Combinatorics and Probability
Part IV: Probability Spaces

Lecture notes, version 1.2 – Nov. 2023

Hebert Pérez Rosés
Grau d'Enginyeria Matemàtica i Física
Universitat Rovira i Virgili

Contents

1. Axiomatic definition of probability
   - Sample space, event space, and probability axioms
   - Some consequences of the axioms
   - Laplacian spaces
   - Geometric probability
   - Monte Carlo methods

2. Conditional probability and independence

Different interpretations of probability

1. Quantitative measure of the “degree of certainty” of the observer (psychological interpretation)
2. Probability as the relative frequency of the occurrence of an event in a large number of trials (statistical interpretation)
3. Definitions of probability that take the notion of “equal likelihood” of events as the point of departure (classical mathematical definition)
4. Axiomatic definition based on set theory and measure theory (Andrei Nikolaevich Kolmogorov, 1903 – 1987)

Random experiment and sample space

Random experiment: Some action or measurement whose outcome cannot be predicted with absolute certainty.

Sample space: The set Ω of all possible outcomes of the random experiment.

Example 1: The action of casting a (fair) die is a random experiment. Here the sample space now is $\Omega_1 = \{1, \ldots, 6\}$.

Example 2: The action of flipping (tossing) a coin is also a random experiment. $\Omega_2 = \{ \text{HEADS} , \text{ TAILS} \}$.

Example 3: The action of flipping a coin three times is a random experiment. If $\Omega_2 = \{ \text{HEADS} , \text{ TAILS} \}$ then our sample space is $\Omega_3 = \Omega_2^3$.

Random experiment and sample space

Example 4: Another random experiment consists in counting the number of coin tosses that are necessary to obtain the first 'HEADS'. In this case $\Omega_4 = \{1, 2, \ldots, \infty\} = \mathbb{N} \cup \{\infty\}$.

Example 5: Throwing a dart is a random experiment. If we assume that the diameter of the tip of the dart is negligible, then our sample space is $\Omega_5 = \mathbb{R}^2$.

Definition: A sample space is *discrete* if it is finite or denumerable, otherwise it is continuous. In the previous examples $\Omega_1, \ldots \Omega_4$ are discrete, while Ω_5 is continuous.

Events and event space

We want to be able to define a probability measure on subsets of Ω, but in order to do so our subsets must have certain properties:

σ**-Algebra**: Let $\Omega \neq \emptyset$. A collection $\mathcal{F}$ of subsets of Ω is a σ-Algebra over Ω if

1. $\Omega \in \mathcal{F}$
2. If $A \in \mathcal{F}$ then $\overline{A} \in \mathcal{F}$, where $\overline{A}$ denotes the complement of A
3. If $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

The elements of $\mathcal{F}$ are called *events*.

Events and event space

Example 6: Let $\Omega \neq \emptyset$. Then the collections $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$ are both σ-algebras over Ω. The collection $\mathcal{F}_0$ is called *the trivial σ-algebra*, and $\mathcal{P}(\Omega)$ is called *the total σ-algebra* over Ω.

Example 7: Let $\Omega = \{1, 2, 3\}$. Then $\mathcal{F}_1 = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ is a σ-algebra over Ω, while $\mathcal{F}_2 = \{\emptyset, \{1\}, \{2\}, \{3\}, \Omega\}$ is not.

Example 8: If $\Omega \neq \emptyset$ and $\mathcal{F}_1, \mathcal{F}_2, \ldots$ are σ-algebras over Ω, then $\mathcal{F} = \bigcap_{i=1}^{\infty} \mathcal{F}_i$ is also a σ-algebra over Ω.

Definition

Let $\Omega \neq \emptyset$ and L a collection of subsets of Ω. The intersection of all σ-algebras over Ω that contain L is the smallest σ-algebra containing L, and it is called *the σ-algebra generated by L*.

Events and event space

Definition

Let $\Omega \neq \emptyset$ and $\mathcal{F}$ a σ-Algebra over Ω. The pair $(\Omega, \mathcal{F})$ is a *measurable space*.

Over a measurable space we can define a *measure*, which is a function having certain properties. Examples of measures are the geometric measures such as length, area, volume, etc., cardinality (in the case of finite sets), some physical quantities such as mass, as well as probability.

The event $\emptyset$ is called *the impossible event*, and Ω is called *the certain event*. Two events A and B are called *mutually exclusive* if $A \cap B = \emptyset$.

Axioms of probability

Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mathbb{P} : \mathcal{F} \longrightarrow \mathbb{R}$ is a *probability* over $\mathcal{F}$ if

1. For all events $E \in \mathcal{F}$, $\mathbb{P}(E) \geq 0$
2. $\mathbb{P}(\Omega) = 1$
3. For any countable sequence of mutually exclusive events (disjoint sets) $E_1, E_2, \ldots$ we have

$$\mathbb{P}\left(\bigcup_{i=0}^{\infty} E_i\right) = \sum_{i=0}^{\infty} \mathbb{P}(E_i)$$

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

Remarks

- A special case of the third axiom occurs when we have a finite collection of mutually exclusive events $E_1, E_2, \ldots, E_n$:

$$\mathbb{P}\left(\bigcup_{i=0}^{n} E_i\right) = \sum_{i=0}^{n} \mathbb{P}(E_i)$$

Examples of probability spaces

Example 9: Let $\Omega = \{1,2,3\}$, $\mathcal{F}_1 = \{\emptyset, \{1\}, \{2,3\}, \Omega\}$ and $\mathbb{P}$ defined as follows

$$\mathbb{P}(A) = \begin{cases} 1 & \text{if } 3 \in A \\ 0 & \text{if } 3 \notin A \end{cases}$$

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Example 10: Let $\Omega = \{1,2\}$, $\mathcal{F}_1 = \mathcal{P}(\Omega)$ and $\mathbb{P}$ defined as follows

$$\mathbb{P}(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \frac{1}{3} & \text{if } A = \{1\} \\ \frac{2}{3} & \text{if } A = \{2\} \\ 1 & \text{if } A = \{1,2\} \end{cases}$$

Then $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.

Consequences 1

- $\mathbb{P}(\emptyset) = 0$
- If $A, B \in \mathcal{F}$ and $A \cap B = \emptyset$ then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
- $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$
- $\mathbb{P}(A) \leq 1$ for all $A \in \mathcal{F}$
- If $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$ and $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A)$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- $$\begin{aligned} \mathbb{P}(A \cup B \cup C) = {} & \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) \\ & - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\ & + \mathbb{P}(A \cap B \cap C) \end{aligned}$$

Proofs - I

First property: $\mathbb{P}(\emptyset) = 0$

$$\Omega = \Omega \cup \emptyset \text{ and } \Omega \cap \emptyset = \emptyset, \quad \text{hence}$$

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset).$$

But since $\mathbb{P}(\Omega) = 1$ we conclude that $\mathbb{P}(\emptyset) = 0$.

Third property (complement rule): $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup \overline{A})$$

Since A and $\overline{A}$ are disjoint, we have

$$\mathbb{P}(A \cup \overline{A}) = \mathbb{P}(A) + \mathbb{P}(\overline{A})$$

Proofs - II

(Partial) fifth property (monotonicity): If $A \subseteq B$ then $\mathbb{P}(A) \leq \mathbb{P}(B)$

For any two events A and B we have

$$B = B \cap \Omega = B \cap (A \cup \overline{A}) = (B \cap A) \cup (B \cap \overline{A})$$

where $B \cap A$ and $B \cap \overline{A}$ are disjoint.

Now, by the definition of inclusion of sets we have

$$A \subseteq B \implies A \cap B = A$$

Therefore

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

Now, since $\mathbb{P}(B \cap \overline{A}) \geq 0$, we have

$$\mathbb{P}(A) \leq \mathbb{P}(B)$$

Consequences 2 - Bonferroni's inequality

Bonferroni's inequality:

If $A_1, A_2, \ldots, A_n \in \mathcal{F}$ then

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} \mathbb{P}(A_i) - (n-1)$$

Carlo Emilio Bonferroni, 1892–1960

Proof: By induction.

Laplacian spaces

Definition

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called Laplacian if Ω is finite, $\mathcal{F} = 2^{\Omega}$, and $\mathbb{P}(\omega) = \dfrac{1}{|\Omega|}$ for all $\omega \in \Omega$.

In a Laplacian space, the probability measure is also called Laplacian, or uniform, or classical – Pierre Simon de Laplace, 1749 – 1827.

In a Laplacian space, if $A \subseteq \Omega$, then

$$\mathbb{P}(A) = \frac{\text{Num. of favorable cases for } A \text{ to occur}}{\text{Total number of cases}} = \frac{|A|}{|\Omega|}$$

Whence the need to count cases (Combinatorics)

Example 11 – Laplacian probability spaces

Problem: Tom and Jerry sit randomly at a round table with n seats (with no distinguished seat), together with other $n-2$ people. What is the probability that they sit side by side?

Answer: This is a Laplacian probability space where Ω is the set of all distinct arrangements of the n guests. Our event of interest, A, consists of all the arrangements where Tom and Jerry are sitting side by side. Then

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{2!(n-2)!}{(n-1)!} = \frac{2}{n-1}$$

Example 12 – Laplacian probability spaces

Problem: A fair coin is tossed three consecutive times. This is a Laplacian probability space with $\Omega = \{(x, y, z) : x, y, z \in \{\text{HEADS}, \text{TAILS}\}\}$. Let A, B be the events

- $A =$"The first toss is HEADS "
- $B =$"The third toss is TAILS "

Calculate the probability of the following events: $A \cap B$, $A \cup B$, $\overline{A}$ and $\overline{A} \cap \overline{B}$.

Answer:

$$\mathbb{P}(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{2}{8} = \frac{1}{4}$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = \frac{4}{8} + \frac{4}{8} - \frac{1}{4} = \frac{3}{4}$$

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A) = 1 - \frac{1}{2} = \frac{1}{2}$$

Example 13 - part 1

Problem: A fair coin is tossed until two consecutive heads appear. What is the probability P_k that the number of tosses is k?

Answer: Obviously, the probability that $k = 1$ is 0. The probability that $k = 2$ is $\dfrac{1}{4}$. For $k \geq 3$ we see that the sequence of tosses must either start with TAILS or with HEADS–TAILS. So, a sequence of length k is either a sequence of length $k - 1$ preceded by TAILS, or a sequence of length $k - 2$ preceded by HEADS–TAILS. The probability that the first toss is TAILS is $\dfrac{1}{2}$, and the probability that the first two tosses correspond to the combination HEADS–TAILS is $\dfrac{1}{4}$. All together, we get the recurrence

$$P_k = \frac{1}{2}P_{k-1} + \frac{1}{4}P_{k-2},$$

with $P_1 = 0$ and $P_2 = \dfrac{1}{4}$.

Example 13 - part 2

The characteristic equation is $x^2 - \frac{1}{2}x - \frac{1}{4} = 0$, whose roots are $\frac{\varphi}{2}$ and $\frac{\hat{\varphi}}{2}$ (**Check !!**). Hence

$$P_k = \alpha \left(\frac{\varphi}{2}\right)^k + \beta \left(\frac{\hat{\varphi}}{2}\right)^k,$$

where $\alpha = \frac{1}{\sqrt{5}\varphi}$ and $\beta = -\frac{1}{\sqrt{5}\hat{\varphi}}$. Thus,

$$P_k = \frac{\varphi^{k-1} - \hat{\varphi}^{k-1}}{2^k\sqrt{5}} = \frac{F_{k-1}}{2^k},$$

where F_i is the i-th Fibonacci number (**Check !!**). By the way, we already saw that $\sum_{k=1}^{\infty} \frac{F_k}{2^k} = 2$. Dividing the whole sum by 2 we get that $\sum_{k=2}^{\infty} \frac{F_{k-1}}{2^k} = 1$, as we should expect.

Geometric probability

Let's now suppose that there is a geometric measure m defined over $(\Omega, \mathcal{F})$ – e.g. length, area, volume, etc. Then we can define a probability measure $\mathbb{P}$ over $(\Omega, \mathcal{F})$ as follows: For any event $A \in \mathcal{F}$,

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

Example 14 – The meeting

Problem: Tom and Jerry want to meet at a park between 6:00 PM and 7:00 PM, and they have agreed not to wait for one another more than 10 minutes. What is the probability that they will meet?

Note: We assume that both of them arrive independently at random at any moment within the time frame that has been agreed.

Solution: We may consider that our sample space Ω is the set of all pairs (x, y), with $0 \le x, y \le 60$, where x represents Tom's arrival, and y represents Jerry's arrival. They will meet if and only if $x - 10 \le y \le x + 10$.

The meeting – Graphical representation

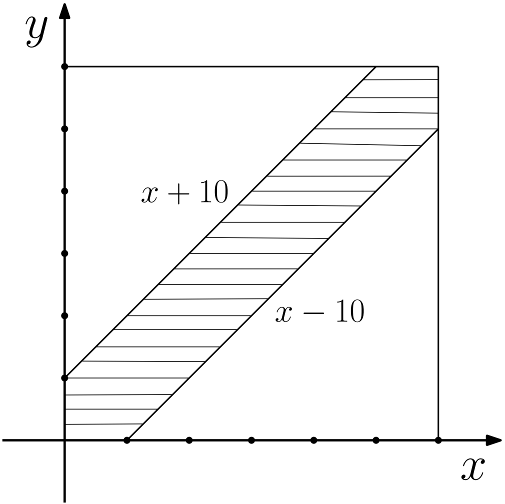

Figure: Graphical representation of the meeting

The meeting – Solution

Thus, Ω is the 60×60 square, and our event of interest, A, is represented by the diagonal shaded region. Hence

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

The area of Ω is obviously equal to 3600 units (min^2). As for A, the easiest way to calculate the area of the shaded region is by calculating the area of the two empty triangles, and then use the formula $\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A})$. Now, the two empty triangles can be combined to form a 50×50 square, hence $\text{area}(\overline{A}) = 2500$ min^2. Finally,

$$\mathbb{P}(A) = 1 - \frac{2500}{3600} = \frac{1100}{3600} = \frac{11}{36} \approx 0,3055$$

Example 15 – Buffon's needle

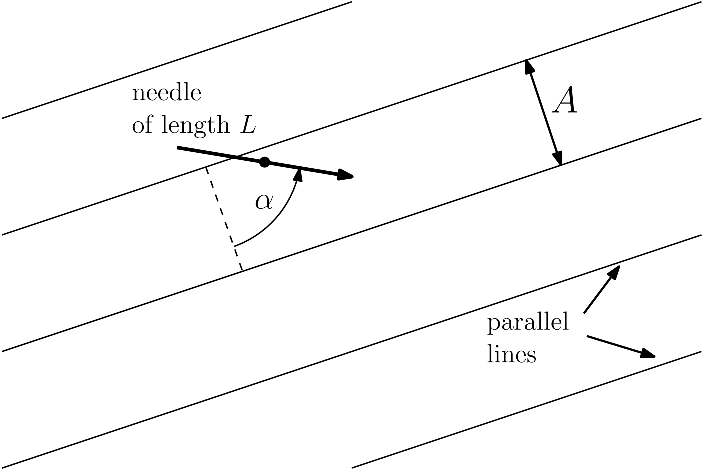

Figure: Buffon's experiment

Buffon's needle – Analysis

We throw the needle at random, and we want to know the probability that it intersects some line (George-Louis Leclerc, comte de Buffon, 1707 – 1788).

Let D denote the distance between the center of the needle and the nearest line, and α be the angle that forms the needle with the normal to the line. We may assume without loss of generality that all angles in $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right]$ are equally likely, as well as all the distances $D \in \left[0; \frac{A}{2}\right]$.

The needle intersects its nearest line iff D is less than or equal to the projection of one half of the needle onto the normal to the lines. I.e.

$$D \leq \frac{L}{2} \cos \alpha$$

Buffon's needle – Geometric representation

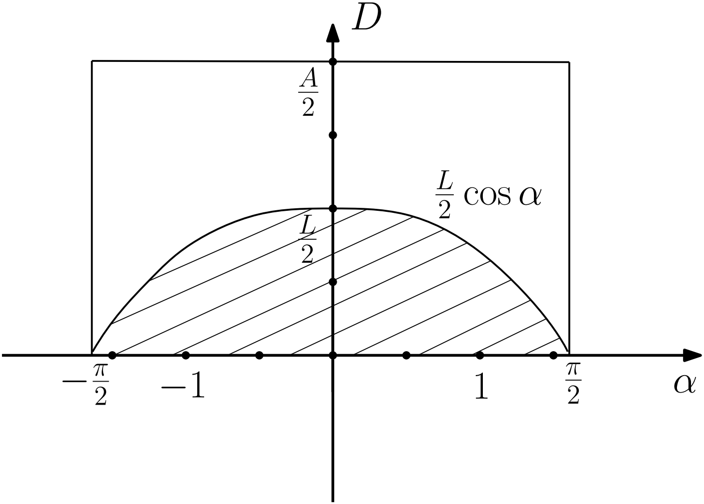

Figure: Geometric representation of the intersection probability

Buffon's needle – Solution

Thus, the probability that the needle intersects some line is the area under the curve divided by the total area of the rectangle. Now, the area under the curve is

$$\frac{L}{2}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos\alpha \, d\alpha = \frac{L}{2}\sin\alpha\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = L$$

Therefore, the desired probability is

$$P = \frac{L}{\frac{A}{2}\pi} = \frac{2L}{A\pi}$$

1. Axiomatic definition of probability
   - Sample space, event space, and probability axioms
   - Some consequences of the axioms
   - Laplacian spaces
   - Geometric probability
   - Monte Carlo methods

2. Conditional probability and independence
   - Independence
   - Total probability
   - Bayes' rule

Monte Carlo approximation of π

The previous solution to Buffon's needle problem suggests a mechanical method for computing π. Indeed, from the previous equation we get

$$\pi = \frac{2L}{AP},$$

therefore, if we can estimate P we also get an estimate for π.

We can estimate P by running the experiment several times, and dividing the number of times that the needle intersects some line by the total number of throws. This is an example of a **Monte Carlo method**.

The Italian mathematician Mario Lazzarini performed Buffon's needle experiment in 1901. He tossed the needle 3 408 times, obtaining the approximation $\pi \approx \dfrac{355}{113}$.

Another Monte Carlo method for computing π

- Construct a unit square and inscribe a circular sector of radius one, as shown here.
- Uniformly generate random points in the unit square.
- Count the number of points inside the circular sector, i.e. those having distance < 1 to the origin.
- The ratio of the inside-count and the total num. of points is an estimate of the ratio of the two areas, which is $\frac{\pi}{4}$.

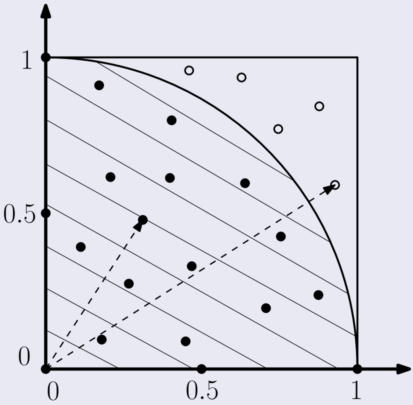

Figure: Monte Carlo method for π

Another application – Monte Carlo integration

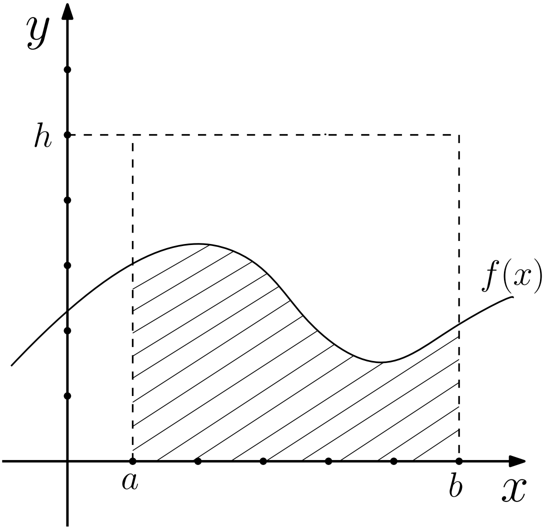

Figure: Area under the curve of $f(x)$

Monte Carlo integration procedure

1. Generate a set of N random points x_i uniformly distributed over the interval of integration $[a; b]$.
2. $\int_a^b f(x)dx \approx \frac{(b-a)}{N} \sum_{i=1}^{N} f(x_i)$
3. In general, the larger the number of points, the greater the accuracy.
4. We could devise an 'adaptive' procedure, that keeps generating points and updating the approximation until a desired level of accuracy is achieved.
5. For improper integrals other strategies must be devised (see for instance Monte Carlo integration: Improper integrals).

Conditional probability I

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A, B \in \mathcal{F}$, with $\mathbb{P}(A) > 0$. We define the probability of event B under the condition A as follows:

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

Remarks:

1. A now plays the role of Ω
2. If $\mathbb{P}(A) = 0$ then the conditional probability $\mathbb{P}(B|A)$ is undefined

Conditional probability II

1. We can also write $\mathbb{P}(A \cap B) = \mathbb{P}(A)\, \mathbb{P}(B|A)$
2. Three events: $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\, \mathbb{P}(B|A)\, \mathbb{P}(C \mid A \cap B)$
3. n events:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\, \mathbb{P}(A_2|A_1)\, \mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}\left(A_n | \bigcap_{i=1}^{n-1} A_i\right)$$

Example 16 – Conditional probability

An experiment consists of casting two fair dice. This is a Laplacian space with $\Omega = \{(i;j) : 1 \leq i, j \leq 6\}$. Let A and B be the events

- A: "The outcome of both dice is different"
- B: "At least one of the outcomes is 6"

We have

- $\mathbb{P}(A) = \frac{30}{36}$
- $\mathbb{P}(B) = \frac{11}{36}$
- $\mathbb{P}(A \cap B) = \frac{10}{36}$

Therefore

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\frac{10}{36}}{\frac{30}{36}} = \frac{1}{3}$$

Independence – Two events

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A, B \in \mathcal{F}$. A and B are called **independent** if

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

Alternatively, $\mathbb{P}(B|A) = \mathbb{P}(B)$ (and also $\mathbb{P}(A|B) = \mathbb{P}(A)$).

Independence – Multiple events I

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $A = \{A_1, A_2, \ldots, A_n\}$ a finite set of events. The events of A are called **pairwise independent** if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\,\mathbb{P}(A_j) \text{ for all } 1 \leq i, j \leq n,\ i \neq j$$

The events of A are called **mutually independent** (or simply, independent) if every event of A is independent of any intersection of the other events, i.e. if for every subset of indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$$

Example 17 – Three events

Three events, A, B and C, are independent (mutually independent) iff

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \mathbb{P}(A)\,\mathbb{P}(B)\\
\mathbb{P}(A \cap C) &= \mathbb{P}(A)\,\mathbb{P}(C)\\
\mathbb{P}(B \cap C) &= \mathbb{P}(B)\,\mathbb{P}(C)\\
\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C)
\end{aligned}$$

Example 17 – Three events (continued)

Remark: If A, B, C are pairwise independent, they are not necessarily independent.

Example: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a Laplacian probability space, with $\Omega = \{1, \ldots, 8\}$, and let $A = \{1, 2, 3, 4\}$, $B = \{2, 4, 6, 8\}$, and $C = \{1, 3, 6, 8\}$. Then

$$\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2},$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \cap C) = \mathbb{P}(B \cap C) = \frac{1}{4}.$$

Therefore, they are pairwise independent. However,

$$\mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C) = \frac{1}{8} \quad \text{and} \quad \mathbb{P}(A \cap B \cap C) = 0.$$

Therefore, they are not mutually independent.

Consequence of independence – I

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $A, B \in \mathcal{F}$ be two incompatible events with non-zero probability, i.e. $A \cap B = \emptyset$, and $\mathbb{P}(A) \neq 0$ and $\mathbb{P}(B) \neq 0$. Then A and B cannot be independent.

Proof: (By reductio ad absurdum) A and B are independent if $\mathbb{P}(A|B) = \mathbb{P}(A)$. However,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = 0,$$

since $\mathbb{P}(A \cap B) = 0$. That means $\mathbb{P}(A) = 0$, which is a contradiction.

Consequence of independence – II

Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $A, B \in \mathcal{F}$ be two independent events. Then

- *A and $\overline{B}$ are independent (and also $\overline{A}$ and B, by symmetry)*
- *$\overline{A}$ and $\overline{B}$ are independent*

Proof: We only need to prove that A and $\overline{B}$ are independent. We have

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap \overline{B}) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A \cap \overline{B}) + \mathbb{P}(A)\,\mathbb{P}(B) \end{aligned}$$

Therefore

$$\mathbb{P}(A \cap \overline{B}) = \mathbb{P}(A)\,(1 - \mathbb{P}(B)) = \mathbb{P}(A)\,\mathbb{P}(\overline{B})$$

1 Axiomatic definition of probability
- Sample space, event space, and probability axioms
- Some consequences of the axioms
- Laplacian spaces
- Geometric probability
- Monte Carlo methods

2 Conditional probability and independence
- Independence
- Total probability
- Bayes' rule

Total probability

Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{A_i\}_{i\geq 1}$ be a finite or countably infinite collection of subsets of Ω (with $A_i \in \mathcal{F}$ for all i). Then $\{A_i\}_{i\geq 1}$ is a *partition* of Ω if

1. $A_i \neq \emptyset$ for all i
2. $A_i \cap A_j = \emptyset$ for $i \neq j$
3. $\bigcup_{i\geq 1} A_i = \Omega$

Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{A_i\}_{i\geq 1}$ a partition of Ω, and $B \in \mathcal{F}$. Then

$$\mathbb{P}(B) = \sum_{i\geq 1} \mathbb{P}(B \cap A_i) = \sum_{i\geq 1} \mathbb{P}(B|A_i)\,\mathbb{P}(A_i)$$

Example 18 – Total probability

Problem:

We have three urns with 100 balls each, distributed as follows:

1. Urn 1 contains 75 red balls and 25 blue balls
2. Urn 2 contains 60 red balls and 40 blue balls
3. Urn 3 contains 45 red balls and 55 blue balls

We choose one of the urns at random, and then pick a ball from the chosen urn, also at random. What is the probability that the ball is red?

Example 18 – Total probability (continued)

Solution:

The three urns form a partition of the set of balls, since the urns are disjoint, and all the balls are in one of the three urns. Let U_i be the event that the i-th urn is chosen, and R the event that the extracted ball is red. We know that

$$\mathbb{P}(U_i) = \frac{1}{3}, \quad \mathbb{P}(R|U_1) = \frac{75}{100}, \quad \mathbb{P}(R|U_2) = \frac{60}{100}, \quad \mathbb{P}(R|U_3) = \frac{45}{100}$$

Hence

$$\begin{aligned}\mathbb{P}(R) &= \sum_{i=1}^{3} \mathbb{P}(U_i)\,\mathbb{P}(R|U_i) \\ &= \frac{1}{3}\cdot\frac{75}{100} + \frac{1}{3}\cdot\frac{60}{100} + \frac{1}{3}\cdot\frac{45}{100} \\ &= \frac{60}{100} = 0.6\end{aligned}$$

Graphical representation

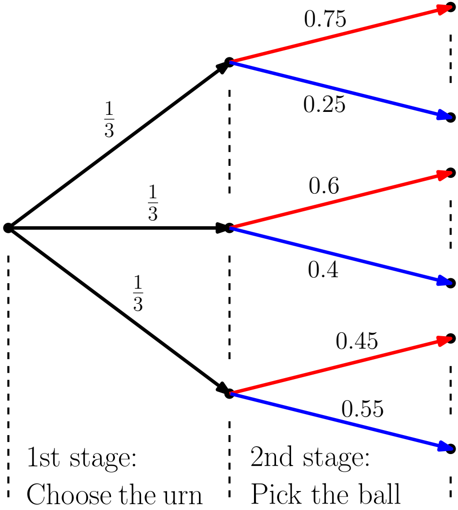

Figure: Tree diagram for the urn problem

Bayes' rule

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $A, B \in \mathcal{F}$.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\ \mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\ \mathbb{P}(A)}{\mathbb{P}(B|A)\ \mathbb{P}(A) + \mathbb{P}(B|\overline{A})\ \mathbb{P}(\overline{A})}$$

Generalization: If $\{A_i\}_{i\geq 1}$ is a partition of Ω, and $B \in \mathcal{F}$, then

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B|A_k)\ \mathbb{P}(A_k)}{\sum_{i\geq 1} \mathbb{P}(B|A_i)\ \mathbb{P}(A_i)}$$

Bayes' rule allows us to calculate the probability of an event *a posteriori*.

Thomas Bayes, 1702 – 1761

Proof of Bayes' rule for the case of two events

Proof:

We know that

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

But also

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

Equating the two expressions above we get

$$\mathbb{P}(A)\,\mathbb{P}(B|A) = \mathbb{P}(B)\,\mathbb{P}(A|B)$$

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

Prove it for the case when Ω is the disjoint union of A_1 and A_2.

Example 19 – Bayes' rule

Problem:

Again we have the three urns with red and blue balls (see the example of total probability). Suppose that somebody else has gone through the procedure, and the extracted ball turns out to be red. What is the probability that Urn # 1 was chosen?

Solution:

We know $\mathbb{P}(R|U_i)$ for all $1 \leq i \leq 3$, but now we are interested in $\mathbb{P}(U_1|R)$, so this is precisely a scenario for Bayes' rule:

$$\mathbb{P}(U_1|R) = \frac{\mathbb{P}(R|U_1)\,\mathbb{P}(U_1)}{\mathbb{P}(R)} = \frac{\frac{75}{100} \cdot \frac{1}{3}}{\frac{6}{10}} = \frac{5}{12}$$

Remark: $\mathbb{P}(R)$ must be calculated by the law of total probability, but we did that before. Note also that $\mathbb{P}(U_1|R) > \dfrac{1}{3}$.

Example 20 – The Monty Hall problem

- The name comes from a popular TV show.
- There are three doors, and behind one of them there is a reward.
- The contestant chooses a door, and before opening it, the host opens another door that he knows has no reward, and asks the contestant if he/she wants to switch.
- **Question**: Should the contestant switch?

Example 20 – The Monty Hall problem (continued)

Let's define the events in chronological order:

- C_i is the event "The CAR is behind door i"
- D_i is the event "The contestant chooses DOOR i"
- H_i is the event "The HOST opens door i"

A priori we may assume that $\mathbb{P}(C_i) = \mathbb{P}(D_i) = \dfrac{1}{3}$ for $1 \leq i \leq 3$, and D_i is independent from C_j for all $1 \leq i, j \leq 3$.

So, let's suppose that the contestant picks a particular door, say D_1, and then the host opens door number 3, showing there is no prize. We want to compare

$$\mathbb{P}(C_1|H_3) = \frac{\mathbb{P}(H_3|C_1)\,\mathbb{P}(C_1)}{\mathbb{P}(H_3)} \quad \text{and} \quad \mathbb{P}(C_2|H_3) = \frac{\mathbb{P}(H_3|C_2)\,\mathbb{P}(C_2)}{\mathbb{P}(H_3)}$$

Since these are the only possible scenarios, $\mathbb{P}(C_1|H_3) + \mathbb{P}(C_2|H_3) = 1$.

Example 20 – The Monty Hall problem (continued)

Obviously, we don't know $\mathbb{P}(H_3)$, but it's the same in both formulas. We also know that $\mathbb{P}(C_1) = \mathbb{P}(C_2) = \frac{1}{3}$. Hence we just have to compare $\mathbb{P}(H_3|C_1)$ and $\mathbb{P}(H_3|C_2)$.

Now, $\mathbb{P}(H_3|C_1) = \frac{1}{2}$, since the host could have equally opened door number 2 or door number 3. None of them contains the prize anyway.

On the other hand, $\mathbb{P}(H_3|C_2) = 1$, since door number 3 would be the only choice left to the host.

So, $\mathbb{P}(C_2|H_3) = 2\,\mathbb{P}(C_1|H_3)$, and since they must add up to 1, we have that $\mathbb{P}(C_2|H_3) = \frac{2}{3}$ and $\mathbb{P}(C_1|H_3) = \frac{1}{3}$.

CONCLUSION: The contestant should switch to the other door in order to double the chances of winning the prize !!

Example 20 – The Monty Hall problem (continued)

This result looks counterintuitive, or even paradoxical to most people, who tend to think that the probabilities remain the same after the host opens a door, i.e. $\frac{1}{3}$, and stick to their original choice. There may also be some psychological reasons involved.

For the history and a more detailed explanation of the problem you can watch the videos

- `https://www.youtube.com/watch?v=4Lb-6rxZxx0&ab_channel=Numberphile` (in English), or
- `https://www.youtube.com/watch?v=1BpTBzDQuRE&ab_channel=DateunVlog` (in Spanish).

Example 21 – The false positive paradox

The false positive paradox – taken from `https://www.probabilitycourse.com/chapter1/1_4_3_bayes_rule.php`

A certain disease affects about 1 out of 10 000 people. There is a test to check whether the person has the disease. The test is quite accurate. In particular, we know that

- The probability that the test result is positive (suggesting the person has the disease), given that the person **does not have** the disease, is 2%
- The probability that the test result is negative (suggesting the person does not have the disease), given that the person **has** the disease, is 1%

A random person gets tested for the disease and the result is positive. What is the probability that the person actually has the disease?

Example 21 – Solution

Let D be the event that the person has the disease, and let T be the event that the test result is positive. We know that

$$\mathbb{P}(D) = \frac{1}{10\,000}, \quad \mathbb{P}(T|\overline{D}) = \frac{2}{100} = 0.02, \quad \mathbb{P}(\overline{T}|D) = \frac{1}{100} = 0.01$$

We want to compute $\mathbb{P}(D|T)$. Again, we use Bayes' rule:

$$\begin{aligned}
\mathbb{P}(D|T) &= \frac{\mathbb{P}(T|D)\,\mathbb{P}(D)}{\mathbb{P}(T|D)\,\mathbb{P}(D) + \mathbb{P}(T|\overline{D})\,\mathbb{P}(\overline{D})} \\
&= \frac{(1-0.01)\times 0.0001}{(1-0.01)\times 0.0001 + 0.02\times(1-0.0001)} \\
&= 0.0049
\end{aligned}$$

Example 21 – Some remarks

Thus, the chance that the person has the disease is less than 1%. This might seem somewhat counterintuitive as we know the test is quite accurate. The point is that the disease is also very rare. Thus, there are two competing forces here, and since the rareness of the disease (1 out of 10 000) is stronger than the accuracy of the test (98 or 99 percent), there is still good chance that the person does not have the disease.

Bibliography Probability I

1. Ash, R.B.: *Basic Probability Theory*. Dover Publications 2008 (in English) `https://faculty.math.illinois.edu/~r-ash/BPT.html`
2. Blanco Castañeda, L.: *Probabilidad*. Univ. Nacional de Colombia 2004 (in Spanish) `https://repositorio.unal.edu.co/handle/unal/53471?show=full`
3. Lueker, George S.: "Some techniques for solving recurrences", *Computing Surveys*, vol. 12, num. 4, 1980.
4. Tsitsiklis, J. *Introduction to Probability*, MIT Open Courseware 2018 (Video lectures in English). `https://ocw.mit.edu/courses/res-6-012-introduction-to-probability-spring-2018/`