

# Introduction to Statistics - Young Researchers Fellowship Program

## Lecture 3 - Introduction to Probability

Daniel Sánchez Pazmiño

### Table of contents

<b>Intro to Probability</b>	<b>1</b>
Experiments, Events and Sample Spaces . . . . .	2
Basic probability operations . . . . .	3
The Complement . . . . .	3
The Union . . . . .	4
The Intersection . . . . .	4
Mutual exclusiveness and the addition rule . . . . .	4
Conditional Probability . . . . .	5
Independence . . . . .	6
The Multiplication Rule . . . . .	7
Summary . . . . .	7

### Intro to Probability

We recently introduced the issue of not having access to the population data, which complicates our understanding of that population. Statisticians have tried to solve that problem by using samples to estimate population parameters, such as the population mean, by using a sample statistic, such as the sample mean.

However, how do we know that the sample mean is a good estimate of the population mean? How do we know that the sample does not contain people who are fundamentally different of the people in the population? Is there a possibility, or likelihood, that the sample mean is very different from the population mean? All of those questions can be answered by using probability.

Probability is a numerical measure of how likely an event is to happen. For instance, when we see clouds which are very gray, we know that it is very *likely* that it will rain. However, we do not know for sure that it will rain. In other words, there is a *probability* that it will rain, but we cannot be fully certain that it will rain for sure.

How do we compute the measures of probability? There are several approaches to do this, but we will focus on the *frequentist* approach. This approach is based on the idea that the probability of an event is equal to the number of times that the event occurs divided by the total number of possible outcomes. Implicitly, a frequentist assumes that the probability of an event is equal to the *long-run* frequency of that event, for which we need historical data. There is another major approach to probability, which is the *Bayesian* approach. This approach is based on the idea that the probability of an event is equal to the degree of belief that we have that the event will occur. In other words, a Bayesian assumes that the probability of an event is equal to the *subjective* belief that the event will occur, for which we need to ask people about their beliefs. Using bayesian methods is beyond the scope of these sessions, as knowing a lot frequentist statistics is a prerequisite to understand bayesian statistics.

For instance, if we want to know the probability of getting a head when we flip a coin, we can divide the number of times that we get a head by the total number of possible outcomes. In this case, the total number of possible outcomes is 2, since we can either get a head or a tail. If we flip the coin once, and we get a head, then the probability of getting a head is  $1/2 = 0.5$ . If we flip the coin 10 times, and we get 5 heads, then the probability of getting a head is  $5/10 = 0.5$ . If we flip the coin 100 times, and we get 50 heads, then the probability of getting a head is  $50/100 = 0.5$ .

Thus, in this sense, probability is always a number between 0 and 1, where 0 means that the event will never happen, and 1 means that the event will always happen. In the case of the coin, the probability of getting a head is 0.5, which means that we will get a head half of the times that we flip the coin.

## Experiments, Events and Sample Spaces

In order to compute probabilities, we need to define the *experiment*, the *event* and the *sample space*. The experiment is the process that we are interested in, such as flipping a coin. The event is the outcome of the experiment that we are interested in, such as getting a head. The sample space is the set of all possible outcomes of the experiment, such as getting a head or a tail.

We assign probabilities to certain events happening based on the following formula.

$$P(\text{event}) = \frac{\text{number of times that the event occurs}}{\text{total number of possible outcomes}}$$

We often note an event with a capital letter, such as  $A$ . The probability of an event is often noted as  $P(A)$ . For instance, the probability of getting a head when we flip a coin is  $P(A) = 0.5$ . The probability of getting a tail when we flip a coin is  $P(B) = 0.5$ .

From this approach, two rules emerge:

1. The probability of an event is always between 0 and 1.
2. The sum of the probabilities of all possible events is equal to 1.

## Basic probability operations

Probabilities of events can be manipulated using operations. The most common operations are the following:

### The Complement

The *complement* of an event is the probability of the event not happening. For instance, the complement of getting a head is getting a tail. The complement of getting a tail is getting a head.

Imagine we have data on the grades and demographic characteristics of 500 students. In a four month long academic term, some of the students have passed the course and some others have failed the course. We can compute the probability of passing the course by dividing the number of students who passed the course by the total number of students. We can also compute the probability of failing the course by dividing the number of students who failed the course by the total number of students. The probability of passing the course plus the probability of failing the course is equal to 1.

If 100 students failed the course, then the probability of failing the course is  $100/500 = 0.2$ . The complement of failing the course is passing the course, so the probability of the complement is  $(500 - 100)/500 = 0.8$ . As you can see, if we sum the probability of the event of failing the course with the probability of passing the course, we get  $0.2 + 0.8 = 1$ . This is a general relationship between the probability of an event and the probability of the complement of that event. We write it down as follows:

$$P(event) + P(complement) = 1$$

So, we can quickly compute the complement of any probability as follows:

$$P(complement) = 1 - P(event)$$

## The Union

The *union* of two events is the probability of either event happening. For instance, the union of getting a head and getting a tail is the probability of getting a head or a tail. We denote it as  $P(A \cup B)$ . We also define the union of an event with another event as the event which contains all of the possibilities or sample points that are in either event. For instance, the union of getting a head and getting a tail is the event of getting a head or a tail.

## The Intersection

The *intersection* of two events is the probability of both events happening. For instance, the intersection of getting a head and getting a tail is the probability of getting a head and a tail. We denote it as  $P(A \cap B)$ . However, if we are using a normal coin, we know that it is impossible to get a head and a tail at the same time. Thus, the intersection of getting a head and getting a tail is 0.

We also define the intersection of an event with another event as the event which contains all of the possibilities or sample points that are in both events. For instance, the intersection of getting a head and getting a tail is the event of getting a head and a tail. Again, this is impossible.

## Mutual exclusiveness and the addition rule

When we want to know the probability of either event happening, we can use the union. For instance, we can use the union to know the probability of getting a head **or** a tail when we flip a coin. We can also use the union to know the probability of getting a student who *either* passed or failed a course.

In this case, knowing the probability of either event happening is a little trivial, since there are just two possibilities. However, if we have more than two possibilities, we can still use the union to know the probability of either event happening. For instance, we want to know whether or not a student passed a course, failed a course or withdrew from the course. In this case, we can use the union to know the probability of a student *either* failing or withdrawing from the course.

If the probability of failing the course is 0.2, and the probability of withdrawing from the course is 0.1, then the probability of either failing or withdrawing from the course is  $0.2 + 0.1 = 0.3$ .

So far, we have computed the union of two events by simply summing the probabilities of the events. This is because events are *mutually* exclusive, meaning that they cannot happen at the same time. For instance, we cannot get a head and a tail at the same time. We cannot get a student who passed and failed a course at the same time.

For mutually exclusive events, the union is computed as follows:

$$P(A \cup B) = P(A) + P(B)$$

However, if the events are not mutually exclusive, we might run in a little problem. For instance, we might be interested of a student exhibiting mental health issues and failing a course. In this case, the student can exhibit mental health issues and fail a course at the same time, or only exhibit one of the two events. Thus, the events are not mutually exclusive.

If the probability of a student exhibiting mental health issues is 0.5, and the probability of a student failing a course is 0.5, then you might be tempted to say that the probability of a student exhibiting mental health issues and failing a course is  $0.5 + 0.5 = 1$ . A probability of 1 means that the event will always happen, which we know is not true. What happened?

The problem is that we counted the probability of a student exhibiting mental health issues and failing a course twice. We counted it when we computed the probability of a student exhibiting mental health issues, and we counted it when we computed the probability of a student failing a course. In the probability of a student exhibiting mental health issues, we included some students who both had mental health issues and failed a course. In the probability of a student failing a course, we included some students who both had mental health issues and failed a course. Thus, we counted the probability of a student exhibiting mental health issues and failing a course twice.

To deal with these cases accurately, we need the *addition rule*. The addition rule is the following:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If we know that 50% of the students exhibit mental health issues, 50% of the students fail a course, and 20% of the students both exhibit mental health issues and fail a course, then the probability of a student exhibiting mental health issues or failing a course is  $50\% + 50\% - 20\% = 80\%$ . This seems more reasonable than 100%.

When two events are mutually exclusive, the probability of the intersection is 0. That is why for mutually exclusive events, the addition rule is simply the sum of the probabilities of the events, as the intersection is zero.

## Conditional Probability

The *conditional probability* of an event is the probability of that event happening given that another event has happened. This is often a source of confusion, so let us go through an example.

If we know that 50% of students in our 500 sample exhibit mental health issues, the probability of a student exhibiting mental health issues is 0.5. If we know that 50% of students failed a course, the probability of a student failing a course is 0.5. What if we want to focus *only* on the students who failed a course? What is the probability of a student exhibiting mental health issues *given* that the student failed a course?

The probability of an event happening under the condition that some other event has already happened is called the *conditional probability*. We denote it as  $P(A|B)$ , which is read as the probability of  $A$  given  $B$ . In this case,  $A$  is the event of a student exhibiting mental health issues, and  $B$  is the event of a student failing a course. Thus, we want to know the probability of a student exhibiting mental health issues given that the student failed a course.

The conditional probability is computed as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

So, if we know that out of all students, 20% of them exhibit mental health issues and fail a course, and 50% of them fail a course, then the probability of a student exhibiting mental health issues given that the student failed a course is  $20\%/50\% = 0.4$ . This conditional probability can be interpreted as follows: out of all students who failed a course, 40% of them exhibit mental health issues. So, we can say that failing a course is associated with exhibiting mental health issues, or that exhibiting mental health issues seems to be dependent on failing a course.

## Independence

Two events are *independent* if the probability of one event happening does not depend on the probability of the other event happening. When two events are unrelated, they are independent. We can be pretty sure that the probability of a student failing a course does not depend on the probability of a raccoon finding a bag of trash 10 km away from the university. Thus, we can say that the event of a student failing a course is independent of the event of a raccoon finding a bag of trash 10 km away from the university.

The concept of independence is closely related to the concept of conditional probability. If two events are independent, the fact that one of them happens does not affect the other. In that sense, conditional probabilities essentially do not exist: the probability of event  $A$  happening given that event  $B$  has happened is the same as the probability of event  $A$  happening.

So, the following properties apply to independent events:

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

If this is not true, the events are dependent.

## The Multiplication Rule

From the expression of conditional probability, we can derive the *multiplication rule*. The multiplication rule is the following:

$$P(A \cap B) = P(A|B)P(B)$$

$$P(A \cap B) = P(B|A)P(A)$$

This rule is useful when we want to know the probability of two events happening at the same time. For instance, we can use the multiplication rule to know the probability of a student exhibiting mental health issues and failing a course, if we know the probability of failed students exhibiting mental health issues and the probability of a student failing a course.

In cases where we know that two events are independent, we can still use the multiplication rule with a small adjustments which makes things very convenient. Since we know that under independency the conditional probability is equal to the probability of the event, we can use the multiplication rule as follows:

$$P(A \cap B) = P(A|B)P(B) = P(A)P(B)$$

This is very convenient, since we do not need to know the conditional probability of one event given the other event. We can simply multiply the probabilities of the events.

If we know that the probability of a raccoon finding a bag of trash 10 km away from the university is 0.1, and the probability of a student failing a course is 0.5, then the probability of a student failing a course and a raccoon finding a bag of trash 10 km away from the university is  $0.1 * 0.5 = 0.05$ .

## Summary

- Probability is a numerical measure of how likely an event is to happen.
- The frequentist approach to probability is based on the idea that the probability of an event is equal to the number of times that the event occurs divided by the total number of possible outcomes.
- The probability of an event is always between 0 and 1.
- The sum of the probabilities of all possible events is equal to 1.
- The complement of an event is the probability of the event not happening.
- The union of two events is the probability of either event happening.
- The intersection of two events is the probability of both events happening.

- The addition rule is the following:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- The conditional probability of an event is the probability of that event happening given that another event has happened.
- The multiplication rule is the following:  $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- Two events are independent if the probability of one event happening does not depend on the probability of the other event happening.