that  $X$  attains that value.

$$p(a) = \Pr[X = a].$$

**Examples 1.1.** (a) Suppose we roll a pair of dice and  $X$  is the sum of the values shown. Then  $X$  can take any integer value between 2 and 12 and its probability mass function is

$$\begin{array}{ll} p(2) = p(12) = \frac{1}{36} & p(3) = p(11) = \frac{2}{36} \\ p(4) = p(10) = \frac{3}{36} & p(5) = p(9) = \frac{4}{36} \\ p(6) = p(8) = \frac{5}{36} & p(7) = \frac{6}{36}. \end{array}$$

- (b) Suppose Alice is communicating with Bob by sending him bits (0's or 1's) one by one. Suppose each bit Alice sends has probability  $p$  of successfully getting to Bob and each transmission is independent of the others. If  $X$  is the first time a bit fails to transmit properly (maybe there's too much noise on the channel), then  $X$  is a geometric random variable with probability mass function

$$p(n) = p^{n-1}(1 - p).$$

Situations naturally arise where we might want to look at two discrete random variables at the same time. For example, if we roll two dice and  $X$  is the smaller roll and  $Y$  is the larger one, can we define an analogue of the probability mass function?

**Definition 1.2.** Suppose  $X$  and  $Y$  are two discrete random variables taking values in the sets  $A$  and  $B$ , respectively. Then their *joint probability mass function* is the function  $p : A \times B \rightarrow [0, 1]$  defined by

$$p(a, b) = \Pr[X = a, Y = b].$$

**Example 1.3.** Say we roll two dice and  $X$  is the largest value shown and  $Y$  is the sum of the two values. Let's compute a few values of the joint probability mass function of  $X$  and  $Y$ . We have

$$\begin{aligned} p(3, 5) &= \Pr[X = 3, Y = 5] \\ &= \Pr[\{(3, 2), (2, 3)\}] \\ &= \frac{2}{36}. \end{aligned}$$

This is because the only way for the largest value to be 3 and the sum to be 5 is for one of the dice to show 2 and the other to show 3. We also have

$$p(1, 8) = 0$$

since there's no way for two dice to sum to 8 and the largest value be a 1.

How does the joint mass function of  $X$  and  $Y$  relate to the *marginal* probability mass functions? Well if we just specify that  $X = a$ , then we haven't put any restrictions on  $Y$ . This gives us

$$\begin{aligned} p_X(a) &= \Pr[X = a] \\ &= \Pr[X = a, Y < \infty] \\ &= \sum_{y \in B} \Pr[X = a, Y = y] \\ &= \sum_{y \in B} p(a, y). \end{aligned} \tag{1}$$

Similarly, we have

$$p_Y(b) = \sum_{x \in A} p(x, b).$$

**Example 1.4.** Say 100 people are asked for their handedness (right-handed or left-handed) and sex (male or female). The survey produces the following table.

	L	R
M	4	44
F	9	43

If we randomly select one of these people and let  $X$  be their sex and  $Y$  be their handedness, then we can obtain the joint probability mass function by just reading off values from the table.

$$\begin{aligned} p(M, L) &= 4/100 & p(M, R) &= 44/100 \\ p(F, L) &= 9/100 & p(F, R) &= 43/100. \end{aligned}$$

Let's compute the marginal probability mass functions. For  $X$  we have

$$\begin{aligned} p_X(M) &= p(M, L) + p(M, R) = \frac{4}{100} + \frac{44}{100} = \frac{48}{100} \\ p_X(F) &= p(F, L) + p(F, R) = \frac{9}{100} + \frac{43}{100} = \frac{52}{100}. \end{aligned}$$

For  $Y$  we have

$$\begin{aligned} p_Y(L) &= p(M, L) + p(F, L) = \frac{4}{100} + \frac{9}{100} = \frac{13}{100} \\ p_Y(R) &= p(M, R) + p(F, R) = \frac{44}{100} + \frac{43}{100} = \frac{87}{100}. \end{aligned}$$

**Example 1.5.** In the previous example we determined the marginal mass function from the joint mass function. Can we go the other way? That is, if we know the marginal mass functions for  $X$  and  $Y$ , can we determine the joint mass function? Well here's another possible outcome of the same survey from the previous example.

	L	R
M	3	45
F	10	42

It's easy to check that we get the same marginal mass functions in this modified example. So if we started with the marginals, how would we know whether the survey outcome was given by this table or the previous one? Since we can't really tell, it looks like the marginals don't determine the joint.

Let's be a little more specific. Suppose the marginals are specified by these equations

$$\begin{aligned} p_X(M) &= p_{ML} + p_{MR} = 48/100 \\ p_X(F) &= p_{FL} + p_{FR} = 52/100 \\ p_Y(L) &= p_{ML} + p_{FL} = 13/100 \\ p_Y(R) &= p_{MR} + p_{FR} = 87/100. \end{aligned}$$

Finding the joint mass function amounts to solving this system for the variables  $p_{ML}, p_{MR}, p_{FL}, p_{FR}$ . This is a linear system with four equations and four unknowns, so this sounds promising. The corresponding matrix equation is

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{ML} \\ p_{MR} \\ p_{FL} \\ p_{FR} \end{bmatrix} = \begin{bmatrix} 48/100 \\ 52/100 \\ 13/100 \\ 87/100 \end{bmatrix}.$$

If we go through the usual procedure of row-reduction, the coefficient matrix reduces to

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This matrix doesn't have full rank, so the system does *not* have a unique solution. In particular, there isn't just one joint mass function corresponding to these marginals.

Let's move on to continuous random variables. Remember that every (real-valued) random variable  $X$  gives us a function  $F_X : \mathbb{R} \rightarrow [0, 1]$  called its *(cumulative) distribution function*:

$$F_X(t) = \Pr[X \leq t]. \quad (2)$$

Likewise, if we have two random variables  $X$  and  $Y$ , we can define their *joint (cumulative) distribution function*.

**Definition 1.6.** Let  $X$  and  $Y$  be two random variables. Then their *joint cumulative distribution function*,  $F : \mathbb{R}^2 \rightarrow [0, 1]$  is defined by

$$F(a, b) = \Pr[X \leq a, Y \leq b].$$

If there's any possibility for ambiguity, we might write  $F_{X,Y}$  to remind us that  $F$  is the cumulative distribution function for  $X$  and  $Y$ .

How is the joint distribution function related to the *marginal* distribution functions of  $X$  and  $Y$ ? Like in the discrete case, if we just specify that  $X \leq a$ , then we haven't put any restrictions on  $Y$ . This gives us

$$\begin{aligned} F_X(a) &= \Pr[X \leq a] \\ &= \Pr[X \leq a, Y < \infty]. \end{aligned} \quad (3)$$

Now the events  $\{X \leq a, Y \leq t\}$  form an increasing sequence of events as  $t$  increases. That is, if  $t_1 < t_2$ , then we have the inclusion

$$\{X \leq a, Y \leq t_1\} \subseteq \{X \leq a, Y \leq t_2\}.$$

This is helpful because probabilities play nicely with increasing (or decreasing) sequences of events. Namely, if  $E_1 \subseteq E_2 \subseteq \dots$  is an increasing sequence of events, then

$$\Pr \left[ \bigcup_{n=1}^{\infty} E_n \right] = \lim_{n \rightarrow \infty} \Pr[E_n].$$

Using this, (3) becomes

$$\begin{aligned} F_X(a) &= \Pr[X \leq a, Y < \infty] \\ &= \Pr \left[ \bigcup_{b \geq 0} \{X \leq a, Y \leq b\} \right] \\ &= \lim_{b \rightarrow \infty} \Pr[X \leq a, Y \leq b] \\ &= \lim_{b \rightarrow \infty} F(a, b) \end{aligned}$$

The same idea tells us that

$$F_Y(b) = \lim_{a \rightarrow \infty} F(a, b).$$

When working with continuous random variables, we often work with their *density functions*. Specifically, if  $X$  is a continuous random variable, there is some function  $f$  such that for (pretty much)<sup>1</sup> any set  $B \subseteq \mathbb{R}$ ,

$$\Pr[X \in B] = \int_B f(x) \, dx.$$

Here's the analogue for multiple variables.

**Definition 1.7.** Let  $X$  and  $Y$  be continuous random variables. We say  $X$  and  $Y$  have a *continuous joint distribution* if there is some function  $f : \mathbb{R}^2 \rightarrow [0, 1]$  such that

$$\Pr[(X, Y) \in C] = \int_C f(x, y) \, dydx.$$

In this case, we call  $f$  the *joint probability density function (pdf)* of  $X$  and  $Y$ .

In the discrete case we were able to start with a joint mass function and sum over one of the variables to obtain the marginal of the other variable. Here's the analogue for continuous random variables.

**Proposition 1.8.** Suppose  $X$  and  $Y$  are jointly continuous random variables with joint probability density function  $f$ . Then  $X$  and  $Y$  are continuous random variables with density functions

$$\begin{aligned} f_X(x) &= \int_{\mathbb{R}} f(x, y) \, dy \\ f_Y(y) &= \int_{\mathbb{R}} f(x, y) \, dx, \end{aligned}$$

respectively.

*Proof.* Suppose  $B \subseteq \mathbb{R}$  is measurable (don't worry too much about this assumption). Then

$$\begin{aligned} \Pr[X \in B] &= \Pr[X \in B, Y \in \mathbb{R}] \\ &= \int_B \left( \int_{\mathbb{R}} f(x, y) \, dy \right) dx. \end{aligned}$$

So the function

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy$$

plays the role of the density function for  $X$ . The same idea gives the density function for  $Y$ .  $\square$

**Example 1.9.** Let  $X$  and  $Y$  be random variables with joint pdf

$$f(x, y) = \begin{cases} kxy, & \text{if } x, y \geq 0, \, x + y \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

where  $k$  is some constant.

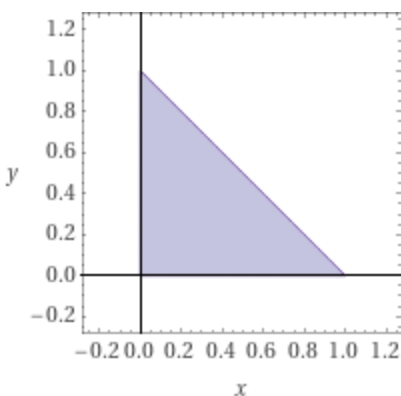
Let's determine the actual value of  $k$ . We must have the following

$$1 = \Pr[(X, Y) \in \mathbb{R}^2] = \int_{\mathbb{R}^2} f(x, y) \, dydx,$$

so we're just going to have to evaluate this integral and then solve for  $k$ . It's usually a good idea to draw the region in question when computing double integrals like this.

---

<sup>1</sup>Technically,  $B$  needs to be what's called a *measurable* set. Pretty much any set you'd care about is measurable, but we need this restriction for the theory to hold up.



As  $x$  ranges from 0 to 1,  $y$  ranges from 0 to  $1 - x$ . To see this, draw a vertical slice upwards from any point on the  $x$ -axis until it intersects the line  $x + y = 1$ . Our integral then becomes

$$\begin{aligned}
 k \int_0^1 \int_0^{1-x} xy \, dy \, dx &= k \int_0^1 x \left[ \frac{1}{2} y^2 \right]_{y=0}^{1-x} dx \\
 &= \frac{k}{2} \int_0^1 x(1-x)^2 \, dx \\
 &= \frac{k}{2} \left[ \frac{1}{4} x^4 - \frac{2}{3} x^3 + \frac{1}{2} x^2 \right]_{x=0}^{x=1} \\
 &= k/24.
 \end{aligned}$$

Since this expression must be equal to 1, we have  $k = 12$ .

Now let's compute the marginal pdf's of  $X$  and  $Y$ . To get the marginal for  $X$ , we “integrate out the  $y$ .” For any fixed  $x$ , the value of  $y$  ranges between 0 and  $1 - x$ , so we have

$$f_X(x) = \int_{\mathbb{R}} f(x, y) \, dy = \int_0^{1-x} kxy \, dy = \frac{k}{2} x(1-x)^2 = 12x(1-x)^2.$$

Similarly, for any fixed  $y$ , the value of  $x$  ranges between 0 and  $1 - y$ .

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) \, dx = \int_0^{1-y} kxy \, dx = \frac{k}{2} y(1-y)^2 = 12y(1-y)^2.$$

**Example 1.10.** The joint density function of  $X$  and  $Y$  is given by

$$f(x, y) = \begin{cases} 2e^{-x}e^{-2y}, & \text{if } 0 < x < \infty, 0 < y < \infty \\ 0, & \text{otherwise.} \end{cases}$$

Let's compute  $\Pr[X > 1, Y < 1]$ . To compute the probability of *any* event, we simply integrate the joint density function over that event. In this case we have

$$\begin{aligned}
 \Pr[X > 1, Y < 1] &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y} \, dx \, dy \\
 &= \int_0^1 2e^{-2y} [-e^{-x}]_{x=1}^{x=\infty} \, dy \\
 &= 2e^{-1} \int_0^1 e^{-2y} \, dy \\
 &= e^{-1}(1 - e^{-2}).
 \end{aligned}$$

In the case of a single random variable, the density and distribution functions are related by differentiation. That is, if  $X$  has density function  $f$  and distribution function  $F$ , then

$$f(x) = \frac{d}{dx}F(x).$$

The multivariable analogue is what you would probably expect. If  $X$  and  $Y$  are jointly continuous with density function  $f(x, y)$  and distribution function  $F(x, y)$ , then we have by Fubini's theorem

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^2}{\partial y \partial x} F(x, y). \quad (4)$$

We aren't going to angst over the proof here, but this technically only holds for the values of  $(x, y)$  where the partial derivatives are defined and continuous.

## 1.2 Independent Random Variables

Remember that we said that two *events*  $A$  and  $B$  are independent if

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B].$$

We can carry this definition over to random variables.

**Definition 1.11.** Let  $X$  and  $Y$  be random variables. Then  $X$  and  $Y$  are *independent* if for any measurable sets  $A$  and  $B$  we have

$$\Pr[X \in A, Y \in B] = \Pr[X \in A] \cdot \Pr[Y \in B].$$

Let's show that this definition plays nicely with the machinery we defined in the previous section.

**Proposition 1.12.** *The discrete random variables  $X$  and  $Y$ , taking values in  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively, are independent if and only if*

$$p(x, y) = p_X(x)p_Y(y) \quad (5)$$

for all  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ .

*Proof.* First let's suppose that  $X$  and  $Y$  are independent. Then we can consider the singleton sets  $\{x\}$  and  $\{y\}$  for any  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ .

$$p(x, y) = \Pr[X = x, Y = y] = \Pr[X = x] \Pr[Y = y] = p_X(x)p_Y(y).$$

Now suppose that equation (5) holds. Then for any sets  $A \subseteq \mathcal{X}$  and  $B \subseteq \mathcal{Y}$  we have

$$\begin{aligned} \Pr[X \in A, Y \in B] &= \sum_{x \in A, y \in B} p(x, y) \\ &= \sum_{x \in A} p_X(x) \sum_{y \in B} p_Y(y) \\ &= \Pr[X \in A] \Pr[Y \in B], \end{aligned}$$

so  $X$  and  $Y$  are independent. □

**Example 1.13.** Suppose we perform  $n + m$  independent trials, each having common success probability  $p$ . Let  $X$  be the number of successes in the first  $n$  trials and let  $Y$  be the number of successes in the next  $m$  trials. Are  $X$  and  $Y$  independent? Intuitively, knowing what happens in the first  $n$  trials shouldn't tell us anything about what happens in the next  $m$  trials, so we hope that  $X$  and  $Y$  are independent. Indeed, we have

$$\Pr[X = x, Y = y] = \binom{n}{x} p^x (1 - p)^{n-x} \binom{m}{y} p^y (1 - p)^{m-y} = \Pr[X = x] \Pr[Y = y].$$

Let's define a new random variable  $Z$  to be the total number of successes in all  $m + n$  trials. Are  $X$  and  $Z$  independent? Well if we know there are some successes in the first  $n$  trials, then we definitely know that there are at least that many successes in total, so we suspect that these aren't independent. We have

$$\begin{aligned} p(x, z) &= \Pr[x \text{ successes in the first } n \text{ trials, } z \text{ successes total}] \\ &= \Pr[x \text{ successes in the first } n \text{ trials, } z - x \text{ successes in the next } m \text{ trials}] \\ &= \binom{n}{x} p^x (1 - p)^{n-x} \binom{m}{z-x} p^{z-x} (1 - p)^{m-z+x}. \end{aligned}$$

However,

$$\begin{aligned} p_X(x) &= \binom{n}{x} p^x (1 - p)^{n-x} \\ p_Z(z) &= \binom{n+m}{z} p^z (1 - p)^{n+m-z}. \end{aligned}$$

It's easily seen that the product of these two quantities does not match up with the previous quantity, so  $X$  and  $Z$  are *not* independent.

So we can check to see if two discrete random variables are independent by looking at their mass functions. What about continuous random variables? We should think of a discrete random variable's mass function as being analogous to a continuous random variable's density function, and this informs the next proposition.

**Proposition 1.14.** *If  $X$  and  $Y$  are continuous random variables, then they are independent if and only if*

$$f(x, y) = f_X(x) f_Y(y).$$

*Proof.* Suppose  $X$  and  $Y$  are independent. This is really a statement about *distribution* functions, not density functions, so we have

$$\Pr[X \leq x, Y \leq y] = \Pr[X \leq x] \Pr[Y \leq y], \tag{6}$$

which is equivalent to

$$F(x, y) = F_X(x) F_Y(y).$$

Now let's take the mixed  $x$  and  $y$  partial derivatives of both sides to obtain

$$\frac{\partial^2}{\partial x \partial y} F(x, y) = \frac{\partial^2}{\partial x \partial y} (F_X(x) F_Y(y)) = \frac{\partial}{\partial x} F_X(x) \frac{\partial}{\partial y} F_Y(y).$$



We arrived at the last equality using the fact that  $F_X(x)$  is constant with respect to  $y$  and  $F_Y(y)$  is constant with respect to  $x$ . Since differentiation distributions gives us densities, we have

$$f(x, y) = f_X(x)f_Y(y)$$

as desired.

Conversely, suppose that equation (6) holds. We pretty much copy the proof of the previous proposition with integrals in place of sums. For any sets  $A$  and  $B$  we have

$$\begin{aligned}\Pr[X \in A, Y \in B] &= \int_A \int_B f(x, y) \, dy dx \\ &= \int_A f_X(x) \, dx \int_B f_Y(y) \, dy \\ &= \Pr[X \in A] \Pr[Y \in B].\end{aligned}$$

□

So we have have independence if our joint density (or mass) function factors into the product of the marginals. We can actually say more – factoring into *any* product of functions, each depending on just one variable, is enough.

**Proposition 1.15.** *Let  $X$  and  $Y$  be continuous random variables. Then  $X$  and  $Y$  are independent if and only if the joint density function can be factored as*

$$f(x, y) = h(x)g(y) \tag{7}$$

for some functions  $g$  and  $h$ .

*Proof.* If  $X$  and  $Y$  are independent, then the previous proposition tells us that we can just take  $g = f_X$  and  $h = f_Y$ .

Conversely, suppose that  $f(x, y) = h(x)g(y)$ . Then

$$\begin{aligned}1 &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) \, dx dy \\ &= \int_{\mathbb{R}} h(x) \, dx \cdot \int_{\mathbb{R}} g(y) \, dy.\end{aligned}$$

But these last two integrals have to each be equal to some constants,  $c_1$  and  $c_2$ , respectively. We also have

$$\begin{aligned}f_X(x) &= \int_{\mathbb{R}} f(x, y) \, dy \\ &= \int_{\mathbb{R}} h(x)g(y) \, dy \\ &= c_1 h(x).\end{aligned}$$

Similarly,  $f_Y(y) = c_2 g(y)$ . Putting it all together, we have

$$\begin{aligned}f_X(x)f_Y(y) &= c_1 h(x) \cdot c_2 g(y) \\ &= (c_1 c_2) h(x) g(y) \\ &= f(x, y).\end{aligned}$$

□

**Example 1.16.**

$$f(x, y) = \begin{cases} kxy, & \text{if } (x, y) \in [0, 1]^2 \\ 0, & \text{otherwise.} \end{cases}$$

Factors

**Example 1.17.**

$$f(x, y) = \begin{cases} kxy, & \text{if } x, y \geq 0, x + y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Doesn't look like we can factor it – check using marginals.

## References

- [1] Ross, Sheldon. *A First Course in Probability*. Ninth Edition. Pearson. 2014.