Foundations of probability theory

Joaquín Rapela

Gatsby Computational Neuroscience Unit
University College London

June 9, 2023

Contents

Main reference

I will mainly follow chapters seven *Foundations of probability theory* and eight *Conditional probability and Bayes* from Tijms (2012).

Contents

Historical notes

- explain the frequency-based interpretation of probability.
- constructing the mathematical foundations of probability theory has proven to be a long-lasting process of trial an error.
- the approach of defining probability as relative frequencies of repeatable experiments lead to unsatisfactory theory (why?) https://www.jstor.org/stable/pdf/20115155.pdf
- the frequency view of probability has a long history that goes back to Aristotle.
- in 1933 the Russian mathematician Andrej Kolmogrov (1903-1987) laid a satisfactory mathematical foundation of probability theory.

He created a set of axioms. Axioms state a number of minimal requirements that the probability objects should satisfy. From these few algorithms all claims of probability can be derived, as we will see.

Probability model

Definition 1 (Probability model)

A **probability model** *is a matematical representation of a random experiment.*

Probability model

Definition 1 (Probability model)

A **probability model** *is a matematical representation of a random experiment. It consists of a description of all possible outcomes of the experiment (i.e.,* **sample space***), a set of subsets of the sample space (i.e.,* **events***), and an assigment of probability to events (i.e.,* **probability measure***).*

Sample space

Definition 2 (Sample space)

The set of all samples in an experiment is called the **sample space**. *It is denoted by* Ω.

Event

Definition 3 (Event)

An **event** *is a subset of the sample space. We denote the collection of all events by* $\mathcal{F}$.

Event

Definition 3 (Event)

An **event** *is a subset of the sample space. We denote the collection of all events by $\mathcal{F}$.*

Notes:

1. We will only assign probabilities to events (i.e., to sets $A \in \mathcal{F}$).
2. For finite or countable sample spaces, we can assign probabilities to any subset of the sample space. Thus, any subset of a finite or countable sample space can be an event.
3. For uncountable sample spaces, we can only assign probabilities to well behaved subsets of the sample space (i.e., to elements in a σ algebra of subsets of the sample space). Only well-behaved subsets of an uncountable sample space can be events.

Probability measure

Definition 4 (Probability measure)

A **probability measure** *is a function that assigns numbers between zero and one to events (i.e., $P : \mathcal{F} \rightarrow [0, 1]$).*

Probability model: definitions

Definition (Sample space)

The set of all samples in an experiment is called the **sample space**. *It is denoted by* Ω.

Definition (Event)

An **event** *is a subset of the sample space. We denote the collection of all events by* $\mathcal{F}$.

Definition (Probability measure)

A **probability measure** *is a function that assigns numbers between zero and one to events (i.e.,* $P : \mathcal{F} \rightarrow [0, 1]$*).*

Definition (Probability model)

A **probability model**, $\mathcal{M}$, *is a mathemtatical representation of a random experiment consiting of a sample space,* Ω, *a set of events,* $\mathcal{F}$, *and a probability measure,* P *(i.e.,* $\mathcal{M} = \{\Omega, \mathcal{F}, P\}$*).*

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once.

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H,\ T]$.

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, T]$. We assign a probability of 0.5 to each element of the sample space.

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair dice and count the number of rolls until the first six shows up.

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair dice and count the number of rolls until the first six shows up. The sample space is the set of positive integers.

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair dice and count the number of rolls until the first six shows up. The sample space is the set of positive integers. The probabilities $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \ldots$ can be assigned to the outcomes $1, 2, 3, \ldots$.

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, \ T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair dice and count the number of rolls until the first six shows up. The sample space is the set of positive integers. The probabilities $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \dots$ can be assigned to the outcomes $1, 2, 3, \dots$.
3. The experiment is to measure the time between the first and second spikes in an experimental trial.

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H,\ T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair dice and count the number of rolls until the first six shows up. The sample space is the set of positive integers. The probabilities $\frac{1}{6}, \frac{5}{6}\times\frac{1}{6}, \left(\frac{5}{6}\right)^2\times\frac{1}{6}, \ldots$ can be assigned to the outcomes $1, 2, 3, \ldots$.
3. The experiment is to measure the time between the first and second spikes in an experimental trial.

Examples of probability models

For each of the following examples, let's find the sample space and propose a probability measure.

1. The experiment is to toss a fair coin once. The sample space is the set $[H, \; T]$. We assign a probability of 0.5 to each element of the sample space.
2. The experiment is to repeately roll a fair dice and count the number of rolls until the first six shows up. The sample space is the set of positive integers. The probabilities $\frac{1}{6}, \frac{5}{6} \times \frac{1}{6}, \left(\frac{5}{6}\right)^2 \times \frac{1}{6}, \ldots$ can be assigned to the outcomes $1, 2, 3, \ldots$.
3. The experiment is to measure the time between the first and second spikes in an experimental trial. The sample space is the set $(0, \infty)$ of positive real numbers. We can assign a probability of $1 - \exp(-\lambda t)$ to the event that the second spike is fired less than t seconds after the first spike.

Axioms of probability theory

Axiom 1

$P(A) \geq 0, \quad \forall A \in \mathcal{F}$

Axiom 2

$P(\Omega) = 1$

Axiom 3

$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ *for every collection of pairwise disjoint events* $A_1, A_2, \ldots$

Experiment with equally likely outcomes

An experiment with equally likely outcomes is one with a finite number of outcomes $\omega_1, \ldots, \omega_N$, where all outcomes have the same probability (i.e., $P(\omega_i) = \frac{1}{N}$).

Experiment with equally likely outcomes

An experiment with equally likely outcomes is one with a finite number of outcomes $\omega_1, \ldots, \omega_N$, where all outcomes have the same probability (i.e., $P(\omega_i) = \frac{1}{N}$).

Claim 1

For any event A, $P(A) = \frac{N(A)}{N}$, where $N(A)$ is the number of outcomes in the set A.

Example: equally likely outcomes – discrete Ω

Example 1

John, Pedro and Rosita each roll on fair die. How do we calculate the probability that the score of Rosita is equal to the sum of the scores of John and Pedro?

Some basic rules

Rule 1

For any finite number of mutually exclusive events $A_1, \ldots, A_N$*,*

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_N)$$

Rule 2

For any event A*,*

$$P(A) = 1 - P(A^c)$$

where the event A^c *consists of all outcomes that are not in* A*.*

Rule 3

For any two events A *and* B*,*

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

Rule 1

Proof.

Denote by $\emptyset$ the empty set. We will firt prove that $P(\emptyset)=0$. Take $A_i=\emptyset$ for $i=1,2,\ldots$. Then $\emptyset=\bigcup_{i=1}^{\infty}A_i$. Next, by Axiom 3, $P(\emptyset)=\sum_{i=1}^{\infty}P(A_i)=\sum_{i=1}^{\infty}P(\emptyset)$. This implies that $P(\emptyset)=0$. Define $A_{N+j}=\emptyset$ for $j=1,2,\ldots$. Then

$$
\begin{aligned}
P(A_1\cup A_2\cup\ldots\cup A_N) &= P(A_1\cup A_2\cup\ldots\cup A_N\cup A_{N+1}\cup A_{N+2}\cup\ldots)\\
&= P(\bigcup_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)\\
&= \sum_{i=1}^{N}P(A_i)+\sum_{j=1}^{\infty}P(A_{N+j})=\sum_{i=1}^{N}P(A_i)
\end{aligned}
$$

$\square$

Notes:

1. the last equality in the second line holds by Axiom 1
2. the last equality in the third line holds because $P(A_{N+j})=P(\emptyset)=0$.

Proof of rule 2

Proof.

$\Omega = A \cup A^c$. A and A^c are disjoint. Then, by Rule 1, $P(\Omega) = P(A) + P(A^c)$. From Axiom 2, $P(\Omega) = 1$. Thus, $1 = P(A) + P(A^c)$.

□

Proof of rule 3

Proof.

$$\begin{aligned} A\cup B &= (A\setminus B)\cup(B\setminus A)\cup(A\cap B)\\ A &= (A\setminus B)\cup(A\cap B)\\ B &= (B\setminus A)\cup(A\cap B)\end{aligned}$$

Since the sets in the right-hand-side of the above equations are pairwise disjoint, by rule 1, we obtain

$$\begin{aligned} P(A\cup B) &= P(A\setminus B)+P(B\setminus A)+P(A\cap B)\\ P(A) &= P(A\setminus B)+P(A\cap B)\rightarrow P(A\setminus B)=P(A)-P(A\cap B)\\ P(B) &= P(B\setminus A)+P(A\cap B)\rightarrow P(B\setminus A)=P(B)-P(A\cap B)\end{aligned}$$

Replazing the equations on the right of the second and third line in the equation on the first line the rule is proved.

$$\begin{aligned} P(A\cup B) &= P(A)-P(A\cap B)+P(B)-P(A\cap B)+P(A\cap B)\\ &= P(A)+P(B)-P(A\cap B)\end{aligned}$$

□

Example: Chevalier de Mere to Blaise Pascal

The gambler Chevalier de Mere posed the following problem to the famous French mathematician Blaise Pascal in 1654. This problem marks the begining of probability theory.

Example 2

How many rolls of a fair die are required to have at least a 50% chance of rolling at least one six? How many rolls of two fair dice are required to hoave at least a 50% chance of rolling at least one double six?

Example:

- example 7.8 (rule 7-3, addition rule, easy)

Example:

- example 7.9 (rule 7-3): uses counting tools (binomial coefficient)
- wrong, but simple, approach
- correct, but more complicated, approach
- sampling approach

Example:

- example 7.10 (rule 7-1, birthday problem, used in example 8.6): uses counting tools (binomial coefficient)

Contents

Conditional probability

- p. 256: good motivation of conditional probability in the cards example
- Definition 8.1
- interpretation of condition probability with relative frequencies

Example:

- Example 8.1 (first ask students their intuition, as the problem is counter intuitive)
- do NOT present example 8.2 at this point, as it requires the concept of independence

Assigning probabilities by conditioning

Rule 4 For any sequence of events $A_1, \ldots, A_n$,

$$P(A_1, ..., A_n) = P(A_n | A_{n-1}, ..., A_1) \ldots P(A_1)$$

Example:

- redo Example 7.9 (solution following Rule 4)

Example:

- probability that it takes 10 or more cards before the first ace appears

Independent events

- motivation of independence definition with conditional probabilities
- Definition 8.2

Example

- Example 8.5

Example

- Example 8.6 (uses birthday problem, example 7.10)

Law of conditional probability

- example of dice followed by coin tosses

Rule 5 law of conditional probability. Let A be an event that can only occur if one of the mutually exclusive events $B_1, \dots, B_n$ occurs. Then

$$P(A) = P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)$$

Example

- example 8.6: tour the France (difficult!)

Bayes rule in odds form

- true/false hypothesis

Rule 6 The posterior probability $P(H|E)$ satisfies

$$\frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(H)}{P(\bar{H})} \frac{P(E|H)}{P(E|\bar{H})}$$

- interpretation of rule 6
- avoid need of P(E)
- prior odds + likelihood ratio or Bayes factor
- prior odds update with new evidence
- sequential update (mention Bayesian linear regression)

Example

- example 8.8

Example

- example 8.11
- add to the problem statement:
- in 1992, 4936 women were murdered in the US, of which roughly 1430 were murdered by their (ex)husbands or boyfriends
- 5% of the married women in the US have at some point been physically abused by their husbands.
- assume that a woman who has been murdered by some other than her husband had the same same chance of being abused by her husband as a randomly selected woman
- Alan Dershowitz admitted that a substantial percentage of the husbands who murder their wives, previous to the murder, also physically abuse their wives. Given this statement, we assume that the proability that a husband physically abused his wife, given that he killed her, is 50 percent.

Bayesian inference – discrete case

- explain posterior sequential update

Example

- example 8.13 (solve it analytically and by sampling)

References

Tijms, H. (2012). *Understanding probability*. Cambridge University Press.