Jointly distributed random variables

Chrysafis Vogiatzis

Lecture 11

Learning objectives

After these lectures, we will be able to:

- Describe and recognize jointly distributed random variables.
- Define joint, marginal, and conditional probability mass functions for discrete random variables.
- Define joint, marginal, and conditional probability distribution functions for continuous random variables.
- Use joint, marginal, and conditional probability mass and distribution functions to calculate probabilities.

Motivation: "Can you hear me now'?'

Not all of us pay attention all the time in a Zoom call; sometimes it is our fault (we are distracted or busy), but others it is not (technical difficulties, bad reception). So the question becomes: how many times does something need to be repeated before you hear it? Note that it does not only depend on whether you are paying attention (which is a random variable), but also on whether you are having a clear connection (another random variable).

Jointly distributed random variables

Real life and its outcomes can be viewed as a combination of random events, rather than a single random event. Succeeding in an exam has many factors that do not rely on only your preparation: you need to be healthy and well-rested, you need to be focused during the exam, you need to have luck at your side, you need to have a calculator whose batteries are still working. And even when all of these things align, you also need to be there on time, which means that you need to catch a bus, that there is no construction causing traffic jams, etc. We can go on like this *a lot*.

The truth is that in this class we have focused on single random variables that are distributed their own way. What about the case where two random variables are distributed alongside each other?

But, wait? Did we not discuss the probability of two events happening hand-in-hand during the first lectures of the course? And did we not discuss *specifically what happens if those two events are independent*[1] *or not?*

You are correct. We have discussed what happens when two events are happening at the same time. We also discussed what the probability is that an event happens given another event happening. However, there are two caveats in our discussion earlier:

1. We only focused on discrete (countable) events: it is time we see what this implies in the continuous space too.

2. We saw this in terms of events and sets. We are now going to have that discussion in terms of distributions, probability mass/density functions.

[1] Recall independence: it implies that knowledge of one event happening does not affect our probabilities of the other event happening.

Definition

We begin with the definition of jointly distributed random variables.

Definition 1 (Jointly distributed random variables) *Let X and Y be two random variables. The probability distribution that defines their simultaneous behavior is referred to as a* ***joint probability distribution****. The two random variables X and Y are then called* ***jointly distributed random variables****.*

Examples of jointly distributed random variables

Here are some examples of jointly distributed random variables.

- The times you have to repeat yourself on the phone and your signal reception.
- The grade you receive in an exam and the amount of sleep you've had the night before.
- The performance of two or more stocks in your portfolio.
- The box office of a movie and the critical reception.

We observe here that jointly distributed does not imply immediate effect. For example, a student could get a very high grade in an exam, even if they slept very little the night before; or a movie could make a lot of money in the box office, despite being universally hated by reviewers. However, jointly distributed random variables imply that what we see is a combination of random variables rather than outcome of a single random variable.

Are the following better modeled as a single random variable or as jointly distributed random variables?

- Getting a higher grade in an exam than the person sitting next to you?
- Throwing a die?
- Throwing two dies and having the first die land on a higher number than the second one?

An example

Securing a position after college might require some effort.. If the economy is doing well ("is good"), then a student could get more job interview invitations, and consequently there are more chances for a job opportunity. If the economy is average, or if it is outright bad, then a student may struggle to get interviews and/or a job..

Based on our definitions, the state of the economy is a random variable. The same can be said about the number of job interviews that a student gets invited to. In the end of the day, the number of interviews that a student needs to go on before they secure a position after graduation is a **jointly distributed random variable**. Let's assume that the probabilities are as given in Table 1.

Table 1: Number of job interviews required to get a job depending on the state of the economy.

	Y=state of the economy		
X=job interviews to get a job	Bad	Average	Good
1	0.01	0.05	0.20
2	0.03	0.05	0.18
3	0.03	0.12	0.08
≥ 4	0.08	0.12	0.05

Jointly distributed discrete random variables

If X and Y are discrete random variables, then (X, Y) is called a jointly discrete bivariate random variable.

Definition 2 (Joint probability mass function) *The* ***joint probability mass function*** *is defined as:*

$$f_{XY}(x, y) = P(X = x, Y = y).$$

It follows the next three properties:

1. $f_{XY}(x,y) \geq 0, \forall x, y$.
2. $\sum_x \sum_y f_{XY}(x,y) = 1$.
3. $P((X,Y) \in A) = \sum\sum_{(x,y)\in A} f_{XY}(x,y)$.

A couple of quick notes about the notation here. You will observe that the joint probability mass function is given by $f_{XY}(x,y)$. We had previously reserved $f(\cdot)$ for continuous random variables, keeping $p(\cdot)$ for discrete ones. For convenience, we only use $f(\cdot)$ for joint probability distributions.

Additionally, we notice that there is a subscript in the function. The subscript is supposed to reveal which random variables the function is including. For example $f_{XY}(3,4)$ would imply that the function is considering random variables X and Y and is asking for them to be equal to 3 and 4, respectively.

Note that this definition is easily generalized for more than two variables: if X_i are discrete random variables for $i = 1, \ldots, n$, then $(X_1, \ldots, X_n)$ is called a **jointly distributed discrete multivariate random variable** with joint pmf:

$$f_{X_1 X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n).$$

Following the same notation as before, we see from the subscript of the function that this distribution contains random variables $X_1, X_2, \ldots, X_n$.

Getting a job after college

Let's see: do the probabilities provided in the example earlier satisfy the first two properties?

1. $f_{XY}(x,y) \geq 0, \forall x, y$. This is true, as all entries for the 12 cases are all positive.
2. $\sum_x \sum_y f_{XY}(x,y) = 0.01 + 0.05 + 0.20 + \ldots + 0.05 = 1$. This is also true.

Let's dwell a little on the third property now.

Getting a job after college

What is the probability that:

1. a student gets a job in 1 interview and that the economy is good?

	Y		
X	Bad	Average	Good
1	0.01	0.05	**0.20**
2	0.03	0.05	0.18
3	0.03	0.12	0.08
≥ 4	0.08	0.12	0.05

The probability is 20%.

2. a student gets a job in less than or equal to 3 interviews and the economy is average?

	Y		
X	Bad	Average	Good
1	0.01	**0.05**	0.20
2	0.03	**0.05**	0.18
3	0.03	**0.12**	0.08
≥ 4	0.08	0.12	0.05

The probability is 22%.

3. a student gets a job in more than 3 interviews?

4. the economy is good?

5. a student gets a job in 1 interview if we know that the economy is good?

Questions 3, 4, and 5 seem to require a little different logic. Could we add all the outcomes that include the specific clause we are after? For example, could we simply add all the probabilities of a good economy and say that this is the probability that the economy is good? But for the last one, we know that the economy is good. How can we use this fact to calculate the required probability? Could we use conditional probabilities?

Marginal probability mass function

Definition 3 (Marginal probability mass function) *The* ***marginal probability mass function*** *(marginal pmf) of a discrete random variable is computed by summing over all possible values of the other random variable. For two random variables, X and Y:*

1. *The marginal distribution of X:*

$$f_X(x) = P(X = x) = \sum_y f_{XY}(x, y)$$

2. *The marginal distribution of Y:*

$$f_Y(y) = P(Y = y) = \sum_x f_{XY}(x, y)$$

The marginal distribution of a random variable answers the question: "what is the probability that X takes a certain value, regardless of Y?" [2]

Going back to the motivation from earlier:

[2] And vice versa for the marginal distribution of Y.

Getting a job after college

What is the probability that:

1. a student gets a job in 1 interview and that the economy is good? The probability is 20%.

2. a student gets a job in less than or equal to 3 interviews and the economy is average. The probability is 22%.

3. a student gets a job in more than 3 interviews?

   We are after the probability of $P(X > 3)$. Based on the definition of marginal distributions, we have that:

$$P(X = x) = \sum_y f_{XY}(x, y) \implies$$

$$\implies P(X > 3) = P(X \geq 4) = 0.08 + 0.12 + 0.05 = 0.25.$$

	Y		
X	Bad	Average	Good
1	0.01	0.05	0.20
2	0.03	0.05	0.18
3	0.03	0.12	0.08
≥ 4	**0.08**	**0.12**	**0.05**

The probability is 25%.

Getting a job after college (cont'd)

4. the economy is good?

   We are after the probability of $P(Y = \text{Good})$. Following a similar logic:

$$P(Y = y) = \sum_x f_{XY}(x, y) \implies$$

$$\implies P(Y = \text{Good}) = 0.20 + 0.18 + 0.08 + 0.05 = 0.51.$$

	Y		
X	Bad	Average	Good
1	0.01	0.05	**0.20**
2	0.03	0.05	**0.18**
3	0.03	0.12	**0.08**
≥4	0.08	0.12	**0.05**

The probability is 51%.

5. a student gets a job in 1 interview if we know that the economy is good?

For calculating marginal distributions in discrete random events given in tabular format, we may also add up the probabilities in the columns and rows and obtain:

	Y=state of the economy			
X=job interviews to get a job	Bad	Average	Good	$f_X(x)$
1	0.01	0.05	0.20	0.26
2	0.03	0.05	0.18	0.26
3	0.03	0.12	0.08	0.23
≥4	0.08	0.12	0.05	0.25
$f_Y(y)$	0.15	0.34	0.51	1

Here the columns are showing the probability of the state of the economy (alone) which are bad with 15%, average with 34%, and good with 51%, whereas the rows are showing the number of interview (26% for 1, 26% for 2, 23% for 3, 25% for more than 3).

Conditional probability mass function

Definition 4 (Conditional probability mass function) *The **conditional probability mass function** (conditional pmf) of a discrete random variable given values for the other ones is computed by dividing the joint pmf of all over the marginal pmf of the others. For two random variables, X and Y:*

1. *The conditional distribution of X given $Y = y$:*

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

2. *The conditional distribution of Y given $X = x$:*

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

Of course, for the conditional pmf to make sense, we need that $f_X(x) > 0$ and $f_Y(y) > 0$.

One more note of notation. Observe how the subscript has changed to reflect the fact that we know what Y or X is. We write:

$$f_{X|Y=y} = f_{X|y},$$

which is read as the "conditional pmf of random variable X given that random variable Y is equal to y" or, simply the "conditional pmf of random variable X given y."

Let us revisit our motivation.

Getting a job after college

What is the probability that:

1. a student gets a job in 1 interview and that the economy is good? The probability is 20%.
2. a student gets a job in less than or equal to 3 interviews and the economy is average. The probability is 22%.
3. a student gets a job in more than 3 interviews? The probability is 25%.
4. the economy is good? The probability is 51%.
5. a student gets a job in 1 interview if we know that the economy is good?

   This is the definition of a conditional probability. Specifically, we want to calculate $P(X = 1|Y = \text{Good})$.

$$P(X = 1|Y = \text{Good}) = \frac{f_{XY}(1, \text{Good})}{f_Y(\text{Good})} = \frac{0.2}{0.51} = 0.3922.$$

	Y			
X	Bad	Average	Good	$f_X(x)$
1	0.01	0.05	**0.20**	0.26
2	0.03	0.05	0.18	0.26
3	0.03	0.12	0.08	0.23
≥4	0.08	0.12	0.05	0.25
$f_Y(y)$	0.15	0.34	**0.51**	1

One full example

As interesting as this example has been, the truth is that in many cases we cannot enumerate easily all cases. In those instances, we turn to calculus. Let us see a similar case:

Jointly distributed discrete random variables

Two discrete random variables X and Y have a joint distribution of $f_{XY}(x,y) = \frac{x+y+1}{c}$, for x and y equal to 0, 1, or 2.

1. What should c be?
2. What is $P(X \leq 1, Y = 1)$?
3. What is $P(Y = 1)$?
4. What is $P(X \leq 1 | Y = 1)$?

For calculating c, we need to use the second property.

Getting the joint pmf

$$\begin{aligned}
&\sum_{x=0}^{2}\sum_{y=0}^{2} f_{XY}(x,y) = \sum_{x=0}^{2}\sum_{y=0}^{2} \frac{x+y+1}{c} = 1 \implies \\
&\implies \frac{1}{c}+\frac{2}{c}+\frac{3}{c}+\frac{2}{c}+\frac{3}{c}+\frac{4}{c}+\frac{3}{c}+\frac{4}{c}+\frac{5}{c} = 1 \\
&\implies c = 27.
\end{aligned}$$

With the full joint pmf, we can answer the remaining questions:

Using the joint pmf

$$P(X \leq 1, Y = 1) = \sum_{x=0}^{1} f_{XY}(x,1) = \sum_{x=0}^{2} \frac{x+2}{27} = \frac{2}{27} + \frac{3}{27} = \frac{5}{27}.$$

We now move our focus to the marginal distribution, which can be found as:

Computing and using the marginal pmf

$$\begin{aligned}
&f_Y(y) = P(Y = y) = \sum_{x=0}^{2} f_{XY}(x,y) = \\
&= \frac{0+y+1}{27} + \frac{1+y+1}{27} + \frac{2+y+1}{27} \implies f_Y(y) = \frac{3y+6}{27}.
\end{aligned}$$

Hence $P(Y = 1) = f_Y(1) = \frac{9}{27} = \frac{1}{3}$.

To conclude this, we may combine the joint and marginal pmf to get the conditional pmf:

Computing and using the conditional pmf

$$f_{X|Y=y}(x) = P(X = x|Y = y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{\frac{x+y+1}{27}}{\frac{3y+6}{27}} = \frac{x+y+1}{3y+6}$$

Finally: $P(X \leq 1|Y = 1) = f_{X|Y=1}(0) + f_{X|Y=1}(1) = \frac{2}{9} + \frac{3}{9} = \frac{5}{9}$.

Jointly distributed continuous random variables

If X and Y are continuous random variables, then (X, Y) is called a jointly continuous bivariate random variable.

Definition 5 (Joint probability distribution function) *The* ***joint probability distribution function*** *is defined as:*

$$f_{XY}(x,y).$$

Like in the simple continuous random variables, f_{XY} reveals a relative likelihood rather than a probability value. It also follows three properties:

1. $f_{XY}(x,y) \geq 0, \forall x, y$.
2. $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = 1$.
3. $P((X,Y) \subset R) = \iint_R f_{XY}(x,y)dxdy$.

Once again, the definitions is easy to generalize to more than two variables: if X_i are continuous random variables for $i = 1, \ldots, n$, then $(X_1, \ldots, X_n)$ is called a **jointly distributed continuous multivariate random variable** with joint pdf:

$$f_{X_1X_2\ldots X_n}(x_1, x_2, \ldots, x_n).$$

In the end of the section, we given an example with more than 2 random variables so that you can practice with it – it will come in hand during Lecture 13.

A chemical mixture

A product is a mixture of two materials: let the volume of material 1 used be represented as X, and the volume of material 2 used be represented as Y. The joint probability density function of the two random variables is

$$f_{XY}(x,y) = c\,(2x+3y)\,, \quad 0 \leq x \leq 1, 0 \leq y \leq 1.$$

1. What is c?

2. What is the probability the first material has volume less than or equal to 0.5, and the second material has volume between 0.25 and 0.5?

Similarly to what we did for discrete random variables (with the main difference that we now need to integrate over the values that X and Y can take), we get:

Computing the joint pdf

1. From the second property of joint pdfs for continuous random variables, we have:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = 1 \implies \int_{0}^{1}\int_{0}^{1} c\,(2x+3y)\,dxdy = 1 \implies$$

$$\implies c\int_{0}^{1}\left(x^2+3xy\right)\Big|_0^1 dy = 1 \implies c\int_{0}^{1}(3y+1)dy = 1 \implies$$

$$\implies c\left(3\frac{y^2}{2}+y\right)\Bigg|_0^1 = 1 \implies c\frac{5}{2} = 1 \implies c = \frac{2}{5}.$$

Using the joint pdf

2. Knowing that $f_{XY}(x,y) = \frac{2}{5}(2x+3y)$, we may calculate the probability $P(X \leq 0.5, 0.25 \leq Y \leq 0.5)$ as follows. Recall that we are talking about continuous random variables, so we will always integrate!

$$P(X \leq 0.5, 0.25 \leq Y \leq 0.5) = \int_{0}^{0.5}\int_{0.25}^{0.5} f_{XY}(x,y)dydx =$$

$$= \int_{0}^{0.5}\int_{0.25}^{0.5} \frac{2}{5}(2x+3y)\,dydx = \frac{2}{5}\int_{0}^{0.5}\left(2xy + 3\frac{y^2}{2}\right)\Big|_{0.25}^{0.5} dx =$$

$$= \frac{2}{5}\int_{0}^{0.5}\left(0.5x + \frac{9}{32}\right)dx = \frac{2}{5}\left(0.5\frac{x^2}{2} + \frac{9}{32}x\right)\Big|_{0}^{0.5} = \frac{13}{160}$$

Marginal probability distribution function

Definition 6 (Marginal probability distribution function) *The **marginal probability distribution function** (marginal pdf) of a continuous random variable is computed by integrating over all possible values of the other random variable. For two random variables, X and Y:*

1. *The marginal distribution of X:*

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

2. *The marginal distribution of Y:*

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx$$

Let us return to our chemical mixture:

Computing and using the marginal pdf

As a reminder, we have $f_{XY}(x,y) = \frac{2}{5}(2x+3y)$ for the joint pdf of two continuous random variables X and Y (volume of material 1 and 2, respectively) taking values between 0 and 1. What is the probability that:

1. the volume of material 1 is less than 0.5?

2. the volume of material 2 is between 0.25 and 0.5?

Computing and using the marginal pdf

1. First, we calculate $f_X(x)$. It is:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy = \int_0^1 \frac{2}{5}(2x+3y)\,dy = \frac{1}{5}(4x+3)\,.$$

   We can now use $f_X(x)$:

$$P(X \leq 0.5) = \int_0^{0.5} f_X(x)dx = \int_0^{0.5} \frac{1}{5}(4x+3)\,dx = 0.4.$$

2. Then, we do the same for $f_Y(y)$:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx = \int_0^1 \frac{2}{5}(2x+3y)\,dx = \frac{1}{5}(6y+2)\,.$$

   And we finish the question by calculating the proper integral:

$$P(0.25 \leq Y \leq 0.5) = \int_{0.25}^{0.5} f_Y(y)dy = \int_{0.25}^{0.5} \frac{1}{5}(6y+2)\,dy = \frac{17}{80}.$$

Conditional probability distribution function

Definition 7 (Conditional probability distribution function) *The **conditional probability distribution function** (conditional pdf) of a continuous random variable given values for the other ones is computed by dividing the joint pdf of all over the marginal pdf of the others. For two random variables, X and Y:*

1. *The conditional distribution of X given $Y = y$:*

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

2. *The conditional distribution of Y given $X = x$:*

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

Once again, the conditional pdf is only defined when $f_X(x) > 0$ and $f_Y(y) > 0$.

Let's go back to our chemical mixture example.

Computing and using the conditional pdf

What is the probability the first material has a proportion less than or equal to 50%, given that the second material has a proportion equal to 30%?

To contrast this with the following question, we shall name this the conditional distribution route.

The conditional distribution route. First, we calculate $f_{X|Y=y}(x)$ as:

$$f_{X|Y=y}(x) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{\frac{2}{5}(2x+3y)}{\frac{1}{5}(6y+2)} = \frac{4x+6y}{6y+2}.$$

Replacing $Y = 0.3$ as is known, we get:

$$f_{X|Y=0.3}(x) = \frac{4x+1.8}{3.8}.$$

Finally, we may calculate $P(X \leq 0.5|Y = 0.3)$ as follows:

$$P(X \leq 0.5|Y = 0.3) = \int_0^{0.5} \frac{4x+1.8}{3.8} dx = 0.3684.$$

And here is one more conditional to practice basic probability theory:

Calculating conditional probabilities

What is the probability the first material has a proportion less than or equal to 50%, given that the second material has a proportion between 25% and 50%?

The basic probability theory route. Remember that $P(A|B) = P(A \cap B)/P(B)$. In our case, we have already calculated $P(A \cap B) = P(X \leq 0.5, 0.25 \leq Y \leq 0.5) = \frac{13}{160}$ and $P(B) = P(0.15 \leq Y \leq 0.5) = \frac{17}{80}$. Combining:

$$P(X \leq 0.5|0.25 \leq Y \leq 0.5) = \frac{13/160}{17/80} = \frac{13}{34}.$$

A multivariate example

Here, we provide a small example for a joint distribution with 4 random variables. More specifically:

A 4-component machine

Suppose that a machine consists of four components, whose lifetimes (in years) are jointly distributed with the following pdf:

$$f_{X_1X_2X_3X_4}(x_1,x_2,x_3,x_4) = c \cdot e^{-2x_1}e^{-x_2}e^{-3x_3}e^{-0.5x_4}.$$

- What should c be for this to be a valid pdf?
- What is the probability the first component survives for more than one year?

For the first question, we want

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X_1X_2X_3X_4}(x_1,x_2,x_3,x_4)dx_4dx_3dx_2dx_1 = 1 \implies$$

$$\implies \int_{0}^{+\infty}\int_{0}^{+\infty}\int_{0}^{+\infty}\int_{0}^{+\infty} f_{X_1X_2X_3X_4}(x_1,x_2,x_3,x_4)dx_4dx_3dx_2dx_1 = 1 \implies$$

$$\implies c\int_{0}^{+\infty}\int_{0}^{+\infty}\int_{0}^{+\infty}\int_{0}^{+\infty} e^{-2x_1}e^{-x_2}e^{-3x_3}e^{-0.5x_4}dx_4dx_3dx_2dx_1 = 1 \implies$$

$$\implies \frac{c}{0.5}\int_{0}^{+\infty}\int_{0}^{+\infty}\int_{0}^{+\infty} e^{-2x_1}e^{-x_2}e^{-3x_3}dx_3dx_2dx_1 = 1 \implies$$

$$\implies \frac{c}{0.5\cdot 3}\int_{0}^{+\infty}\int_{0}^{+\infty} e^{-2x_1}e^{-x_2}dx_2dx_1 = 1 \implies$$

$$\implies \frac{c}{0.5\cdot 3\cdot 1}\int_{0}^{+\infty} e^{-2x_1}dx_1 = 1 \implies \frac{c}{0.5\cdot 3\cdot 1\cdot 2} = 1 \implies c = 3.$$

For the second question, first calculate the marginal pdf $f_{X_1}(x_1)$:

$$f_{X_1}(x_1) = \int_{0}^{+\infty}\int_{0}^{+\infty}\int_{0}^{+\infty} f_{X_1X_2X_3X_4}(x_1,x_2,x_3,x_4)dx_4dx_3dx_2 = 2e^{-2x_1}.$$

Then, the probability it survives for more than one year is

$$P(X_1 > 1) = 1 - P(X_1 \le 1) = 1 - \int_{0}^{1} 2e^{-2x_1}dx_1 = 0.1353.$$

Review

A very brief summary of today's lecture follows. For two random variables X and Y that are *jointly* distributed, we have the following:

Joint pmf/pdf

- TL;DR: How are both variables distributed as *simultaneously*?
  - **Discrete**: what is $P(X = x \text{ and } Y = y)$?
  - **Continuous**: what is the relative likelihood of X having the value of x and Y getting the value of y?
- Denoted by $f_{XY}(x, y)$.
- Follows three properties:
  - **Discrete**:
    1. $f_{XY}(x, y) \geq 0$.
    2. $\sum_x \sum_y f_{XY}(x, y) = 1$.
    3. $P((X, Y) \in A) = \sum\sum_{(x,y)\in A} f_{XY}(x, y)$.
  - **Continuous**:
    1. $f_{XY}(x, y) \geq 0, \forall x, y$.
    2. $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dxdy = 1$.
    3. $P((X, Y) \subset R) = \int\int_R f_{XY}(x, y) dxdy$.

Marginal pmf/pdf

- TL;DR: I am only interested in one of the random variables.
  - **Discrete**: Let's lose Y. What $P(X = x)$?
  - **Continuous**: Let's lose Y. What the relative likelihood of X getting the value of x?
- Denoted by $f_X(x)$ or $f_Y(y)$.

Conditional pmf/pdf

- TL;DR: I am given information on one of the random variables.
  - **Discrete**: I know what Y is! It is equal to y. What is $P(X = x|Y = y)$?
  - **Continuous**: I know what Y is! It is equal to y. What is the relative likelihood of X taking value x now?
- Denoted by $f_{X|y}(x)$ or $f_{Y|x}(y)$.

Recall that all definitions shown here for both discrete and continuous random variables may be extended to more than 2 random variables $X_1, X_2, \ldots$