Data Science Fundamentals

Part I: Probability theory

ISEP 2nd year
2023-2024

Based on the course given by Nathalie Colin & Jean-Claude Guillerot

Probability theory

- **1st session (September 29th 2023):**
- **Chapter 1: NOTION OF EVENT AND OF PROBABILITY**
- **Chapter 2: CONDITIONAL PROBABILITIES & INDEPENDENCE**

Chapitre 1: Introduction – Main definitions

- Introduction
- Three definitions of probability
  - Frequentist probability
  - Definition based on the number of outcomes
  - Definition based on axioms
- Sets and events
  - Reminder of set algebra
  - Corollaries of the axioms

Introduction

Probability : $\neq$ Certainty

Probability : Degree of certainty that an event will occur

Historically …

From VIII and XIII centuries, Arab mathematicians studied cryptography making the first use of permutations and combinations to list all possible Arabic words with and without vowels.

The modern mathematical theory of probability has its roots in attempts to analyze games of chance by Gerolamo Cardano in the XVI century, and by Pierre de Fermat and Blaise Pascal in the XVII century.

Initially, probability theory mainly considered discrete events, and its methods were mainly combinatorial. Later, analytical considerations compelled the incorporation of continuous variables.

Nowadays, It is present in many scientific branches such as *Economics, Physics (statistics), Genetics, Communications, Computer science, etc.*

Three definitions of probability

- Frequentist probability
- Definition based on the number of outcomes
- Definition based on axioms

Vocabulary

We are interested in a random experiment which produces a single outcome among a finite number of possible outcomes denoted: ω_1, ω_2, ..., ω_n. We denote Ω the set of all possible results (sample space):

$$\Omega=\{\omega_1,\omega_2,\ldots,\omega_n\}$$

We call event a subset A of Ω

Example:

- Experiment: rolling a 6-sided-die
- Possible outcomes: $\Omega = \{1,2,3,4,5,6\}$
- Event: A = {the result is even}

We denote the impossible event the empty set $\emptyset$ and the certain event will be the set Ω.

Definition 1 : Frequentist probability

The relative frequency f_k of the outcome ω_k is the ratio of the number n_k of experiments whose result is ω_k to the total number of experiments n: $f_k = \frac{n_k}{n}$

The probability of an event A, denoted P(A), is the limit of its relative frequency in many trials.

$$\boxed{P(A) = \lim_{n \to \infty} \left(\frac{n_A}{n} \right)}$$

Where : n_A : number of times the result of the experiment is the event A

n_A/n : relative frequency of event A

Advantages / Disadvantages of this definition:

- It satisfies the intuitive notion of probability

- It measures the probability of an event using repeated trials

- The expression contains a limit.

Definition 2 : Definition based on the number of outcomes

$$P(A) = \frac{N_A}{N} = \frac{\text{number of outcomes for event A}}{\text{total number of outcomes in the sample space}}$$

All outcomes in the sample space must be equally likely.

Example :*Experiment -> Roll a die*

Event -> A = {1}

Number of possible outcomes = 6

Number of outcomes for event A = 1 then P(A) = 1/6

Avantages/Disadvantages of this definition

- Very simple definition
- Counting might be impossible, for instance, when the set of outcomes is infinity!!!

Definition 3 : Definition based on axioms

Given an experiment, the probability of event A, denoted P(A) is a real number between 0 and 1 that satisfies :

Axiom 1 : P(A) is positive or null.

Axiom 2 : If A is the certain ***Ω*** event : P(A) = 1

Axiom 3 : Additivity

If A and B are disjoint, i.e. $A \cap B = \emptyset$,

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

This property can be extended to an infinite set of disjoint events Ai

$$i \neq j \quad (A_i \cap A_j) = \emptyset \implies P\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} P(A_i)$$

Set Algebra ! => very useful in probability theory

Correspondence between *Set algebra ⇔ Probability theory*

- Sets ⇔ Events
- Elements of sets ⇔ any possible outcome of an experiment
- Empty set $\emptyset$ ⇔ impossible event
- Ω : Universe set or power set⇔ sample space (set of all possible outcomes)

Reminder : Set algebra

The union	∪
The intersection	∩
The Power set	Ω
The empty set	$\varnothing$
The complement of A	$\bar{A}$
Commutative laws	$A \cup B = B \cup A$ and $A \cap B = B \cap A$
Associative laws	$(A \cup B) \cup C = A \cup (B \cup C)$ and $(A \cap B) \cap C = A \cap (B \cap C)$
Distributive laws	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
Laws of complements	$A \cup \bar{A} = \Omega$, $A \cap \bar{A} = \emptyset$, $A \cup \Omega = \Omega$, $A \cap \Omega = \mathrm{A}$, $A \cup \emptyset = \mathrm{A}$, $A \cap \emptyset = \emptyset$
Involution law	$\overline{(\bar{A})} = A$
Idempotence law	$A \cup A = A$ and $A \cap A = A$
Morgan's law	$\overline{(A \cup B)} = \bar{A} \cap \bar{B}$ and $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$

4 corollaries of axioms:

Corollary 1 : $$P(\varnothing) = 0$$

Corollary 2 : $$P(A) = 1 - P(\overline{A}) \leq 1$$

Corollary 3 : $$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Corollary 4 :

$$\text{If } A \subset B \text{ then } P(B) = P(A) + P(\bar{A} \cap B) \geq P(A)$$

$A \subset B$ means A is a subset of B

Chapter 2 : CONDITIONAL PROBABILITIES & INDEPENDENCE

– Definition and Interpretation
– Independence
– Law of total probability
– Bayes’ theorem

CONDITIONAL PROBABILITY

Definition of conditional probability :

Let *A* and *M* be two events and assume that *M* has occurred (a priori information) then $P(M) \neq 0$. The conditional probability of *A* given *M*, denoted $P(A|M)$ is:

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

Is P(A|M) a probability? Let's see if it verifies the 3 axioms:

- It is a positive or zero number.
- The conditional probability of the certain event P(Ω/M) is equal to 1.
- The conditional probability of the union of two disjoint events ($A \cap B \cap M = \emptyset$) is equal to the sum of their conditional probabilities, that is to say:

$$P(A \cup B|M) = P(A|M) + P(B|M)$$

Interpretation by relative frequency

- $n(A)$: number of times the event A occurs
- $n(M)$: number of times the event M occurs
- $n(A \cap M)$: number of times the events A and M occur simultaneously.
- n : total number of trials

$$P(A) = \frac{n(A)}{n};\ P(M) = \frac{n(M)}{n};\ P(A \cap M) = \frac{n(A \cap M)}{n}$$

According to the definition of conditional probability:

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

P(A) calculated with respect to *n(M)* instead of *n*, the total number of trials

Independent events

Two events A and B are independent if and only if:

$$\boxed{P(A \cap B) = P(A)P(B)}$$

Two events A and B are disjoint if and only if:

$$\boxed{P(A \cap B) = P(\emptyset) = 0}$$

Remark : if A and B are independent : P(A/B) = P(A)

Interpretation by relative frequency :

$$P(A) = \frac{n(A)}{n}; P(A|B) = \frac{n(A \cap B)}{n(B)}$$ which implies that …

if A and B are independent $$\frac{n(A)}{n} = \frac{n(A \cap B)}{n(B)}$$

If events A and B are independent, the relative frequency of event A is the same whether we consider the total number of trials or just those for which event B has occurred.

Example

Consider a deck of 52 cards (13 cards of four suits: clubs (♣), diamonds (♦), hearts (♥) and spades (♠). Each suit has 1 king, 1 Queen and 1 Jack card). Are the following two events A and B independent?

A = {we draw a queen} B = {we draw a heart}

if A and B are independent, the fact that B occurred does not modify at all the probability of A !

Example: SOLUTION!

Consider a deck of 52 cards. Are the following two events A and B independent?

A = {we draw a queen} B = {we draw a heart}

According to the 2nd definition of probability :

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

Besides A∩B = {we draw a queen of heart}, that is : $\frac{1}{52}$

if A and B are independent, the fact that B occurred does not modify at all the probability of A!

Law of total probability

Consider a set of N pairwise disjoint events denoted $M_1, \ldots, M_N$ whose union is the entire sample space $\boldsymbol{\Omega}$, that is:

$$\bigcup_{i=1}^{N} M_i = \Omega \quad \text{with } M_i \cap M_j = \emptyset \,, \; i \neq j$$

mutually exclusive and exhaustive

then for any event A:

$$\boxed{P(A) = \sum_{i=1}^{N} P(\, A \cap M_i) = \sum_{i=1}^{N} P(\, A|M_i)\, P(M_i)}$$

Example :

Experiment: Draw a coin from a set of 4500 coins where 800 are fake and 3700 fair. The coins were put into 4 boxes:

Box	Total	Fair	Fake
B1	2 000	1 600	400
B2	500	300	200
B3	1 000	900	100
B4	1 000	900	100

Question : What is the probability of drawing a fake coin?

$$P(F) = \sum_{i=1}^{4} P(F \mid \mathrm{B_i}) . P(B_i)$$

$$P(F) = \frac{1}{4}(0{,}2 + 0{,}4 + 0{,}1 + 0{,}1) = 0{,}2$$

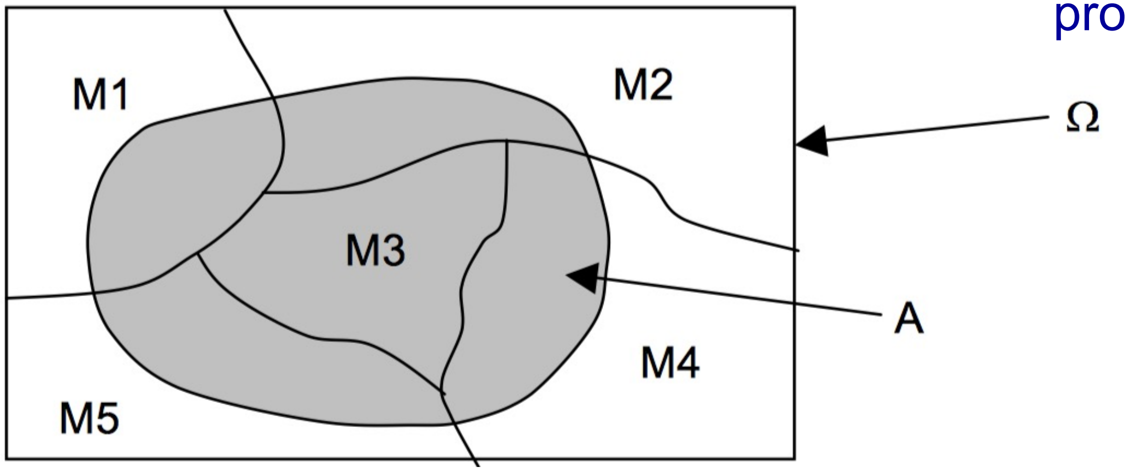
Bayes' theorem

Given a set of N pairwise <u>disjoint</u> events denoted $M_1, \ldots, M_N$ whose union is the entire sample space with prior probabilities $P(M_i)\ \forall\ i \in \{1, \ldots, N\}$, then given an event A for which $P(A) > 0$, the posterior probability of M_k given that A has occurred is:

$$\boxed{P(M_k|A) = \frac{P(A|M_k)P(M_k)}{\sum_{i=1}^{N} P(A|M_i)\,P(M_i)}}$$

Law of total probabilities

Bayes' theorem scheme



Example :

Experiment : Draw a coin from a set of 4500 coins where 800 are fake and 3700 fair. The coins were put into 4 boxes. We drew a fake coin.

Question : What is the probability that fake coin comes from box B_2?

$$P(B_2 \mid \mathrm{F}) = \frac{P(F \mid B_2)P(B_2)}{P(F)} = \frac{0{,}4 \,.\, 0{,}25}{0{,}2} = 0{,}5$$

Remark : *$P(B_2) = 0{,}25$ (prior probability)*

$P(B_2/F) = 0{,}5$ (posterior probability)

Appendix: Counting Methods

Type	Formulas	Explanation of Variables	Example
Permutation with repetition (Use permutation formulas *when order matters* in the problem.)	n^r	Where *n* is the number of things to choose from, and you choose *r* of them.	A lock has a 5 digit code. Each digit is chosen from 0-9, and a digit can be repeated. How many different codes can you have? n = 10, r = 5 $10^5 = 100{,}000$ codes
Permutation without repetition (Use permutation formulas *when order matters* in the problem.)	$\frac{n!}{(n-r)!}$	Where *n* is the number of things to choose from, and you choose *r* of them. Sometimes you can see the following notation for the same concept: $P(n,r) = {}^nP_r = {}_nP_r = \frac{n!}{(n-r)!}$	How many ways can you order 3 out of 16 different pool balls? n = 16, r = 3 $\frac{16!}{(16-3)!} = 3{,}360\ ways$
Combination with repetition (Use combination formulas *when order doesn't matter* in the problem.)	$\frac{(n+r-1)!}{r!\,(n-1)!}$	Where *n* is the number of things to choose from, and you choose *r* of them.	If there are 5 flavors of ice cream and you can have 3 scoops of ice cream, how many combinations can you have? You can repeat flavors. n = 5, r = 3 $\frac{(5+3-1)!}{3!\,(5-1)!} = 35\ combinations$
Combination without repetition (Use combination formulas *when order doesn't matter* in the problem.)	$\frac{n!}{r!\,(n-r)!}$	Where *n* is the number of things to choose from, and you choose *r* of them. Sometimes you can see the following notation for the same concept: $C(n,r) = {}^nC_r = {}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$	The state lottery chooses 6 different numbers between 1 and 50 to determine the winning numbers. How many combinations are possible? n = 50, r = 6 $\frac{50!}{6!\,(50-6)!} = 15{,}890{,}700\ combinations$