STAT 305: Chapter 5

Part IV

Amin Shirazi

Course page:
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 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
- $\sum_{x,y} f(x. y) = 1$

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 $f_Y(y)$

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

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

Discrete RVs

Calculate:

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

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

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

Discrete RVs

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:

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

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)**.

Chapter 5.5: Functions of Random Variables

Results and Theorems

Functions of Random Variables

A random variable can be thought of as a function whose input is an outcome and whose output is a real number. When we take a function of the value the random variable takes, the resulting value is still depends on the outcome of a random experiment - in other words: functions of random variables are random variables.

This means that a function of a random variable will have probabilities attached to the value it takes, based on the value taken by the random variable. It also means functions of random variables will have:

> - probability functions (if discrete) or probability density functions (if continuous)
> - cumulative probability functions (if discrete) or cumulative density functions (if continuous)
> - expected values and variances …

Linear combinations

For engineering purposes, it often suffices to know the mean and variance for a function of several random variables, $U = g(X_1, X_2, \dots, X_n)$ (as opposed to knowing the whole distribution of U). When g is **linear**, there are explicit functions.

Proposition: If $X_1, X_2, \dots, X_n$ are n **independent** random variables and $a_0, a_1, \dots, a_n$ are $n+1$ constants, then consider

$$U = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

> U is itself a random variable as it is a linear combination of n *independent* random variables

Linear combinations [cont'd]

U, as a random variable has mean

> ‘$\mathrm{E}U = a_0 + a_1\mathrm{E}X_1 + a_2\mathrm{E}X_2 + \cdots + a_n\mathrm{E}X_3$

and variance

> ‘$\mathrm{Var}U = a_1^2\mathrm{Var}X_1 + a_2^2\mathrm{Var}X_2 + \cdots + a_n^2\mathrm{Var}X_3$

Linear Combinations

Example:

Say we have two independent random variables X and Y with $\mathrm{E}X = 3.3, \mathrm{Var}X = 1.91, \mathrm{E}Y = 25$, and $\mathrm{Var}Y = 65$. Find the mean and variance for

- U = 3 + 2X - 3Y

- V = -4X + 3Y

Functions of RVs

Linear Combinations

Example:

Say $X \sim Binomial(n = 10, p = 0.5)$ and $Y \sim Poisson(\lambda = 3)$. Calculate the mean and variance of $Z = 5 + 2X - 7Y$.

Functions of RVs

Linear Combinations

Sample Mean

A particularly important use of functions of random variables concerns n iid random variables where each $a_i = \frac{1}{n}$ for $i = 1, 2, \cdots, n$. Then we can define the random variable $\overline{X}$ as follows

$$\overline{X} = \frac{1}{n}X_1 + \cdots \frac{1}{n}X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

> Note that $\overline{X}$ is a random variable

We can then find the mean and variance of this random variable.

Functions of RVs

Linear Combinations

Sample Mean

$$\overline{X} = \frac{1}{n}X_1 + \cdots \frac{1}{n}X_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

as they relate to the population parameters $\mu = \mathrm{E}X_i$ and $\sigma^2 = \mathrm{Var}X_i$.

For **independent** variables $X_1, \ldots, X_n$ with common mean μ and variance σ^2,

$\mathrm{E}(\overline{X})$:

$\mathrm{V}(\overline{X})$:

Functions of RVs

Linear Combinations

Sample Mean

What is the point?

It does not matter if we are working with discrete or continuous random variables, as long as we have an independent and identically distributed (iid) sample of size n with the same mean μ and the same variance σ^2, the random variable $\overline{X}$ has

> $\mathrm{E}(\overline{X}) : \mu$

and

> $$\mathrm{V}(\overline{X}) : \frac{\sigma^2}{n}$$
>
> The point is that the variance of a sample mean of size n is the population variance devided by the sample size n which makes it smaller
>
> i.e. as the sample size increases, the variability of the sample mean decreases.

Example:[Seed lengths]

One botanist measured the length of 10 seeds from the same plant. The seed lengths measurements are $X_1, X_2, \ldots, X_{10}$. Suppose it is known that the seed lengths are iid with mean $\mu = 5$ mm and variance $\sigma^2 = 2$ mm.

Calculate the mean and variance of the average of 10 seed measurements.

Central Limit Theorem

The Most Important Result in Statistics

Central limit theorem

One of the most frequently used statistics in engineering applications is the sample mean. We can relate the mean and variance of the probability distribution of the sample mean to those of a single observation when an iid model is appropriate.

In the case of the sample mean, if the sample size (n) is large enough, we can also approximate the *shape* of the *probability distribution function* of the sample mean!

Central limit theorem

If $X_1, \ldots, X_n$ are **independent** and **identically** distributed (iid) random variable (with mean μ and variance σ^2), then for large n, the variable $\overline{X}$ is approximately normally distributed. That is,

$$\overline{X} \dot{\sim} Normal\left(\mu, \frac{\sigma^2}{n}\right)$$

This is one of the **most important** results in statistics.

Example: [Tool serial numbers]

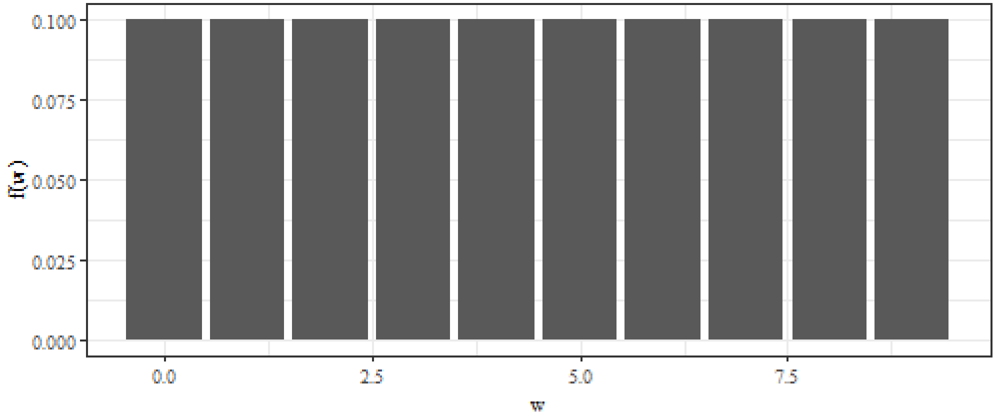
Consider selecting the last digit of randomly selected serial numbers of pneumatic tools. Let

$$\begin{aligned} W_1 &= \text{the last digit of the serial number} \\ &\quad \text{observed next Monday at 9am} \\ W_2 &= \text{the last digit of the serial number} \\ &\quad \text{observed the following Monday at 9am} \end{aligned}$$

A plausible model for the pair of random variables W_1, W_2 is that they are independent, each with the marginal probability function

$$f(w) = \begin{cases} .1 & w = 0, 1, 2, \ldots, 9 \\ 0 & \text{otherwise} \end{cases}$$

Example: [Tool serial numbers]

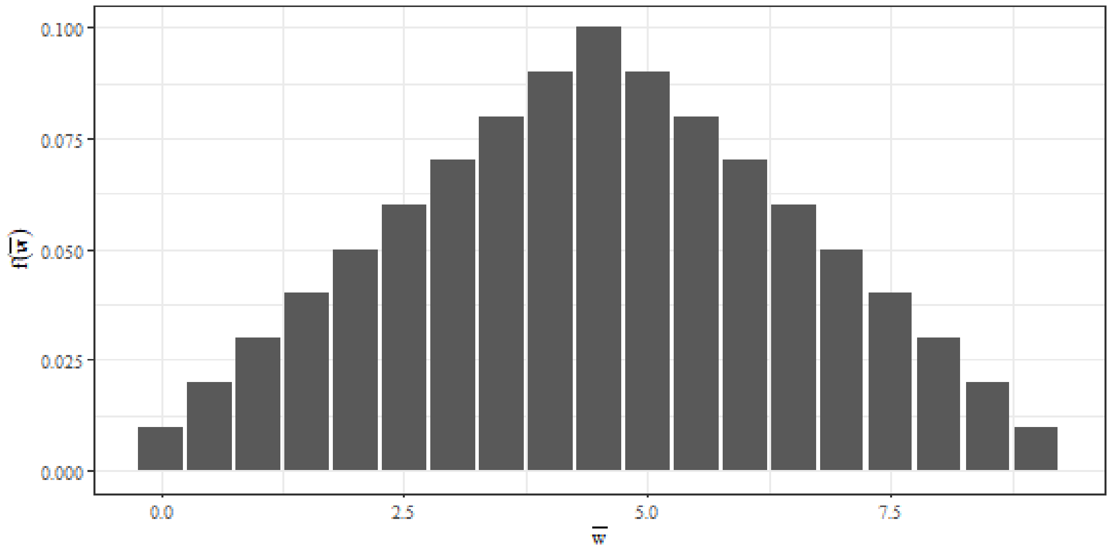


With $\mathrm{E}W = 4.5$ and $\mathrm{Var}W = 8.25$.

Using such a distribution, it is possible to see that $\overline{W} = \frac{1}{2}(W_1 + W_2)$ has probability distribution

$\overline{w}$	$f(\overline{w})$	$\overline{w}$	$f(\overline{w})$	$\overline{w}$	$f(\overline{w})$	$\overline{w}$	$f(\overline{w})$	$\overline{w}$	$f(\overline{w})$
0.00	0.01	2.00	0.05	4.00	0.09	6.00	0.07	8	0.03
0.50	0.02	2.50	0.06	4.50	0.10	6.50	0.06	8.5	0.02
1.00	0.03	3.00	0.07	5.00	0.09	7.00	0.05	9	0.01
1.50	0.04	3.50	0.08	5.50	0.08	7.50	0.04		

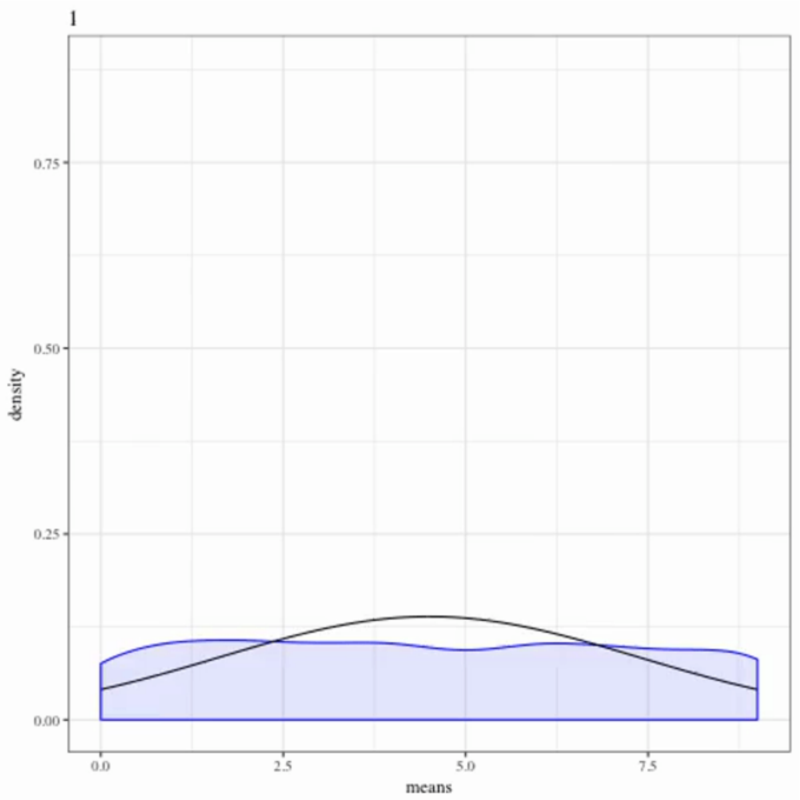
Example: [Tool serial numbers]



Comparing the two distributions, it is clear that even for a completely flat/uniform distribution of W and a small sample size of $n = 2$, the probability distribution of W looks more bell-shaped than the underlying distribution.

Now consider larger and larger sample sizes, $n = 1, \ldots, 40$:

Watch how CLT works here



Example: [Stamp sale time]

Imagine you are a stamp salesperson (on eBay). Consider the time required to complete a stamp sale as S, and let

$$\overline{S} = \text{the sample mean time required to complete the next 100 sales}$$

Each individual sale time should have an $Exp(\alpha = 16.5s)$ distribution. We want to consider approximating $P[\overline{S} > 17]$.

Example: [Cars]

Suppose a bunch of cars pass through certain stretch of road. Whenever a car comes, you look at your watch and record the time. Let X_i be the time (in minutes) between when the i^{th} car comes and the $(i+1)^{th}$ car comes for $i = 1, \ldots, 44$. Suppose you know the average time between cars is 1 minute.

Find the probability that the average time gap between cars for the next 44 cars exceeds 1.05 minutes.

Example: [Baby food jars, cont'd]

The process of filling food containers appears to have an inherent standard deviation of measured fill weights on the order of $1.6g$. Suppose we want to calibrate the filling machine by setting an adjustment knob and filling a run of n jars. Their sample mean net contents will serve as an indication of the process mean fill level corresponding to that knob setting.

You want to choose a sample size, n, large enough that there is an 80% chance the sample mean is within $.3$g of the actual process mean.

Example: [Printing mistakes]

Suppose the number of printing mistakes on a page follows some unknown distribution with a mean of 4 and a variance of 9. Assume that number of printing mistakes on a printed page are iid.

- What is the approximate probability distribution of the average number of printing mistakes on 50 pages?

- Can you find the probability that the number of printing mistakes on a single page is less than 3.8?

Example: [Printing mistakes]

- Can you find the probability that the average number of printing mistakes on 10 pages is less than 3.8?

- Can you find the probability that the average number of printing mistakes on 50 pages is less than 3.8?