Parametric Statistics

Sofia Triantafillou

sof.triantafillou@gmail.com

University of Crete
Department of Mathematics and Applied Mathematics

October 11, 2021

Today: Probability

- DeGroot and Schervish, Chapters 1-2.
- Wasserman, Chapter 1.

What is probability?

Probability

is a language for quantifying uncertainty. It is a way to quantify how likely something is to occur.

Experiment

An experiment is any real or hypothetical process, in which the **possible outcomes** can be identified ahead of time. Events are **sets** of possible outcomes. Probability is then a way to describe how likely each event is.

Possible experiments

- We toss a coin 2 times.
  Possible Outcomes:
  Sample Space:
  Examples of Events:
  Probability of each event:
- We measure the temperature.
  Possible Outcomes:
  Sample Space:
  Examples of events:
  Probability of each event:

Sample Spaces

- The sample space is Ω is the set of possible outcomes of an experiment.
- $\omega \in \Omega$ is are called **sample outcomes**, or **elements**.
- Subsets of Ω are called **events**.

Example:

Coin tossing: If you toss a coin twice then

$$\Omega = \{HH, HT, TH, TT\}$$

The event that both tosses are heads are:
The event that the first toss is heads is:

Sample Space: Examples

Let ω be the outcome of measuring temperature. A sample space for this experiment is $\Omega = (-\infty, \infty)$. Is this accurate?

- What are the elements of Ω?
- Example events: temperature is 15.5.
- Example events: temperature is at least 10 but lower than 20 is $A = [10, 20)$.

Probability

We want to assign a real number $\mathbb{P}(A)$ to every event A which represents how likely event A is to occur. This is called the probability of A.

- Axiom 1: $\mathbb{P}(A) \geq 0$ for every A
- Axiom 2: $\mathbb{P}(\Omega) = 1$
- Axiom 3: for a finite sequence $A_1, A_2, \ldots, A_n$ of disjoint events

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Are all subsets of Ω events?

- Events are subsets of Ω.
- Are all subsets of Ω events?

In general, we will need to restrict ourselves to a class of subsets of Ω that are called a σ-algebra. The sets in a σ-algebra have the property of being measurable. This is necessary because some crazy things can happen if you try to assign probabilities to non-measurable sets. Example: We cannot assign probabilities to every element of the real line (why do you think?).

Some nice notes on σ-algebra

σ-algebra

Consider a set Ω. A σ-algebra or a σ-field on Ω is a set of subsets $\mathcal{A}$ of Ω that sastisfy the following conditions:

- $\emptyset \in \mathcal{A}$
- If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$ (closed under complement)
- if $A_1, A_2, \dots \in \mathcal{A}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ (closed under countable unions)

The sets in $\mathcal{A}$ are said to be measurable. We $(\Omega, \mathcal{A})$ a measurable space.

σ-algebras are necessary to ensure that we can assign probabilities to subsets of uncountable sample spaces (e.g. the real line) without running into problems (e.g., probabilities not summing to 1). More on this topic on your probability/measure theory courses :)

Another reason for using $\sigma-$ algebras

- Reason 1: We want to be able to assign probabilities to all the events we are interested in without problems.
- Reason 2: We may be interested in a specific set of events (we do not need all events) or we may want to represent partial information.
- Example: Assume someone flips a coin twice, but you only observe the first flip. What is the sigma-algebra that can represent the information you have?

Generating $\sigma-$ algebras

You can always construct the smallest σ-algebra for a set of subsets $\mathcal{M}$. This is called the σ-algebra generated by $\mathcal{M}$, and we denote it $\sigma(\mathcal{M})$

Example

$$\Omega = \{a, b, c, d\}, \mathcal{M} = \{\{a\}, \{b\}\}$$

$$\sigma(\mathcal{M}) = ?$$

Probability Measure

A function from a σ-algebra $\mathbb{A}$ to $[0, 1]$ is called a probability measure if it satisfies the following:

- Axiom 1: $\mathbb{P}(A) \geq 0$ for every A
- Axiom 2: $\mathbb{P}(\Omega) = 1$
- Axiom 3: for an infinite sequence $A_1, A_2, \ldots$ of disjoint events

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Probability Space

Ok so now we have:

- A sample space Ω for our experiment.
- A set of events $\mathcal{A}$ (measurable subsets of Ω to which we can assign a probability without running into problems with the axioms of probability).
- A function from the $\sigma-$algebra $\mathcal{A}$ to $[0, 1]$ that indicates how likely an element of $\mathcal{A}$ is to occur.

The triplet $(\Omega, \mathcal{A}, \mathbb{P})$ is called a **probability space**.

What is probability?

Interpretation of Probability:

- Frequency
- Degree of belief

Properties of Probabilities (1)

Based on the axioms of probability, we can derive several properties:

- The probability of an impossible event is 0:

$$P(\emptyset) = 0$$

- Axiom 3 also holds for finite sequences of events:

$$P(\bigcup_{i=1}^{N} A_i) = \sum_{i=1}^{N} P(A_i)$$

(page 17 of DGS)

Properties of Probabilities (2)

The law of total probability

Let $B_1, \ldots, B_n$ be a partition of the sample space. Then for any event A,

$$P(A) = \sum_i P(A \cap B_i)$$

Let's prove it for a very simple partition: B, B^c. Reminder:

- Set partitioning: $A = (A \cap B) \cup (A \cap B^c)$.
- Distribution law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- $(A \cap B), (A \cap B^c)$ are disjoint.

Properties of Probabilities (3)

Lemma

For any events A and B,

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

Proof.

□

Example.

Two coin tosses. Let H_1 be the event that heads occurs on toss 1 and let H_2 be the event that heads occurs on toss 2.If all outcomes are equally likely, what is the $\mathbb{P}(H_1 \cup H_2)$?

Continuity of Probabilities

Theorem

If $A_n \to A$*, then* $\mathbb{P}(A_n) \to \mathbb{P}(A)$

Discrete Probability Spaces.

For uncountable sample spaces, we need σ-algebras (which do not include all subsets of Ω) to avoid mathematical difficulties. Finite and countable sample spaces are much easier to think about:

Examples

- Consider a single toss of a coin. If we believe that heads (H) and tails (T) are equally likely, find an appropriate probability distribution.
- Consider a single roll of a die. if we believe that all six outcomes are equally likely, find an appropriate probability model.
- We toss an unbiased coin n times. What is an appropriate probability model?

Counting in a uniform (simple) probability space

In discrete sample spaces, if all outcomes of an experiments are considered equally likely, then for each event A,

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

Example:

We toss a (fair) coin twice. Ω has 36 elements: $11, 12, \ldots, 16, 21, \ldots, 26, \ldots, 61, \ldots, 66$

Let's say we are interested in the event "at least one 6" To assign a probability to each possible event, we need to be able to count: The number of points in Ω and A. To do so, we need some combinatorial methods.

Multiplication rule

General counting rule:

- r steps
- n_r choices at each step
- Then the number of choices are $n_1 \times n_2 \times \cdots \times n_r$

Permutations/Combinations

Permutations

Number of distinct ways to order n objects: $n!$

Combinations

Number of distinct ways of choosing k elements from a collection of n objects:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!}$$

Independent events

Sometimes we know that the events are independent based on the construction of the experiment (e.g., we know that the

Definition

Two events A and B are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

We denote this as

$$A \perp\!\!\!\perp B$$

A set of events $\{A_i : i \in I\}$ is independent if

$$\mathbb{P}(\cap_{i \in J} A_i) = \prod \mathbb{P}(A_i)$$

for every finite subset J of I.

Independent events

We already discussed how if we flip a coin twice, the probability of two heads is $\frac{1}{2} \times \frac{1}{2}$. This is because we consider the two tosses as independent. This means, the outcome of the first coin flip does not affect the outcome of the second flip.

Definition

Two events A and B are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

We denote this as

$$A \perp\!\!\!\perp B$$

A set of events $\{A_i : i \in I\}$ is independent if

$$\mathbb{P}(\bigcap_{i \in J} = \prod \mathbb{P}(A_i)$$

for every finite subset J of I.

Independence is not always intuitive

- If we flip a coin twice, we typically assume that the flips are independent (i.e., the coin has no memory of previous flip).
- Sometimes, independence just comes up. Example: We are roll a fair die and we are interested in the following two events: A : "The outcome is an even number" B : "The outcome is one of the numbers $\{1, 2, 3, 4\}$"

Independent events

We already discussed how if we flip a coin twice, the probability of two heads is $\frac{1}{2} \times \frac{1}{2}$. This is because we consider the two tosses as independent. This means, the outcome of the first coin flip does not affect the outcome of the second flip.

Definition

Two events A and B are **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

We denote this as

$$A \perp\!\!\!\perp B$$

A set of events $\{A_i : i \in I\}$ is independent if

$$\mathbb{P}(\bigcap_{i \in J} = \prod \mathbb{P}(A_i)$$

for every finite subset J of I.

Conditional Probability

One way to interpret the independence of events is as follows:

- Consider again the following two events:
  A : "The outcome is an even number"
  B : "The outcome is one of the numbers $\{1, 2, 3, 4\}$"
- You want to bet on event A. How much are you willing to bet?
- I roll the die and tell you that event 2 has happened (hence, the outcome is one of $\{1, 2, 3, 4\}$.
- How much are you willing to bet now?
- We just described the **conditional probability** $P(A = true|B = true)$

Conditional Probability

Definition (Conditional Probability of A given B)

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

The probability of event A in the universe (sample space) where event B has already happened.

Conditional Probability

Definition (Conditional Probability of A given B)

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

The probability of event A in the universe (sample space) where event B has already happened.

Bayes Rule

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

Let's prove it

Why is Bayes Rule so important?

Very often, people confuse $P(A|B)$ and $P(B|A)$. These can be VERY different.

Think about it:

You read in the paper: "Half of the people hospitalized with covid-19 are fully vaccinated". Do you think that getting the vaccine lowers your chances of getting hospitalized?

Why is Bayes Rule so important?

- Vacc: Yes if vaccinated, zero otherwise
- Hosp: Yes if hospitalized, zero otherwise.
- $P(Hosp|Vacc) = 0.01$
- $P(Hosp|\neg Vacc) = 0.2$
- Three different possibilities: $P(Vacc) = 0.8, 0.5, 0.99$

Let' s use Bayes rule to compute $P(Vacc|Hosp)$ for all three cases.

Review(1)

What did we talk about today:

- Probability is a way to quantify the probability with which an event occurs.
- For discrete sample spaces, it is pretty easy to define a probability measure over the set of all possible events.
- For continuous sample spaces, we need a $\sigma-$ algebra to make sure everything is well-defined according to the axioms of probability.
- We can use the axioms of probability to prove several properties of probability

Review(2)

What did we talk about today:

- Two events are called independent when knowing the value of one doesn't influence the probability of the value of the other.
- The conditional probability of A given B denotes the probability of event A in a world where B has occured.
- Bayes rule connects $P(A|B)$ and $P(B|A)$. Very often, these two are confused but they are not the same.