Introduction to Statistics - Young Researchers Fellowship Program

Lecture 3 - Introduction to probability

Daniel Sánchez Pazmiño

Laboratorio de Investigación para el Desarrollo del Ecuador

September 2024

Why probability?

- Probability is a foundational concept in inferential statistics.
- It is related to the idea that because it is hard to access the entire population, we can use a sample to make inferences about the population.
- We then take a sample, which is thought to be representative of the population, and use it to make inferences about the population.

Why probability?

- However, even if we ensure the sample is similar to the population based on theoretical knowledge, we would never expect that it is exactly the same.
  - There will always be some difference, potentially due to random variation, which make the sample different from the population.
- Probability is the mathematical tool that allows us to quantify the uncertainty associated with the sample.
  - How far are we from the population?
  - If there are significant differences, are we looking for an unrepresentative sample?
  - Is a sample mean fundamentally different from the population mean?
- All the questions above can be answered using the concepts of probability.

Probability

- A measure of the likelihood through which an event will occur.
- For example: when seeing gray clouds, we might think that it will rain. The probability of rain is higher when we see gray clouds.
    - However, it is not certain that it will rain. It might rain, or it might not (it is uncertain).

Approaches to probability calculation

- There are two main approaches to calculate probability:
  - Frequentist approach
  - Bayesian approach
- The frequentist approach is based on the idea that probability is the long-run frequency of an event.
  - For example, the probability of getting a head when flipping a coin is 0.5 because in the long run, half of the flips will be heads.
- This means that frequentist probability will calculate a probability based on past data

$$\text{Probability} = \frac{\text{Number of times an event occurred}}{\text{Total number of trials}}$$

Approaches to probability calculation

- Another major approach to probability is the Bayesian approach (pronounced “bay-zee-an”).
- The Bayesian approach is based on the idea that probability is a a measure of the degree of belief that an event will occur.
  - Hence, it is like a measure of how “good” our information is.
- We won’t focus Bayesian approaches on this module, but it is good to know that it exists, and researchers have devoted a lot of time to it.
  - Typically, we will want to understand frequentist approaches first and gain mathematical intuition before we move to Bayesian approaches.

Basic probability rules

- Two major rules govern probability:

1. All probabilities are between 0 and 1.
   - A probability of 0 means that the event will never happen (0% chance).
   - A probability of 1 means that the event will always happen (100% chance).
   - This means probabilities are always in proportion form, but can be easily converted to percentages.
2. The sum of the probabilities for all possible outcomes of an event must sum to 1.
   - In an exam, you'd have two outcomes: pass or fail (assuming no other "weird" outcomes).
   - The probability of passing plus the probability of failing must sum to 1.
   - This means 100% of outcomes are covered in the "possible outcomes" set, each with a probability.

Probability: the world of board games

- Two typical examples often emerge when discussing probability:
  - Flipping a "fair" coin
  - Rolling a "fair" die
- A fair coin is a coin that has an equal probability of landing on heads or tails.
- So, because there is 1 head and 1 tail, the probability of getting a head is $1/2$, and the probability of getting a tail is $1/2$.

$$(P(H)) = \frac{Number of heads}{Total number of outcomes} = \frac{1}{2}$$

$$(P(T)) = \frac{Number of tails}{Total number of outcomes} = \frac{1}{2}$$

Probability: the world of board games

- The other common example is rolling a die.
- A die has 6 faces, each with a number from 1 to 6.
- The probability of getting a 1 is $1/6$, the probability of getting a 2 is $1/6$, and so on.

Probability: the world of board games

- It is common to also see drawing a card from a deck of cards as an example of probability.
- Decks of cards have 52 cards, with 4 suits (hearts, diamonds, clubs, and spades) and 13 cards in each suit.
- Special cards include the King, Queen, the Jack, and the Ace.
  - There are 4 of each of these cards in the deck.
  - The probability of drawing a King is $4/52$, the probability of drawing a Queen is $4/52$, and so on.

The lingo: experiments, events, outcomes and sample spaces

- There is specific terminology and definitions in probability that must be understood before we move forward.
- An **experiment** is a process that generates well-defined outcomes, and we're interested in probabilities associated with these outcomes.
  - For example, flipping a coin, rolling a die, or drawing a card from a deck.
- An **event** is a subset of the outcomes of an experiment.
  - For example, in flipping a coin, the event of getting a head is a subset of the outcomes of the experiment.

The lingo: experiments, events, outcomes and sample spaces

- In simple experiments, we often see “events” covering one outcome, and not more.
  - However, events can “group” outcomes in a specific way.
  - E.g. the event of getting an even number when rolling a die. This event covers outcomes 2, 4, and 6.
- The **sample space** is the set of all possible outcomes of an experiment.
  - For example, when flipping a coin, the sample space is {H, T}.
  - When rolling a die, the sample space is {1, 2, 3, 4, 5, 6}.

Set theory and probability

- Probability is closely related to set theory.
  - Sets are collections of objects, and in probability, we often think of events as sets of outcomes.
- Thus, it is common to see notation from set theory in probability.
  - A set is denoted by curly braces, e.g. $\{1, 2, 3, 4, 5, 6\}$.
  - We use letters to denote sets, e.g. A, B, C, etc.
  - We use set operations to calculate probabilities.
  - We use Venn diagrams to visualize probabilities.
- Set theory operators, such as belongs to ($\in$), union ($\cup$), intersection ($\cap$), and complement ($'$) are used in probability.

Basic operations in probability

- There are three basic operations in probability:
  - Union
  - Intersection
  - Complement

Union of