Introduction to Basic Elements of Statistics

Stéphane Rivaud

INRIA Saclay

Septembre 12, 2024

Outline

Outline

What are Statistics ?

Probability of a coin toss: Heads or Tail ?

$$\mathbb{P}(\text{Head}) = 1 - \mathbb{P}(\text{Tail}) = \theta$$

Statistics: is the coin fair ?

$$\mathbb{P}(\theta \mid \text{Set of coin tosses}) = \frac{1}{2} \quad (?)$$

$\longrightarrow$ What can we deduce from partial observation of a phenomenon ?

Probabilities and Statistics are **dual mathematical frameworks**.

Cox-Jaynes Theorem: Quantifying plausibility

Hypothesis on the method:

- **Coherence** $\rightarrow$ If a result can be derived in many ways, all derivation should yield the same result.
- **Continuity of the method** $\rightarrow$ Changing the value of a parameter should not change the computation method.
- **Universality** $\rightarrow$ The computation method should be general and not tied to a specific case.

Hypothesis on the practitioner:

- **Unambiguous specifications** $\rightarrow$ a proposition can only be understood in a unique way.
- **No hidden information** $\rightarrow$ the algorithm is given all relevant information available.

$\Longrightarrow$ **Isomorph** to **probability** theory

Some key points

- What is ***Applied*** *Statistics* ?
- What are the kinds of **questions** we can answer with statistics ?
- What is the importance of **data** (vs statistical theory) ?
- Why is there so much **computer science** in statistics ?
- What is the difference between **Statistics** and **Machine Learning** ?

Example: Basics

Example: Medical Diagnosis

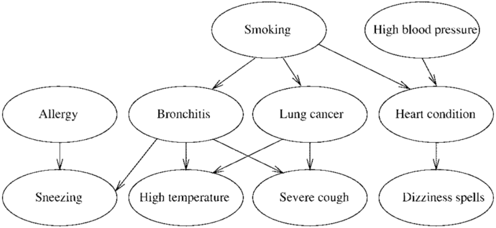
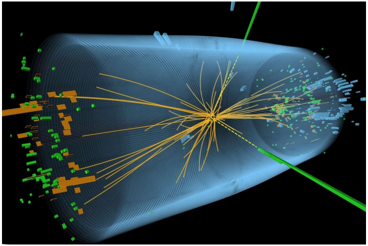


Figure: Bayesian network for differential diagnosis

Physics



Advanced Example: Natural Language Processing

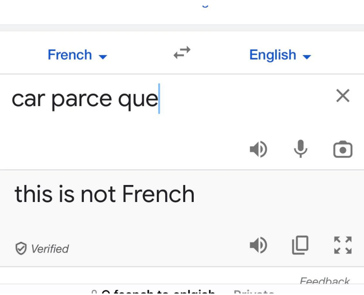


Figure: Automatic translation

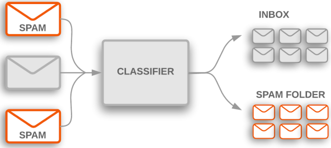


Figure: Spam detection

Outline

Events and Realization

Let Ω be the space of events.

- An event A is a subset of Ω.
- A realization of event A is an element $x \in A$.

Example:

- Dice rolls
- Next opponent move in Chess game.
- Result of a Chess game.

Event Space, Events, and Realizations

- In probability theory, we reason under uncertainty by associating events with probabilities.
- **Event Space** Ω: The set of all possible outcomes of a random process.
- **Event** $A \subseteq \Omega$: A subset of outcomes, representing an occurrence or a condition. The set $\mathcal{F} \subset 2^{\Omega}$ of all events is called a sigma-algebra.
- **Realization** of an event: The actual outcome from Ω, which either belongs to an event A or not.
- Example: If Ω represents the outcomes of a die roll, an event A could be "rolling an even number" (i.e., $A = \{2, 4, 6\}$).

Probabilities on Sets

- A probability measure $\mathbb{P}$ assigns a probability to each event $A \subseteq \Omega$.
- The function $\mathbb{P}$ satisfies the following axioms:
  1. $0 \leq \mathbb{P}(A) \leq 1$ for any event A.
  2. $\mathbb{P}(\Omega) = 1$ (the entire sample space has probability 1).
  3. If $A_1, A_2, \ldots$ are disjoint events, then:

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

- We can manipulate probabilities through basic rules:
  - $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ (complement rule)
  - $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ (union of events)

Probability Space

Definition. A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ where:

- Ω is the sample space; $\mathcal{F} \subseteq 2^{\Omega}$ is a σ-algebra;
- $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a probability measure, $\mathbb{P}(\Omega) = 1$, and $\mathbb{P}$ is countably additive.

Basic properties for events $A, B \in \mathcal{F}$

- Bounds: $0 \leq \mathbb{P}(A) \leq 1$, with $\mathbb{P}(\varnothing) = 0$ and $\mathbb{P}(\Omega) = 1$.
- Complement: $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
- Monotonicity: $A \subseteq B \;\Rightarrow\; \mathbb{P}(A) \leq \mathbb{P}(B)$.
- Union/intersection: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
- Disjoint additivity: if $A \cap B = \varnothing$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$; extends to countable disjoint unions.

Events are the measurable statements about outcomes: elements of $\mathcal{F}$.

Probability Space — Examples

Example: one coin toss

Fair coin (one toss): $\Omega = \{H, T\}$, $\mathcal{F} = 2^\Omega$, and $\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}$. For $A = \{H\}$, $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(A^c) = \frac{1}{2}$, $\mathbb{P}(A \cup A^c) = 1$, $\mathbb{P}(A \cap A^c) = 0$.

Example: two fair dice (ordered outcomes)

Two fair dice: $\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$ with $\mathbb{P}(\{\omega\}) = \frac{1}{36}$ for each ordered outcome. Let A be "the sum is 7" and B be "the first die is even".

$$A = \{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}$$
$$B = \{(i,j) : i \in \{2,4,6\},\ j \in \{1, \dots, 6\}\}.$$

Then $\mathbb{P}(A) = \frac{6}{36} = \frac{1}{6}$, $\mathbb{P}(B) = \frac{18}{36} = \frac{1}{2}$, and $\mathbb{P}(A \cap B) = \frac{3}{36} = \frac{1}{12}$, which are computed by counting outcomes in the sample space and can be used with the union/intersection formulas above.

Outline

Conditional Probability

Definition

For two events A and B with $\mathbb{P}(B) > 0$, the *conditional probability* of A given B is defined as

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

- Intuition: restrict the sample space to B and renormalize probabilities.
- Basic identity (rearrangement):

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

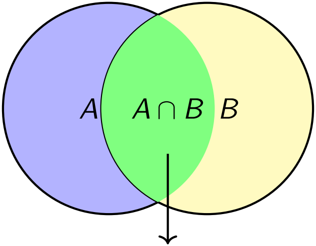
- Chain rule (useful for multiple events):

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

- Law of total probability (partition $\{B_i\}$ of Ω with $\mathbb{P}(B_i) > 0$):

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i).$$

Conditional probability — visual



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

Visual takeaway

Conditioning on B means we restrict attention to the region B (the right circle) and ask what fraction of B is occupied by $A \cap B$ (the shaded overlap). The conditional probability is that fraction:

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

Conditional probability — a hands-on example

Setup (standard deck, 52 cards).
Let $A =$ "card is a King" and $B =$ "card is a face card".

Unconditional probabilities

- $\mathbb{P}(A) = \dfrac{4}{52} = \dfrac{1}{13}$
- $\mathbb{P}(B) = \dfrac{12}{52} = \dfrac{3}{13}$

Conditional probability (definition only)

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

Since $A \cap B =$ "King that is also a face card" has 4 outcomes,

$$\mathbb{P}(A \cap B) = \frac{4}{52} = \frac{1}{13} \quad \Rightarrow \quad \mathbb{P}(A \mid B) = \frac{\frac{1}{13}}{\frac{3}{13}} = \frac{1}{3}.$$

Conditional probability — a hands-on example

Intuition & takeaway

- The event B ("card is a face card") restricts the sample space to 12 cards: $\{J, Q, K\} \times \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$.
- Within this restricted space, there are 4 Kings, so $\mathbb{P}(A \mid B) = 4/12 = 1/3$.
- This shows how conditioning on an event can significantly change the probability of another event.

Bayes Theorem

Bayes Theorem

Let A and B be two events, with $\mathbb{P}(B) > 0$. Then:

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

- $\mathbb{P}(A \mid B)$: The probability of A given B (posterior).
- $\mathbb{P}(B \mid A)$: The probability of B given A (likelihood).
- $\mathbb{P}(A)$: The prior probability of A.
- $\mathbb{P}(B)$: The total probability of B.

Takeaway

- **Bayes Theorem** is a fundamental theorem in probability that allows us to update the probability of a hypothesis based on new evidence.
- It expresses how the probability of an event A, given another event B, is related to the reverse conditional probability $\mathbb{P}(B \mid A)$.

Example — Dice

Compute $\mathbb{P}(A \mid B)$ for the two-dice example

Recall: $A =$ "sum is 7", $B =$ "first die is even" with $\mathbb{P}(A) = 6/36$, $\mathbb{P}(B) = 18/36$, and $\mathbb{P}(A \cap B) = 3/36$.
Direct conditional formula:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{3/36}{18/36} = \frac{3}{18} = \frac{1}{6}.$$

Using Bayes' theorem (alternative view):

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{3/36}{6/36} = \frac{1}{2},$$

hence

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\,\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{\frac{1}{2} \cdot \frac{1}{6}}{\frac{1}{2}} = \frac{1}{6}.$$

Example — Medical diagnosis

Compute $\mathbb{P}(\text{Disease} \mid \text{Positive})$

Assume: $\mathbb{P}(\text{Disease}) = 0.001$, $\mathbb{P}(+ \mid \text{Disease}) = 0.99$, and $\mathbb{P}(+ \mid \text{No Disease}) = 0.01$.
First compute the marginal probability of a positive test (law of total probability):

$$\mathbb{P}(+) = \mathbb{P}(+ \mid \text{Disease})\mathbb{P}(\text{Disease}) + \mathbb{P}(+ \mid \text{No Disease})\mathbb{P}(\text{No Disease})$$

Numerically:

$$\mathbb{P}(+) = 0.99 \times 0.001 + 0.01 \times 0.999 = 0.00099 + 0.00999 = 0.01098.$$

Now apply Bayes' theorem:

$$\mathbb{P}(\text{Disease} \mid +) = \frac{\mathbb{P}(+ \mid \text{Disease})\,\mathbb{P}(\text{Disease})}{\mathbb{P}(+)} = \frac{0.99 \times 0.001}{0.01098} \approx 0.0902.$$

So a positive test yields about a 9.02% chance of actually having the disease.

Example — Medical diagnosis (negative test)

Compute $\mathbb{P}(\text{Disease} \mid \text{Negative})$

Using the same assumptions as before: $\mathbb{P}(\text{Disease}) = 0.001$, $\mathbb{P}(+ \mid \text{Disease}) = 0.99$, $\mathbb{P}(+ \mid \text{No Disease}) = 0.01$. Thus $\mathbb{P}(- \mid \text{Disease}) = 0.01$ and $\mathbb{P}(- \mid \text{No Disease}) = 0.99$. Compute the marginal for a negative test:

$$\mathbb{P}(-) = \mathbb{P}(- \mid \text{Disease})\mathbb{P}(\text{Disease}) + \mathbb{P}(- \mid \text{No Disease})\mathbb{P}(\text{No Disease})$$

Numerically:

$$\mathbb{P}(-) = 0.01 \times 0.001 + 0.99 \times 0.999 = 0.00001 + 0.98901 = 0.98902.$$

Apply Bayes' theorem:

$$\mathbb{P}(\text{Disease} \mid -) = \frac{\mathbb{P}(- \mid \text{Disease})\,\mathbb{P}(\text{Disease})}{\mathbb{P}(-)} = \frac{0.01 \times 0.001}{0.98902} \approx 1.011 \times 10^{-5}.$$

This is about 0.00101% (very small): a negative test makes the disease extremely unlikely.

Outline

Independence Among Events

Definition

Two events A and B are independent if the occurrence of one does not affect the probability of the other. Formally:

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

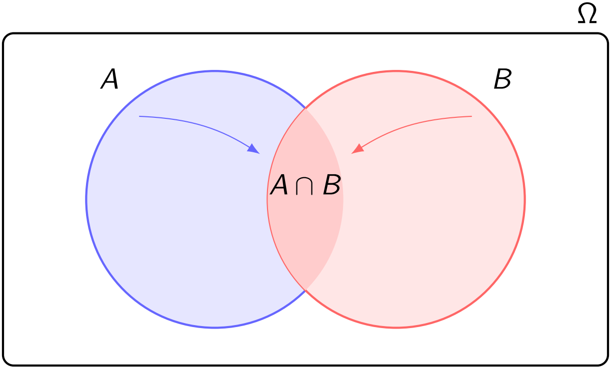
Equivalently (when $\mathbb{P}(A), \mathbb{P}(B) > 0$):

$$\mathbb{P}(B \mid A) = \mathbb{P}(B) \quad \text{and} \quad \mathbb{P}(A \mid B) = \mathbb{P}(A).$$

Takeaway

- Independence means that knowing A occurred gives no information about whether B occurred, and vice versa.
- Independence simplifies computations: intersections factor into products.
- In practice, check independence via $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ or the conditional form.

Independence Among Events — Visual



$\mathbb{P}(A) = 0.5 \quad \mathbb{P}(B) = 0.5 \quad \mathbb{P}(A \cap B) = 0.25 \quad A \perp\!\!\!\perp B$

Visual intuition: events are subsets of Ω; independence here matches

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

Independence of Events — Coin Toss Example

Experiment: Toss two fair coins.

Sample space (equiprobable outcomes):

$$\{HH, HT, TH, TT\}, \quad \mathbb{P}(\text{each}) = \tfrac{1}{4}.$$

Joint probability table:

	Second = H	Second = T	Total
First = H	1/4	1/4	1/2
First = T	1/4	1/4	1/2
Total	1/2	1/2	1

Check independence:

- $A =$ "First coin is Heads" $\Rightarrow$ $\mathbb{P}(A) = 1/2$.
- $B =$ "Second coin is Heads" $\Rightarrow$ $\mathbb{P}(B) = 1/2$.
- Joint: $\mathbb{P}(A \cap B) = 1/4 = (1/2)(1/2)$.
- $\Rightarrow$ A and B are **independent**.

Non-Independence of Events — Coin Toss Example

Experiment: Toss two fair coins.

Sample space (equiprobable outcomes):

$$\{HH, HT, TH, TT\}, \quad \mathbb{P}(\text{each}) = \tfrac{1}{4}.$$

Joint probability table:

	At least one Head	No Head (TT)	Total
First = H	1/2	0	1/2
First = T	1/4	1/4	1/2
Total	3/4	1/4	1

Check independence:

- $A =$ "First coin is Heads" $\Rightarrow \mathbb{P}(A) = 1/2$.
- $D =$ "At least one Head" $\Rightarrow \mathbb{P}(D) = 3/4$.
- Joint: $\mathbb{P}(A \cap D) = 1/2$.
- Product: $\mathbb{P}(A)\,\mathbb{P}(D) = (1/2)(3/4) = 3/8$.
- Since $1/2 \neq 3/8$, A and D are **not independent**.

Mutual Independence — Definition and Intuition

Definition

Formal definition (arbitrary family): Let $\{A_i\}_{i \in I}$ be a (possibly infinite) family of events on $(\Omega, \mathcal{F}, \mathbb{P})$. We say $\{A_i\}_{i \in I}$ is **mutually independent** if for every finite, nonempty subset $J \subseteq I$ with $|J| \geq 2$,

$$\mathbb{P}\Big(\bigcap_{j \in J} A_j\Big) = \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}) = \prod_{j \in J} \mathbb{P}(A_j).$$

Equivalently: all finite subcollections are independent (the product rule holds for every finite intersection).

- No matter which *finite* subset of events you look at, knowing some of them occurred does not change the probability of any combination of the others.
- This is stronger than pairwise independence: it requires factorization for all k-way intersections ($k = 2, 3, \dots$) drawn from the family.
- For infinite families, mutual independence is still checked via *all finite* subfamilies (not an infinite intersection).

Pairwise vs. Mutual Independence

Independence of two events:

$$A, B \text{ independent} \iff \mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B).$$

Pairwise independence of a family:

- Every *pair* (A_i, A_j) is independent.
- Only checks 2-event intersections.

Mutual independence of a family:

- Stronger condition: independence holds for *every sub-collection*, not only pairs.
- Requires all k-way intersections to factorize, for all $k = 2, 3, \ldots, n$.

Key takeaway

Pairwise independence does *not* imply mutual independence.

Example — Pairwise but not Mutual Independence

Experiment: Toss two fair coins.

- $A =$ "First coin is Heads".
- $B =$ "Second coin is Heads".
- $C =$ "The two coins show the same result".

Check pairwise independence:

- $\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(C) = \frac{1}{2}$.
- $\mathbb{P}(A \cap B) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(B)$.
- $\mathbb{P}(A \cap C) = \frac{1}{4} = \mathbb{P}(A)\mathbb{P}(C)$.
- $\mathbb{P}(B \cap C) = \frac{1}{4} = \mathbb{P}(B)\mathbb{P}(C)$.
- $\Rightarrow A, B, C$ are pairwise independent.

But not mutually independent:

$$\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \quad \neq \quad \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C) = \frac{1}{8}.$$

Summary

What we covered

This lesson introduced the foundational concepts of probability theory that underpin statistical analysis.

- We motivated the study of statistics and its applications.
- We introduced event spaces, events, and their realizations.
- We discussed how to define and manipulate probabilities on sets.
- We introduced Bayes theorem, which allows us to update probabilities with new information.
- We explained the concept of independence between events.

References

- Textbook: "Introduction to Probability and Statistics" by Mendenhall, Beaver, and Beaver.
- Online resources:
  `https://www.khanacademy.org/math/statistics-probability`