STAT 305: Chapter 5

Part IV

Amin Shirazi

ashirazist.github.io/stat305.github.io

Chapter 5.4: Joint Distributions and Independence

Working with Multiple Random Variables

Joint Distributions

We often need to consider two random variables together - for instance, we may consider

- the length and weight of a squirrel,
- the loudness and clarity of a speaker,
- the blood concentration of Protein A, B, and C and so on.

This means that we need a way to describe the probability of two variables *jointly*. We call the way the probability is simultaneously assigned the "joint distribution".

Joint Distributions

Discrete RVs

Joint distribution of discrete random variables

For several discrete random variable, the device typically used to specify probabilities is a *joint probability function*. The two-variable version of this is defined.

> A **joint probability function (joint pmf)** for discrete random variables X and Y is a nonnegative function $f(x, y)$, giving the probability that (simultaneously) X takes the values x and Y takes the values y. That is,
>
> $$f(x, y) = P[X = x \text{ and } Y = y]$$

Properties of a valid joint pmf:

- $f(x, y) \in [0, 1]$ for all x, y or $f(x,y) \geq 0 \ \forall x, y$
- $\boxed{\sum_{x,y} f(x. y) = 1}$

similarly: $\int_x \int_y f(x,y)\,dy\,dx = 1$

Joint distribution of discrete random variables

So we have probability functions for X, probability functions for Y and now a probability function for X and Y together - that's a lot of fs floating around though! In order to be clear which function we refer to when we refer to " f ", we also add some subscripts

Suppose X and Y are two discrete random variables.

- we may need to identify the *joint probability function* using $f_{XY}(x, y)$,
- we may need to identify the probability function of X by itself (aka the *marginal probability function* for X) using $f_X(x)$,
- we may need to identify the probability function of Y by itself (aka the *marginal probability function* for Y) using $f_Y(y)$

$f_x(x) = \sum_y f_{x,y}(x,y)$

continuous r.v: $f_y(y) = \int_x f_{x,y}(x,y)\, dx$

Joint pmf

For the discrete case, it is useful to give $f(x, y)$ in a **table**.

Two bolt torques, cont'd

Recall the example of measure the bolt torques on the face plates of a heavy equipment component to the nearest integer. With

$$X = \text{the next torque recorded for bolt } 3$$
$$Y = \text{the next torque recorded for bolt } 4$$

Joint pmf

the joint probability function, $f(x, y)$, is

X $\rightarrow$ x = 14

Y $\downarrow$ y = 19

y\x	11	12	13	14	15	16	17	18	19	20
20	0	0	0	0	0	0	0	2/34	2/34	1/34
19	0	0	0	0	0	0	2/34	0	0	0
18	0	0	1/34	1/34	0	0	1/34	1/34	1/34	0
17	0	0	0	0	2/34	1/34	1/34	2/34	0	0
16	0	0	0	1/34	2/34	2/34	0	0	2/34	0
15	1/34	1/34	0	0	3/34	0	0	0	0	0
14	0	0	0	0	1/34	0	0	2/34	0	0
13	0	0	0	0	1/34	0	0	0	0	0

Joint Distributions

Discrete RVs

Calculate:

- $P[X = 14 \text{ and } Y = 19]$ $= 0$

- $P[X = 18 \text{ and } Y = 17]$ $= 2/34$

Joint Distributions

Discrete RVs

By summing up certain values of $f(x, y)$, probabilities associated with X and Y with patterns of interest can be obtained.

Consider: $P(X \geq Y)$

y\x	11	12	13	14	15	16	17	18	19	20
20										
19										
18										
17										
16										
15										
14										
13										

Joint Distributions

Discrete RVs

$P(|X - Y| \leq 1)$

y\x	11	12	13	14	15	16	17	18	19	20
20										
19										
18										
17										
16										
15										
14										
13										

Discrete RVs

$P(X = 17)$

y\x	11	12	13	14	15	16	17	18	19	20
20										
19										
18										
17										
16										
15										
14										
13										

Marginal Distribution

Marginal distributions

In a bivariate problem, one can add down columns in the (two-way) table of $f(x, y)$ to get values for the probability function of X, $f_X(x)$ and across rows in the same table to get values for the probability distribution of Y, $f_Y(y)$.

The individual probability functions for discrete random variables X and Y with joint probability function $f(x, y)$ are called **marginal probability functions**. They are obtained by summing $f(x, y)$ values over all possible values of the other variable.

Connecting Joint and Marginal Distributions

Discrete RVs

Use: Joint to Marginal for Discrete RVs

Let X and Y be discrete random variables with joint probability function Then the marginal probability function for X can be found by:

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

and the marginal probability function for Y can be found by:

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

`

Example: [Torques, cont'd]

Find the marginal probability functions for X and Y from the following joint pmf.

y\x	11	12	13	14	15	16	17	18	19	20
20	0	0	0	0	0	0	0	2/34	2/34	1/34
19	0	0	0	0	0	0	2/34	0	0	0
18	0	0	1/34	1/34	0	0	1/34	1/34	1/34	0
17	0	0	0	0	2/34	1/34	1/34	2/34	0	0
16	0	0	0	1/34	2/34	2/34	0	0	2/34	0
15	1/34	1/34	0	0	3/34	0	0	0	0	0
14	0	0	0	0	1/34	0	0	2/34	0	0
13	0	0	0	0	1/34	0	0	0	0	0

Getting marginal probability functions from joint probability functions begs the question whether the process can be reversed.

> **Can we find joint probability functions from marginal probability functions?**

Conditional Distribution

Joint Distributions

Discrete RVs

Conditional Distribution

Conditional Distribution of Discrete Random Variables

When working with several random variables, it is often useful to think about what is expected of one of the variables, given the values assumed by all others.

> For discrete random variables X and Y with joint probability function $f(x, y)$, the **conditional probability function of X given $Y = y$** is *a function of x*
>
> $$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f(x,y)}{\sum_x f(x,y)}$$
>
> and the **conditional probability function of Y given $X = x$** is *a function of y*
>
> $$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{f(x,y)}{\sum_y f(x,y)}.$$

Example: [Torque, cont'd]

y\x	11	12	13	14	15	16	17	18	19	20
20	0	0	0	0	0	0	0	2/34	2/34	1/34
19	0	0	0	0	0	0	2/34	0	0	0
18	0	0	1/34	1/34	0	0	1/34	1/34	1/34	0
17	0	0	0	0	2/34	1/34	1/34	2/34	0	0
16	0	0	0	1/34	2/34	2/34	0	0	2/34	0
15	1/34	1/34	0	0	3/34	0	0	0	0	0
14	0	0	0	0	1/34	0	0	2/34	0	0
13	0	0	0	0	1/34	0	0	0	0	0

Find the following probabilities:

- $f_{Y|X}(20|18)$

Joint Distributions

Discrete RVs

Conditional Distribution

Example: [Torque, cont'd]

- $f_{Y|X}(y|15)$

- $f_{Y|X}(y|20)$

- $f_{X|Y}(x|18)$

Independence

Joint Distributions

Discrete RVs

Conditional Distribution

Independence

Let's start with an example. Look at the following joint probability distribution and the associated marginal probabilities.

$y \backslash x$	1	2	3	$f_Y(y)$
3	0.08	0.08	0.04	0.20
2	0.16	0.16	0.08	0.40
1	0.16	0.16	0.08	0.40
$f_X(x)$	0.40	0.40	0.20	1.00

What do you notice?

Joint Distributions

Discrete RVs

Conditional Distribution

Independence

Discrete random variables X and Y are **independent** if their joint distribution function $f(x, y)$ is the product of their respective marginal probability functions. This is,

> independence means that
>
> $$f(x, y) = f_X(x) f_Y(y) \qquad \text{for all } x, y.$$
>
> If this does not hold, then X and Y are **dependent**

Alternatively, discrete random variables X and Y are independent if for all x and y,

If X and$Y$ are not only independent but also have the same marginal distribution, then they are **independent and identically distributed (iid)**.