Foundations of probability theory

Joaquín Rapela

Gatsby Computational Neuroscience Unit
University College London

May 18, 2022

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.

Probabilistic foundation

- sample space
- examples of finite, countable and uncountable
- mention the proof by Cantor that the real numbers are not countable

Axioms of probability theory

- events in countable and uncountable sample spaces (hint about sigma algebras)
- probability measure
- three axioms
- explain what an infinite union means
- probability space: sample space + events + probability measure = probability space
- building a probability measure for a finite or countable sample space.
- probability model = sample space + probability measure

Equally likely outcomes

- examples
- uncountable sample space

Some basic rules

Rule 1 For any finite number of mutually exclusive events $A_1, \ldots, A_n$,

$$P(A_1, \ldots, A_n) = P(A_1) + \ldots + P(A_n)$$

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,

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

Proof of rule 1

Proof of rule 2

Proof of rule 3

Example: Chevalier de Mere to Blaise Pascal

- example 7.7 (rule 7-2): Chevalier de Mere to Blaise Pascal 1654

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 $B_1, \ldots, B_n$ 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.