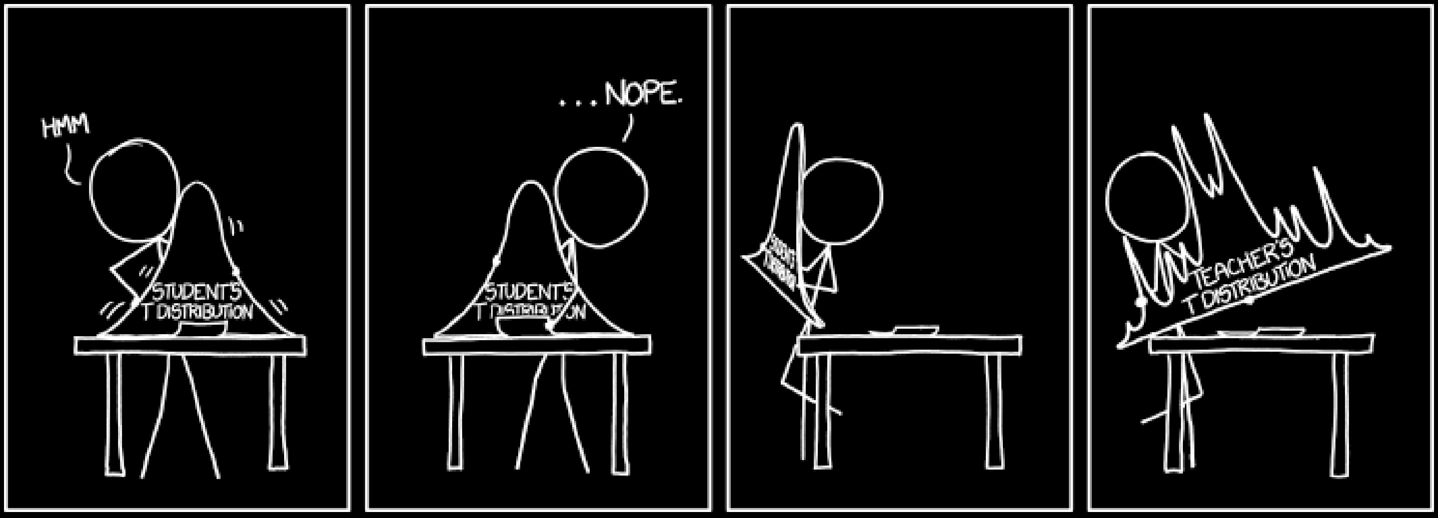
SQ28

Probability and Statistics

Alexis Flesch

Spring 2018

Student edition



http://www.xkcd.com/1347/

Contents

I Definitions

Definitions 1.1. The *sample space* of a random experiment is the set of all possible outcomes. An *event* A is a subset of the sample space. We will say that the event A occurs iff (if and only if) the outcome of the experiment is an element of A.

Example 1.2. Someone tosses a coin. The possible outcomes are "Heads" and "Tails". One can then denote by $\Omega = \{H; T\}$ the sample space of the experiment.

Definitions 1.3. Let Ω be a sample space and $\mathcal{A}$ be a set of subsets of Ω. A probability $\mathbb{P}$ on $(\Omega, \mathcal{A})$ is a mapping $P : \mathcal{A} \to [0, 1]$ satisfying the following axioms:

(i) $\mathbb{P}(A) \geqslant 0$ for all $A \subseteq \Omega$ (positivity).

(ii) $\mathbb{P}(\Omega) = 1$ (finitivity).

(iii) if $A \cap B = \varnothing$ (exclusive events), then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (additivity).

(iv) if for all $n \neq p$, $A_n \cap A_p = \varnothing$ then $\mathbb{P}(\cup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} P(A_i)$ (σ-additivity).

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

Example 1.4. Someone tosses a coin. A sample space for this experiment is $\Omega = \{H; T\}$. The corresponding set of events is:

$$\mathcal{A} = \{\{H\}; \{T\}; \{H; T\}; \varnothing\}.$$

If the coin isn't rigged, then the probability corresponding to this experiment is the map $\mathbb{P}\colon \mathcal{A} \to [0, 1]$ defined by:

$$\mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = .5, \quad \mathbb{P}(\varnothing) = 0 \quad \text{and} \quad \mathbb{P}(\{H; T\}) = 1.$$

Remark 1.5. When the set of outcomes Ω is finite or a subset of $\mathbb{Z}$, the corresponding set of events $\mathcal{A}$ will always be $\mathcal{P}(\Omega)$ for technical reasons.

Example 1.6. You roll a regular die. A sample space for this experiment is $\Omega = [\![1, 6]\!]$. The corresponding set of events is $\mathcal{A} = \mathcal{P}(\Omega)$ and the probability corresponding to this experiment is the map $\mathbb{P}\colon \mathcal{A} \to [0, 1]$ defined by:

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \ldots = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

1 *optional.* You toss two coins and are interested in the number of Heads. Give the probability map corresponding to those two choices of sample spaces:

$$\Omega_1 = \big\{\{H; H\}; \{H; T\}; \{T; T\}\big\} \quad \text{and} \quad \Omega_2 = \big\{(H, H); (H, T); (T, H); (T, T)\big\}.$$

Remark 1.7. When dealing with finite (or countable[1]) sample spaces, it is sufficient to define $\mathbb{P}$ on the sets $\{\omega\}$ for all $\omega \in \Omega$. For example, in the previous experiment, if one knows the values of $\mathbb{P}(\omega)$ for all ω then one can compute the probability that the outcome will be an odd number using the properties of $\mathbb{P}$:

$$\mathbb{P}(\{1; 3; 5\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{3\}) + \mathbb{P}(\{5\})$$

Properties 1.8. Let $\mathbb{P}$ be a probability map and let A and B be two events. Then:

[1] a set E is *countable* iff there is a one-to-one map between E and $\mathbb{N}$. For example, any infinite subspace of $\mathbb{Z}$ is countable, but $[0, 1]$ is not.

(i) $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$;

(ii) $\mathbb{P}(\varnothing) = 0$;

(iii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$;

(iv) if $A \subset B$ then $\mathbb{P}(A) \leqslant \mathbb{P}(B)$;

Proof. Exercise *(optional)*. □

Definition 1.9. A *partition* of Ω is a family $(B_i)_{i \in I}$ of subsets of Ω such that:

(i) $\cup_{i \in I} B_i = \Omega$;

(ii) $\forall i \neq j, B_i \cap B_j = \varnothing$;

Property 1.10. Let $(B_i)_{i \in I}$ be a partition of Ω. Then:

$$\sum_{i \in I} \mathbb{P}(B_i) = 1.$$

Proof. Straightforward. □

Theorem 1.11 (Law of total probabilities)

Let A be an event and $(B_i)_i$ a partition of Ω. Then:

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}\left(A \cap B_i\right).$$

II Conditional probability

You're playing Yatze (dice game) and got three aces on your first throw. Your strategy is now evolving: the knowledge of some events changes your sample space.

Definition 1.12. Let A and B be two events such that $\mathbb{P}(B) \neq 0$. The *conditional probability of A given B* is:

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

Remark 1.13. The map $\mathbb{P}_B$ is a probability map in the sense of definition 1.3. This is left as an exercise *(optional)*.

Definition 1.14. Two events A and B are said to be *independent* if:

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

Warning. Exclusive events $\neq$ independent events.

Remark 1.15. If $\mathbb{P}(B) \neq 0$, then A and B are independent iff $\mathbb{P}_B(A) = \mathbb{P}(A)$.

Proposition 1.16 (Bayes' Formula)

Let A and B be two events with non-zero probabilities. Then:

$$\mathbb{P}_B(A) = \frac{\mathbb{P}_A(B)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

Proof. Exercise *(mandatory)*. □

Remark 1.17. You don't have to remember Bayes' Formula by heart, but you must know that it exists and you must be able to find it.

Theorem 1.18 (Conditional low of total probabilities)

Let A be an event and $(B_i)_i$ a partition of Ω. Then:

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}_{B_i}(A)\mathbb{P}(B_i).$$

Proof. Exercise *(mandatory)*. □

Example 1.19. You invited two friends over, Mike and Anna. There is a 10% chance that Mike will show up if Anna doesn't and a 80% chance that Mike will come if Anna does. There is also a 50% chance that Anna might show up.

1. What is the probability that Mike will be there tonight ?
2. What is the probability that Anna will be there given that Mike will ?

Denote by A and B the events:

- A: Mike shows up;
- B: Anna shows up;

We know that:

$$\mathbb{P}_B(A) = .8, \quad \mathbb{P}_{\overline{B}}(A) = .1 \quad \text{and} \quad \mathbb{P}(B) = .5.$$

Thus:

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) \\
&= \mathbb{P}_B(A)\mathbb{P}(B) + \mathbb{P}_{\overline{B}}(A)\mathbb{P}(\overline{B}) \\
&= .8 \times .5 + .1 \times .5 \\
&= .45
\end{aligned}$$

Now, using Bayes' formula:

$$\mathbb{P}_A(B) = \frac{\mathbb{P}_B(A)\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{.8 \times .5}{.45} \approx 88.9\%.$$

Remark 1.20. The main difficulty in that kind of exercise is the modelisation of the problem. Always take your time to write everything down before doing anything, even if you have no idea how to solve the problem in the first place.

2 *optional.* Box I contains 3 red and 2 blue marbles while Box II contains 2 red and 8 blue marbles. A fair coin is tossed. If the coin turns up heads, a marble is chosen from Box I, otherwise a marble is chosen from Box II. Find the probability that a red marble is chosen:

- using the conditional law of total probabilities;
- using a tree diagram.

Find the probability that the marble came from Box I given that it was a blue one.

III Combinatorics

In some cases, like a non-weighted die roll or a fair coin toss, one can assume equal probabilities for all simple events, that is:

$$\forall \omega \in \Omega, \mathbb{P}(\{\omega\}) = \frac{1}{n},$$

where $n = \sharp\Omega$. Then, and only then, one can write:

$$\forall A \subset \Omega, \quad \mathbb{P}(A) = \frac{\sharp A}{\sharp \Omega}.$$

In this section, we will learn a few techniques to compute $\sharp A$.

III.1 Trees

In some situations, a tree diagram can be helpful.

Example 1.21. Imagin a man has two shirts, a green one and a blue one, and four ties, two of which are blue, one is green and the last one purple. If he picks randomly one shirt and one tie (with equiprobability), what is the probability that his shirt will match his tie? Draw a tree diagram.

III.2 Permutations

3 *mandatory.* Suppose you have three chairs of three different colors and that there are five people in the room. How many different ways of sitting them can you find ?

Definition 1.22. A *permutation* (or partial permutation or sequence without repetition) of lenght k from Ω is an ordered sequence of k distinct elements of Ω.

Example 1.23. Let $\Omega = \{a; b; c\}$ and $k = 2$. The 2-permutations of Ω are:

$$(a, b), (a, c), (b, c), (b, a), (c, a), \text{ and } (c, b).$$

Proposition 1.24

The number of k-permutations of n is:

$$n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

Idea of the proof. Draw a tree: we will do it in class. □

4 *mandatory.* How many different trifectas are there in a horse race with 10 horses?

III.3 Combinations

5 *optional.* Suppose you have three chairs and five guests and you wonder how many different ways of sitting your guests there are, except this time you just care about who is sitting down (and not who is on which chair).

Definition 1.25. A *combination* of k elements out of n is a subset $A \subset \Omega$ such that $\sharp A = k$. The number of combinations is denoted by $\binom{n}{k}$.

Example 1.26. Let $\Omega = \{a; b; c\}$ and $k = 2$. The different combinations of 2 elements of Ω are:

$$\{a, b\}, \{a, c\}, \text{ and } \{b, c\}.$$

Proposition 1.27

Let $(k, n) \in \mathbb{N}^2$ with $k \leqslant n$. Then:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Proof. Try and understand this result using the example 5. We will discuss it in class. □

6 *mandatory.* How many hands of 5 cards are there in a deck of 32 cards?

7 *optional.* How many hands of 5 cards with two aces are there in a deck of 32 cards?

IV Binomial coefficients

Proposition 1.28

Let $(k, n) \in \mathbb{N}^2$ such that $k \leqslant n$. Then:

$$\binom{n}{n-k} = \binom{n}{k}.$$

Proof. Exercise *(mandatory)*. □

Proposition 1.29 (Pascal's rule)

Let $(k, n) \in \mathbb{N}^2$ such that $k \leqslant n$. Then:

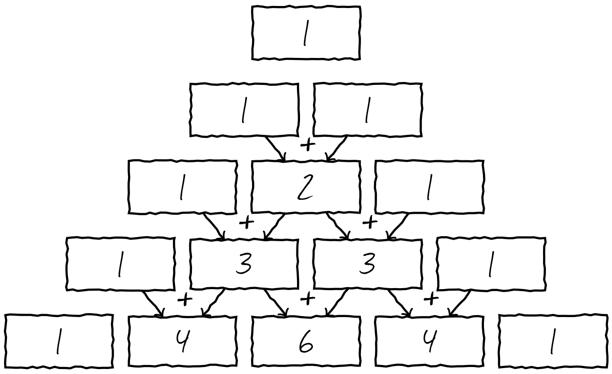
$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}.$$

Proof. It's a straight calculus:

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1} = \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-(k+1))!}$$
$$= \frac{n!}{(k+1)!(n-k)!}\left((k+1)+(n-k)\right)$$
$$= \frac{(n+1)!}{(k+1)!((n+1)-(k+1))!}.$$

□

Illustration 1.30. Pascal's triangle.



Theorem 1.31 (Binomial theorem)

Let $(a, b) \in \mathbb{C}^2$ and $n \in \mathbb{N}^\star$. Then:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Proof. See MTA. □

I Introduction

I.1 First definitions

Suppose you roll a fair die and win 10€ if it lands on a 6 and lose 1€ otherwise. You can define a map by:

$$X : \Omega \to \mathbb{R}, \quad \omega \mapsto \begin{cases} 10 & \text{if } \omega = 6, \\ -1 & \text{otherwise.} \end{cases}$$

This way, X associates a real number to any outcome of the experiment. Such a map is called a *random variable*. In this example, the probability that X takes the value 10 will be by definition:

$$\mathbb{P}_X(\{10\}) = \mathbb{P}(X = 10) = \mathbb{P}\left(\{\omega \in \Omega, X(\omega) = 10\}\right) = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

The probability for X to take the value -1 is:

$$\mathbb{P}_X(\{-1\}) = \mathbb{P}(X = -1) = \mathbb{P}\left(\{\omega \in \Omega, X(\omega) = -1\}\right) = \mathbb{P}(\{1; 2; 3; 4; 5\}) = \frac{5}{6}.$$

If you're only interested in the money you will win/lose then you don't need to "remember" the result of the die roll. The random variable X *transfers* the original probability space $(\Omega, \mathcal{A}, \mathbb{P})$ onto a new one $(X(\Omega), \mathcal{A}', \mathbb{P}_X)$.
From now on, you can forget about this construction. To describe a discrete random variable, we will enumerate the values it can take ($X(\Omega)$) and the probabilities with which it takes them.

Definition 2.1. A (real-valued) *random variable* (rv for short) is a map from Ω to $\mathbb{R}$.

Remark 2.2. We did not give a satisfying definition for $\mathcal{A}$[1] in the previous chapter. Therefore, we won't go into details either on how X should behave with respect to $\mathcal{A}$. For more information on that, you can check Wikipedia for σ-algebras (sigma-algebras) but this is way beyond the scope of this course.

Definition 2.3. A *discrete random variable* is a random variable that takes a finite or countable number of values. Typically, it is a $\mathbb{Z}$-valued function.

Example 2.4. The number of children in a family, the number of cars on road 66, the number of atoms in your body...

Definition 2.5. The *probability mass function* (or pmf for short) of a discrete random variable X is:

$$f_X : \mathbb{R} \to [0, 1], \quad x \mapsto \mathbb{P}(X = x),$$

where $\mathbb{P}(X = x) = 0$ if $x \notin X(\Omega)$.

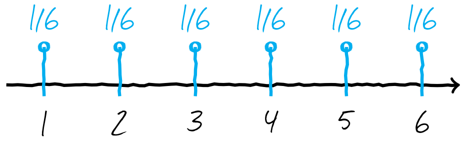
Example 2.6. Let's go back to the first example: you roll a fair die and win 10€ if it lands on a 6 and lose 1€ otherwise. Denote by X your profit. Then, the pmf of X is:

$$f_X : \mathbb{R} \to [0, 1], \quad x \mapsto \begin{cases} 5/6 & \text{if } x = -1, \\ 1/6 & \text{if } x = 10, \\ 0 & \text{otherwise.} \end{cases}$$

[1] $\mathcal{A} = \mathcal{P}(\Omega)$ when Ω is finite or countable and is "a set" of subsets of Ω otherwise.

Example 2.7. The pmf of a fair die is:

$$f : \mathbb{R} \to [0,1], \quad x \mapsto \begin{cases} 1/6 & \text{if } x \in [\![1,6]\!], \\ 0 & \text{otherwise.} \end{cases}$$



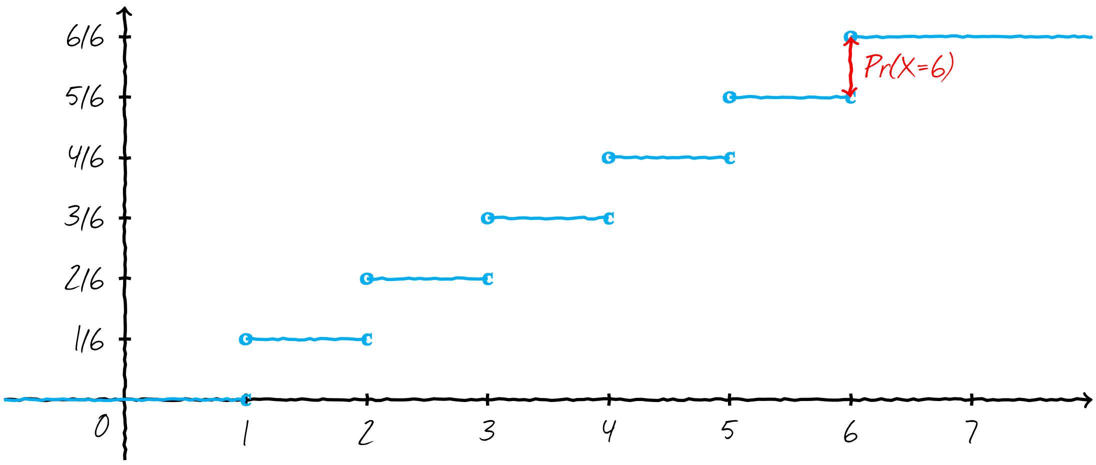
Definition 2.8. The *cumulative distribution function* (cdf) of a random variable X is:

$$F_X : \mathbb{R} \to [0,1], \quad x \mapsto \mathbb{P}(X \leqslant x),$$

Remark 2.9. The cdf F of a discrete random variable is a step function (or staircase function). It is non-decreasing, right-continuous with left limits and satisfies:

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to\infty} F(x) = 1.$$

Example 2.10. Denote by X the result of a fair die roll. The cdf of X is:



Remark 2.11. The cdf of a discrete random variable X describes entirely X. If $x_1, \ldots, x_n$ are the different possible values for X, then the jumps of F_X occur at the x_i's and the height of the jump at x_i is equal to $\mathbb{P}(X = x_i)$.

Remark 2.12. We will often use tabulars to describe discrete random variables. For example, in the case of the dice roll of example 2.10:

x_i	1	2	3	4	5	6
$\mathbb{P}(X = x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

I.2 Functions of random variables

Let X be a random variable. A new random variable Y can be defined by applying a function $\varphi\colon \mathbb{R} \to \mathbb{R}$ to X. This variable $Y = \varphi(X)$ satisfies:

$$Y(\Omega) = \{\varphi(x_1), \varphi(x_2), \ldots, \quad x_i \in X(\Omega)\},$$

and

$$\forall y \in Y(\Omega), \quad \mathbb{P}(Y = y) = \sum_{x, \varphi(x) = y} \mathbb{P}(X = x).$$

Example 2.13. Let X be a random variable satisfying:

x_i	-1	0	1	2
$\mathbb{P}(X = x_i)$	.2	.1	.3	.4

and let $Y = X^2$. Then $Y(\Omega) = \{0; 1; 4\}$ and:

$$\mathbb{P}(Y = 0) = \mathbb{P}(X^2 = 0) = \mathbb{P}(X = 0) = .1$$
$$\mathbb{P}(Y = 1) = \mathbb{P}(X^2 = 1) = \mathbb{P}(X = -1 \text{ or } X = 1) = .2 + .3 = .5$$
$$\mathbb{P}(Y = 4) = \mathbb{P}(X^2 = 4) = \mathbb{P}(X = 2) = .4$$

8 *mandatory.* Let X be the result of a fair dice roll. Suppose you win \$1 if X is odd and lose \$1 if X is even. Denote by Y your winnings. What is the probability distribution of Y?

II Expected value and variance

II.1 Finite case

In this section we consider a real-valued random variable X which is finite and we denote by $x_1, \dots, x_n$ the different possibles values for X:

$$X(\Omega) = \{x_i,\ i \in [\![1, n]\!]\}.$$

Example 2.14. Suppose that each time you roll a die, you win \$1 if you get an ace, \$2 if you get a 6 and lose 60 cents otherwise. Is it in your interest to play this game a lot?

Definition 2.15. Let X be a finite discrete rv with $X(\Omega) = \{x_1, \dots x_n\}$. The *expected value* (or expectation or mean) of X is:

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{P}(X = x_i) x_i.$$

Example 2.16. Let's go back to example 2.14 and denote by X your profit. Then

$$X(\Omega) = \{-.6, 1, 2\},$$

and

$$\mathbb{P}(X = -.6) = \frac{4}{6}, \quad \mathbb{P}(X = 1) = \frac{1}{6}, \quad \text{and} \quad \mathbb{P}(X = 2) = \frac{1}{6}.$$

The expected value of X is then:

$$\mathbb{E}(X) = \frac{1}{6} \times 1 + \frac{4}{6} \times (-.6) + \frac{1}{6} \times 2 = 0.1$$

This doesn't mean that you will win 10 cents most of the time. But if you play «a lot» you can expect to win 10 cents per game in average.

Proposition 2.17

Let X be a finite random variable with $X(\Omega) = \{x_1, \dots x_n\}$, $\varphi\colon \mathbb{R} \to \mathbb{R}$ and $Y = \varphi(X)$. Then:

$$\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{P}(X = x_i) \varphi(x_i).$$

Proof. This proof is a little technical and can be skipped. Denote by $x_1, \dots, x_n$ the different possible values for X and by $y_1, \dots, y_p$ the different possible values for Y. One can rearrange the x_i's into p groups such that:

$$\begin{aligned}
\forall i \in \{1, \dots, n_1\}, &\quad \varphi(x_i) = y_1, \\
\forall i \in \{n_1 + 1, \dots, n_2\}, &\quad \varphi(x_i) = y_2, \\
&\dots \\
\forall i \in \{n_{p-1} + 1, \dots, n_p\}, &\quad \varphi(x_i) = y_p.
\end{aligned}$$

Then:

$$\begin{aligned}
\sum_{i=1}^{n} \mathbb{P}(X = x_i)\varphi(x_i) &= \sum_{i=1}^{n_1} \mathbb{P}(X = x_i)\varphi(x_i) + \dots + \sum_{i=n_{p-1}+1}^{n_p} \mathbb{P}(X = x_i)\varphi(x_i) \\
&= \sum_{i=1}^{n_1} \mathbb{P}(X = x_i)y_1 + \dots + \sum_{i=n_{p-1}+1}^{n_p} \mathbb{P}(X = x_i)y_p \\
&= y_1 \sum_{i=1}^{n_1} \mathbb{P}(X = x_i) + \dots + y_p \sum_{i=n_{p-1}+1}^{n_p} \mathbb{P}(X = x_i) \\
&= y_1 \mathbb{P}(Y = y_1) + \dots + y_p \mathbb{P}(Y = y_p) \\
&= \mathbb{E}(Y).
\end{aligned}$$

□

Remark 2.18. As a consequence of this proposition, using the same notations, one has:

$$\mathbb{E}(X^2) = \sum_{i=1}^{n} \mathbb{P}(X = x_i)x_i^2.$$

Proposition 2.19

Let X and Y be two finite random variables and $(a, b, c) \in \mathbb{R}^3$. Then:

$$\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c.$$

Proof. Admitted. □

Example 2.20. You roll one fair die and denote by X_1 the result. The expected value of X_1 is:

$$\mathbb{E}(X_1) = \sum_{k=1}^{6} \frac{1}{6} k = \frac{1}{6} \sum_{k=1}^{6} k = \frac{1}{6} \cdot \frac{6 \times 7}{2} = 3.5$$

If you roll five dice, $X_1, \dots, X_5$, what is the expected value of the sum of the dice?

$$\mathbb{E}(X_1 + \dots + X_5) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_5) = 5 \times 3.5 = 17.5$$

Definition 2.21. Let X be a finite random variable. The *variance* of X is:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \geqslant 0.$$

Remark 2.22. The variance of X measures the spread of the distribution : when the variance is small, the elements of $X(\Omega)$ are close to the mean.

Proposition 2.23

Let X be a finite random variable, then:

$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2.$$

Proof. Exercise *(mandatory)*: use proposition 2.19. □

Proposition 2.24

Let X be a finite random variable and $(a, b) \in \mathbb{R}^2$. Then:

$$\mathbb{V}\text{ar}(aX + b) = a^2 \mathbb{V}\text{ar}(X).$$

Proof. Exercise *(mandatory)*: use propositions 2.23 and 2.19. □

Remark 2.25. As the variance measures the spread of the distribution, it is no surprise that the variance of X is the same as the one of $X + b$: the distribution is translated but its spread remains unchanged.

Definition 2.26. Let X be a finite random variable. The *standard deviation* of X is:

$$\sigma(X) = \sqrt{\mathbb{V}\text{ar}(X)}.$$

II.2 Countable case

In this section we consider a real-valued random variable X which is countable and we denote by $x_1, x_2, \dots$ the different possibles values for X:

$$X(\Omega) = \{x_n, n \in \mathbb{N}^\star\}.$$

Definition 2.27. An infinite series of real numbers $a_1, a_2, \dots$ is said to be *absolutely convergent* if the limit:

$$\lim_{N\to\infty} \sum_{n=1}^{N} |a_n|$$

exists and is finite.

Example 2.28. Let $a_n = 2^{-n}$ for all $n \in \mathbb{N}$. Then:

$$\sum_{n=0}^{N} a_n = \sum_{n=0}^{N} \left(\frac{1}{2}\right)^n = \frac{1 - \frac{1}{2^{N+1}}}{1 - \frac{1}{2}} = 2 - \frac{1}{2^N} \xrightarrow[N\to\infty]{} 2.$$

Property 2.29. If the infinite series $\sum a_n$ is absolutely convergent, then it is convergent in the sense that the following limit exists and is finite:

$$\lim_{N\to\infty} \sum_{n=1}^{N} a_n.$$

In this case, the limit will be denoted by $\sum_{n=1}^{+\infty} a_n$.

Definition 2.30. The *expected value* of X is:

$$\mathbb{E}(X) = \sum_{n=1}^{+\infty} \mathbb{P}(X = x_n)x_n,$$

provided that the sum is absolutely convergent.

Proposition 2.31

Let $\varphi\colon \mathbb{R} \to \mathbb{R}$ and $Y = \phi(X)$. If the infinite series

$$\sum_{n=1}^{+\infty} \varphi(x_n)\mathbb{P}(X = x_n)$$

is absolutely convergent, then the expected value of Y exists and is equal to the above series.

Proof. Admitted. □

Proposition 2.32

For any two random variables X and Y such that $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ exist and for any $(a, b, c) \in \mathbb{R}^3$:

$$\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c.$$

Proof. Admitted. □

Definition 2.33. Let X be a countably infinite random variable. The *variance* of X is:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) \geqslant 0,$$

provided that the above expected value exists.

Remark 2.34. As with finite random variables, the variance of X measures the spread of the distribution : when the variance is small, the elements of $X(\Omega)$ are close to the mean.

Proposition 2.35

Let X be a countable infinite random variable, then:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - \left(\mathbb{E}(X)\right)^2,$$

provided that $\mathbb{V}\mathrm{ar}(X)$ exists.

Proof. Straightforward (same proof as proposition 2.23). ☐

Proposition 2.36

Let X be a countably infinite random variable and $(a, b) \in \mathbb{R}^2$. Then:

$$\mathbb{V}\mathrm{ar}(aX + b) = a^2\mathbb{V}\mathrm{ar}(X),$$

provided that $\mathbb{V}\mathrm{ar}(X)$ exists.

Proof. Straightforward (same proof as proposition 2.24). ☐

Definition 2.37. Let X be a countably infinite random variable. The *standard deviation* of X is:

$$\sigma(X) = \sqrt{\mathbb{V}\mathrm{ar}(X)},$$

provided that $\mathbb{V}\mathrm{ar}(X)$ exists.

III Usual distributions

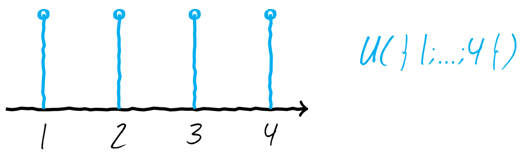
III.1 The Uniform distribution

Definition 2.38. Let X be a random variable with $X(\Omega) = \{x_1, \ldots, x_n\}$. The probability distribution of X is said to be *uniform* if:

$$\forall i \in [\![1, n]\!], \quad \mathbb{P}(X = x_i) = \frac{1}{n}.$$

In this case, we write $X \sim \mathcal{U}(\{x_1, \ldots, x_n\})$.

Illustration 2.39. Probability mass function of a uniform distribution on $\{1, \ldots, 4\}$.



Example 2.40. The distribution of a die roll is uniform on $[\![1, 6]\!]$.

Proposition 2.41

Let $X \sim \mathcal{U}(\llbracket 1, n \rrbracket)$. Then:

$$\mathbb{E}(X) = \frac{n+1}{2} \quad \text{and} \quad \mathbb{V}\text{ar}(X) = \frac{(n+1)(n-1)}{12}$$

Proof. • Let's first compute the mean:

$$\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{P}(X = k)\, k = \sum_{k=1}^{n} \frac{1}{n} k = \frac{1}{n} \sum_{k=1}^{n} k = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

• To compute the variance, we will use the formula:

$$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 .$$

Using proposition 2.23, we have:

$$\mathbb{E}(X^2) = \sum_{k=1}^{n} \frac{1}{n} k^2 = \frac{1}{n} \sum_{k=1}^{n} k^2 = \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$

Thus:

$$\mathbb{V}\text{ar}(X) = \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{(n+1)(n-1)}{12}$$

□

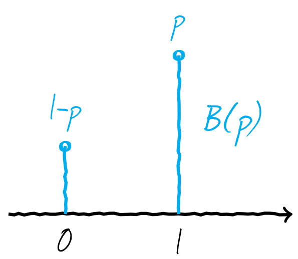
III.2 The Bernoulli distribution

Definition 2.42. A random variable X follows the *Bernoulli distribution* with parameter p if $X(\Omega) = \{0; 1\}$ and:

$$\mathbb{P}(X = 1) = 1 - \mathbb{P}(X = 0) = p.$$

We write $X \sim \mathcal{B}(p)$.

Illustration 2.43. Probability mass function of a Bernoulli variable.



Example 2.44. You toss a coin and denote by:

$$X = \begin{cases} 0 & \text{if it lands on heads,} \\ 1 & \text{otherwise.} \end{cases}$$

If the coin is fair, then $X \sim \mathcal{B}(.5)$ as:

$$\mathbb{P}(X = 1) = \mathbb{P}(X = 0) = .5$$

Proposition 2.45

Let $X \sim \mathcal{B}(p)$. Then:

$$\mathbb{E}(X) = p \qquad \text{and} \qquad \mathbb{V}\text{ar}(X) = p(1-p).$$

Proof. Exercice *mandatory*: take a look at the proof of proposition 2.41 if you don't know where to start. □

9 *mandatory.* You roll a die and denote by:

$$X = \begin{cases} 1 & \text{if you get a 6,} \\ 0 & \text{otherwise.} \end{cases}$$

1. What is the disitribution of the random variable X?
2. What is its expected value?

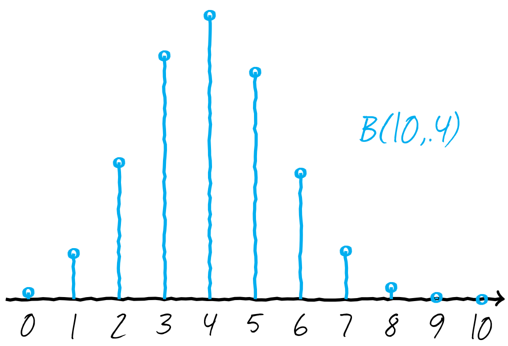
III.3 The binomial distribution

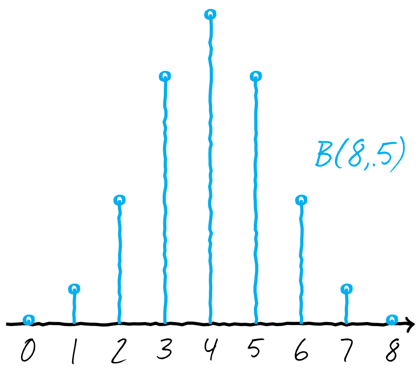
Definition 2.46. A random variable Y follows the *binomial distribution* with parameters n and p if it can be written as a sum of n independant Bernoulli random variables $X_i \sim \mathcal{B}(p)$:

$$Y = \sum_{k=1}^{n} X_i.$$

The binomial distribution counts the number of successes in a sequence of n independant Bernoulli trials.

Illustration 2.47. Probability mass functions of two binomial distributions.





Example 2.48. You toss a coin 10 times and denote by X the number of heads. If the coin is fair then $X \sim \mathcal{B}(10, .5)$.

10 *mandatory.* You roll a fair die 20 times and denote by X the number of aces. What is the distribution of X?

Remark 2.49. A Bernoulli variable with parameter p is a binomial variable with parameters $(1, p)$.

Proposition 2.50

Let $X \sim \mathcal{B}(n,p)$ and $k \in [\![0,n]\!]$. Then:

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Proof. Try and understand this result. We'll explain it in class. □

Proposition 2.51

Let $X \sim \mathcal{B}(n,p)$. Then:

$$\mathbb{E}(X) = np \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}(X) = np(1-p).$$

Proof. For the expected value: exercise *(mandatory)*. You can use definition 2.46 and the linearity of the mean. The result on the variance is admitted for now. □

11 *optional.* You roll 5 fair dice simultaneously and denote by X the number of aces.

a. What is the probability distribution of X?

b. What is the expected value of X?

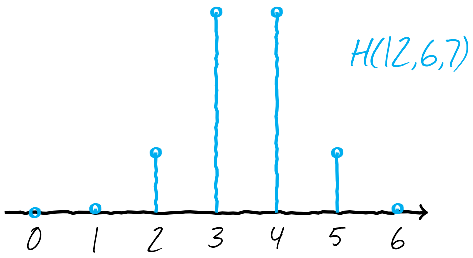
c. Compute $\mathbb{P}(X \leqslant 2)$.

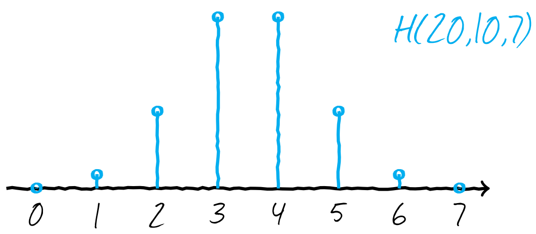
III.4 The hypergeometric distribution

12 *optional.* There are 10 balls in an urn, three of which are blue and seven of which are red. You draw simultaneously five balls. What is the probability that two of them are blue?

Definition 2.52. The hypergeometric distribution with parameters (N, n, m) counts the number of successes in n draws *without replacement* from a population of size N with m successes. When X is distributed as a hypergeometric random variable with parameters (N, n, m) we write $X \sim \mathcal{H}(N, n, m)$.

Illustration 2.53. Probability mass functions of two hypergeometric distributions.





Example 2.54. In an urn containing N balls, m of them being blue and the rest being red, if one draws simultaneously n balls and denote by X the number of blue balls, then $X \sim \mathcal{H}(N, n, m)$.

Proposition 2.55

Let $X \sim \mathcal{H}(N, n, m)$ and $k \in [\![0, n]\!]$. Then:

$$\mathbb{P}(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}.$$

Proof. Try and understand this result. We will discuss it in class. □

Proposition 2.56

Let $X \sim \mathcal{H}(N, n, m)$ and $k \in [\![0, n]\!]$. Then:

$$\mathbb{E}(X) = \frac{nm}{N} \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}(X) = \frac{n \cdot \frac{m}{N}\left(1 - \frac{n}{N}\right)(N - n)}{N - 1}.$$

Proof. Admitted. □

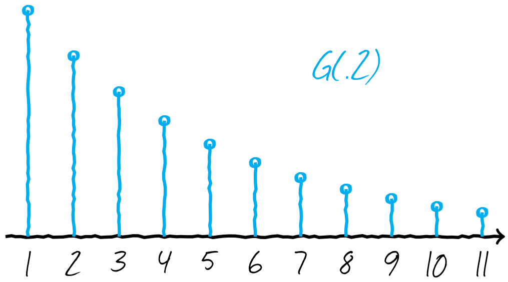
III.5 The geometric distribution

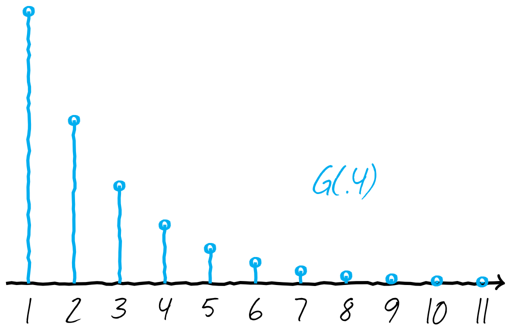
Definition 2.57. A random variable Y follows the geometric distribution with parameter p if it can be written:

$$Y = \min\{n \in \mathbb{N}^\star, X_n = 1\},$$

where $X_1, X_2, \ldots$ are independant Bernoulli random variables with parameter p. The geometric distribution counts the number of independant Bernoulli trials needed to get one success.

Illustration 2.58. Probability mass function of the geometric distribution with parameters .2 and .4





Example 2.59. You roll a die and count the number Y of rolls until it lands on a 6. Then, $Y \sim \mathcal{G}(1/6)$.

Proposition 2.60

Let $Y \sim \mathcal{G}(p)$ and let $n \in \mathbb{N}^\star$. Then:

$$\mathbb{P}(Y = n) = (1-p)^{n-1}p.$$

Proof. Straightforward. □

Proposition 2.61

Let $p \in (0,1)$ and $X \sim \mathcal{G}(p)$. Then:

$$\mathbb{E}(X) = \frac{1}{p} \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}(X) = \frac{1-p}{p^2}.$$

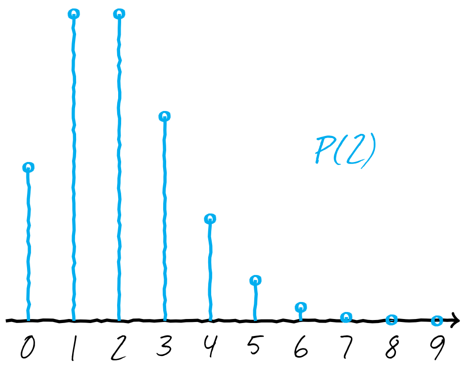
Proof. Admitted. □

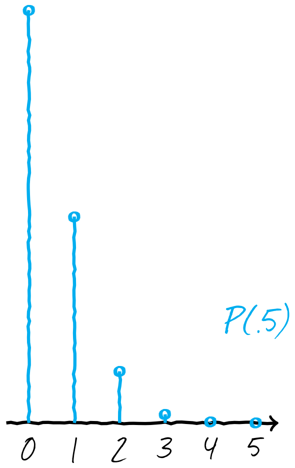
III.6 The Poisson distribution

Definition 2.62. A random variable X is Poisson disitributed with parameter $\lambda > 0$ if for all $k \in \mathbb{N}$:

$$\mathbb{P}(X = k) = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

Illustration 2.63. Probability mass function of the Poisson distribution with parameters 2 and .5





Remark 2.64. The Poisson distribution is sometimes used to count the number of occurences of a certain event in a given time interval. For example, the number of calls at a hotline between 8am and 9am.

Proposition 2.65

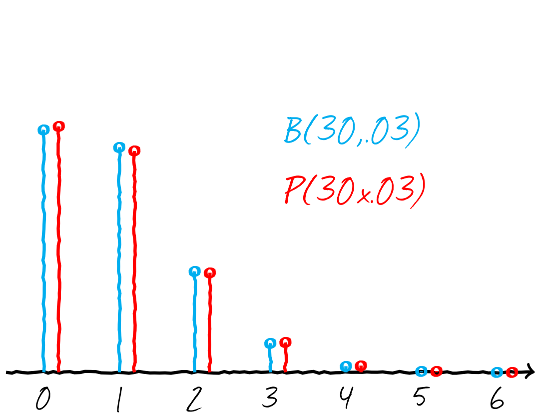
Let $X \sim \mathcal{P}(\lambda)$. Then:

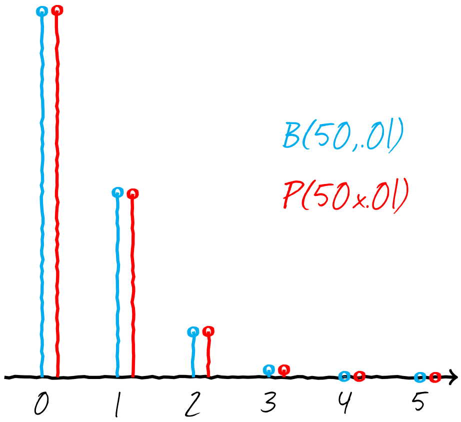
$$\mathbb{E}(X) = \lambda \qquad \text{and} \qquad \mathbb{V}\text{ar}(X) = \lambda.$$

Proof. We will do it in class. □

Remark 2.66. When $n \geqslant 30$, $p < .1$ and $np \leqslant 5$, the Poisson distribution $\mathcal{P}(np)$ is a good approximation of the binomial distribution $\mathcal{B}(n, p)$.

Illustration 2.67. Probability mass functions of the binomial distribution and the Poisson distribution. The red lines have voluntarily been shifted to make the graph more readable.





I Improper integrals

Definition 3.1. Let $f\colon \mathbb{R} \to \mathbb{R}_+$ be continuous except maybe for finitely many points and let $a \in \mathbb{R}$. The *improper integral* $\int_a^{+\infty} f(t)\mathrm{d}t$ is said to be *convergent* if the following limit exists and is finite:

$$\lim_{x\to+\infty} \int_a^x f(t)\mathrm{d}t.$$

In this case, we write:

$$\int_a^{+\infty} f(t)\mathrm{d}t = \lim_{x\to+\infty} \int_a^x f(t)\mathrm{d}t.$$

Example 3.2. Let:

$$\begin{array}{rccc} f & : & [1,+\infty[& \to & \mathbb{R}_+ \\ & & t & \mapsto & \dfrac{1}{t^2} \end{array}$$

Let $x \in \mathbb{R}_+$. Then:

$$\int_1^x f(t)\mathrm{d}t = \left[-\frac{1}{t}\right]_1^x = -\frac{1}{x} + 1$$

Thus:

$$\int_1^{+\infty} \frac{1}{t^2}\mathrm{d}t = 1.$$

13 *mandatory.* Show that the improper integral $\int_1^{+\infty} \frac{1}{t}\mathrm{d}t$ isn't convergent.

14 *mandatory.* Let $\lambda > 0$. Show that the improper integral $\int_0^{+\infty} \lambda \mathrm{e}^{-\lambda t}\mathrm{d}t$ is convergent.

Definition 3.3. Let $f\colon \mathbb{R} \to \mathbb{R}_+$ be continuous except maybe for finitely many points and let $a \in \mathbb{R}$. The improper integral $\int_{-\infty}^{a} f(t)\mathrm{d}t$ is said to be convergent if the following limit exists:

$$\lim_{x\to-\infty} \int_x^a f(t)\mathrm{d}t.$$

In this case, we write:

$$\int_{-\infty}^a f(t)\mathrm{d}t = \lim_{x\to-\infty} \int_x^a f(t)\mathrm{d}t.$$

Definition 3.4. Let $f\colon \mathbb{R} \to \mathbb{R}_+$ be continuous except maybe for finitely many points. The improper integral $\int_{-\infty}^{+\infty} f(t)\mathrm{d}t$ is convergent if the improper integrals $\int_a^{+\infty} f(t)\mathrm{d}t$ and $\int_{-\infty}^{a} f(t)\mathrm{d}t$ are convergent. In this case, we write:

$$\int_{-\infty}^{+\infty} f(t)\mathrm{d}t = \lim_{x\to-\infty}\lim_{y\to+\infty} \int_x^y f(t)\mathrm{d}t.$$

15 *optional.* Show that the following improper integral is convergent:

$$\int_{-\infty}^{+\infty} \frac{1}{1+t^2}\mathrm{d}t.$$

Definition 3.5. When f is not non-negative (i.e. when f can take negative values), the same definitions hold provided that $\int |f|$ is convergent.

II Probability density functions

Definition 3.6. A probability density function is a map $f\colon \mathbb{R} \to \mathbb{R}$ satisfying:

(i) $\forall t \in \mathbb{R}, f(t) \geqslant 0$;

(ii) f is continuous except maybe for finitely many points;

(iii) $\int_{-\infty}^{+\infty} f(t)\mathrm{d}t$ exists and is equal to 1.

Example 3.7. Let:

$$\begin{array}{rcl} f : \mathbb{R} & \to & \mathbb{R} \\ t & \mapsto & \begin{cases} \frac{1}{t^2} & \text{if } t > 1, \\ 0 & \text{otherwise} \end{cases} \end{array}$$

Then, f is a probability density function (see example 3.2).

Definition 3.8. Let E be a set and let $A \subset E$. The *indicator function* or *characteristic function* of A is:

$$\begin{array}{rcl} \mathbb{1}_A : A & \to & E \\ t & \mapsto & \begin{cases} 1 & \text{if } t \in A, \\ 0 & \text{otherwise.} \end{cases} \end{array}$$

Example 3.9. The function f from example 3.7 can be written:

$$\begin{array}{rcl} f : [1, +\infty[& \to & \mathbb{R}_+ \\ t & \mapsto & \frac{1}{t^2}\mathbb{1}_{]1,+\infty[}(t). \end{array}$$

16 *optional.* Find $C \in \mathbb{R}$ such that the following function is a probability density function:

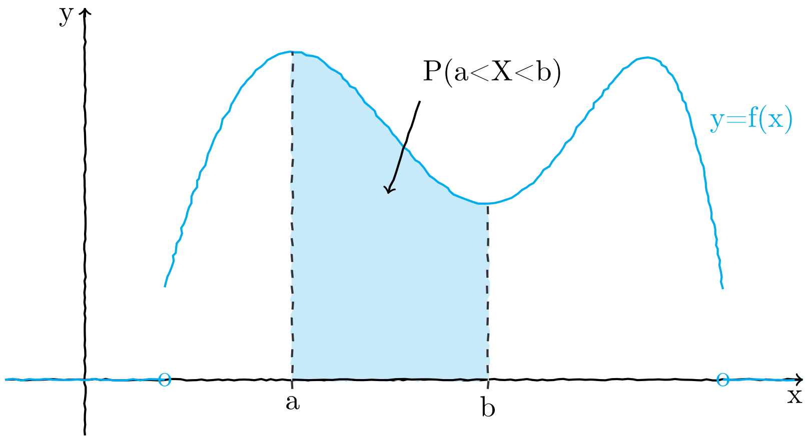
$$\begin{array}{rcl} f : \mathbb{R} & \to & \mathbb{R} \\ t & \mapsto & \frac{C}{1+t^2} \end{array}$$

Definition 3.10. Let f be a probability density function. A *continuous random variable* X with *probability density function* f is a random variable satisfying:

$$\forall (a,b) \in (\mathbb{R} \cup \{\pm\infty\})^2, \quad \mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f(t)\mathrm{d}t.$$

In this case, f is called **a** *density* of X.

Illustration 3.11. The area under f's graph between a and b represents the probability $\mathbb{P}(X \in [a,b])$.



17 *mandatory.* Let X be a continuous random variable with density f and let $a \in \mathbb{R}$. Show that:

$$\mathbb{P}(X = a) = 0.$$

Proposition 3.12

Let X be a continuous random variable and $a < b$ be two real numbers. Then:

(i) $\mathbb{P}(X = a) = 0$.

(ii) $\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(a < X < b)$;

Proof. The first point was dealt with in exercise 17. The second one is a consequence of:

$$\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(X = a) + \mathbb{P}(a < X < b) + \mathbb{P}(X = b).$$

□

18 *optional.* Let:

$$\begin{array}{rcccl} f & : & \mathbb{R} & \to & \mathbb{R} \\ & & t & \mapsto & \cos(t)\,\mathbb{1}_{[0,\pi/2]}(t). \end{array}$$

1. Plot f.
2. Show that f is a probability density function.
3. Compute $\mathbb{P}(X \leqslant \pi/4)$ and illustrate the result (draw the graph of f).

III Cumulative distribution functions

Definition 3.13. The *cumulative distribution function* (cdf) of a continuous random variable X is:

$$\begin{array}{rcccl} F_X & : & \mathbb{R} & \to & [0,1] \\ & & x & \mapsto & \mathbb{P}(X \leqslant x), \end{array}$$

Remark 3.14. This is the same definition as the one for discrete random variables.

Remark 3.15. Let X be a continuous random variable with density f and let $x \in \mathbb{R}$. Then:

$$F(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f(t)\mathrm{d}t.$$

19 *optional.* Let:

$$\begin{array}{rccc} f & : & \mathbb{R} & \to & \mathbb{R} \\ & & t & \mapsto & \mathbb{1}_{[0,1]}(t) \end{array}$$

Show that f is a probability density function and compute the corresponding cumulative distribution function.

Proposition 3.16

Let F be the cumulative distribution function of a continuous random variable X. Then:

$$\mathbb{P}(a < X \leqslant b) = F(b) - F(a)$$

Proof. Exercise *(mandatory)*. □

Proposition 3.17

Let F be the cumulative distribution function of a continuous random variable with density f. Then:

(i) F is non-decreasing;

(ii) $\lim_{-\infty} F = 0$ and $\lim_{+\infty} F = 1$;

(iii) F is continuous;

(iv) F is differentiable except maybe for finitely many points;

(v) If F is differentiable at $t \in \mathbb{R}$ then:

$$F'(t) = f(t).$$

Proof. Point (i) is left as an exercise *(mandatory)*, so are points (ii), (iii), (iv) and (v) *(optional)*. *Hint for (ii):* $\lim_{-\infty} F = 1 - \lim_{+\infty} F$. □

Theorem 3.18

Let X be a random variable with cumulative distribution function F. Suppose that:

(i) F is continuous;

(ii) F is differentiable except maybe for finitely many points;

Then X is a continuous random variable. Moreover, any non-negative map f satisfying $f(x) = F'(x)$ at any point x where F is differentiable is a density of X.

Proof. Admitted. □

Method. To prove that a random variable X is continuous, one can compute its cumulative distribution function F and check that it is continuous everywhere and differentiable except maybe for finitely many points. Then, one can find a density of X by differentiating F.

Example 3.19. Let:

$$f : \mathbb{R} \to \mathbb{R}, \quad t \mapsto \begin{cases} 0 & \text{if } t < 0, \\ \mathrm{e}^{-t} & \text{otherwise.} \end{cases}$$

Then, according to exercise 16, f is a probability density function. Suppose that X has density f. Then, its cumulative distribution function at $t > 0$ is given by:

$$F_X(t) = \mathbb{P}(X \leqslant t) = \int_{-\infty}^{t} f(x)\mathrm{d}x = \int_{0}^{t} \mathrm{e}^{-x}\mathrm{d}x = 1 - \mathrm{e}^{-t}.$$

Let now $Y = X^3$. Then, for any $t \in \mathbb{R}^+$:

$$F_Y(t) = \mathbb{P}(Y \leqslant t) = \mathbb{P}(X^3 \leqslant t) = \mathbb{P}(X \leqslant \sqrt[3]{t}) = F_X(\sqrt[3]{t}) = 1 - \mathrm{e}^{\sqrt[3]{t}}.$$

Thus:

$$F_Y(t) = \left(1 - \mathrm{e}^{\sqrt[3]{t}}\right) \mathbb{1}_{\mathbb{R}_+}(t) = \begin{cases} 0 & \text{if } t < 0, \\ 1 - \mathrm{e}^{\sqrt[3]{t}} & \text{otherwise.} \end{cases}$$

As F_Y is continuous everywhere and differentiable (except maybe at 0), Y is a continuous random variable. A probability density function for Y is given by:

$$f_Y(t) = -\frac{1}{3} t^{-2/3} \mathrm{e}^{\sqrt[3]{t}} \mathbb{1}_{\mathbb{R}_+}(t) = \begin{cases} 0 & \text{if } t < 0, \\ -\frac{1}{3} t^{-2/3} \mathrm{e}^{\sqrt[3]{t}} & \text{otherwise.} \end{cases}$$

IV Expected value and variance

IV.1 Expected value

Definition 3.20. Let X be a continuous random variable with density f. If the improper integral $\int_{-\infty}^{+\infty} |t| f(t)\mathrm{d}t$ is convergent, then we say that X has an *expected value* (or *mean* or *expectation*), which is given by:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t f(t)\mathrm{d}t.$$

Example 3.21. Let X be a continuous random variable with density:

$$f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto \begin{cases} 0 & \text{if } x < 0, \\ \dfrac{2}{\pi(1+t^2)} & \text{otherwise.} \end{cases}$$

One can easily show that f is a probability density function. Let's check whether X has an expected value:

$$\int_{-\infty}^{+\infty} |t| f(t)\mathrm{d}t = \int_{0}^{+\infty} \frac{2t}{\pi(1+t^2)}.$$

X has an expected value iff the above improper integral is convergent. But:

$$\int_0^x \frac{2t}{\pi(1+t^2)} = \frac{1}{\pi}\left[\log(|1+t^2|)\right]_0^x = \frac{1}{\pi}\log(1+x^2) \underset{x\to\infty}{\longrightarrow} +\infty.$$

Thus, X doesn't have an expected value.

Example 3.22. Let X be a continuous random variable with density:

$$f : \mathbb{R} \to \mathbb{R}, \quad t \mapsto \begin{cases} 1 & \text{if } 0 < t < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Then, X admits an expected value as:

$$\int_{-\infty}^{+\infty} |t| f(t) \mathrm{d}t = \int_0^1 |t| \mathrm{d}t < \infty.$$

Thus:

$$\mathbb{E}(X) = \int_0^1 t \, \mathrm{d}t = \frac{1}{2}.$$

Remark 3.23. Let X be a continuous random variable with density f. If f is equal to zero in the neighborhood of $\pm\infty$, then X has an expected value as the corresponding integral is not improper.

Proposition 3.24 (linearity of the mean)

Let X and Y be two continuous random variables admitting expected values and let $(a, b) \in \mathbb{R}^2$. Then:

$$\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c$$

Proof. Admitted. □

Remark 3.25. It is not surprising that $\mathbb{E}(c) = c$ as the value one can expect from a constant is the constant itself.

IV.2 Variance

Proposition 3.26

If $\mathbb{E}(X^2) < +\infty$, then $\mathbb{E}(X)$ exists.

Proof. Admitted. □

Proposition 3.27 (law of the unconscious statistician[1])

Let X be a continuous random variable with probability density function f and let $\varphi \colon \mathbb{R} \to \mathbb{R}$. If $\mathbb{E}(\varphi(X))$ exists then:

$$\mathbb{E}(\varphi(X)) = \int \varphi(t) f(t) \mathrm{d}t.$$

Proof. Admitted. □

[1] According to Sheldon M. Ross' book "Introduction to Probability Models", "this law got its name from "unconscious" statisticians who have used it as if it were the definition of $\mathbb{E}(\varphi(X))$". In French, it is known as the "transfer theorem".

Definition 3.28. Let X be a continuous random variable with density f. If the improper integral $\int_{-\infty}^{+\infty} t^2 f(t)\mathrm{d}t$ is convergent, then it is called the *second moment* of X and it is denoted by $\mathbb{E}(X^2)$.

Definition 3.29. Let X be a continuous random variable. The *variance* of X when it exists is:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

The variance of a (continuous) random variable measures the spread of the variable. It is always non-negative.

Remark 3.30. If X has a second order moment, then $\mathbb{V}\mathrm{ar}(X)$ exists.

Proposition 3.31

If $\mathbb{V}\mathrm{ar}(X)$ exists, then it can be expressed as:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Proof. It's the same proof as the one for the discrete case. □

Example 3.32. Let X be a continuous random variable with density:

$$\begin{aligned} f \ : \ \mathbb{R} &\to \mathbb{R} \\ t &\mapsto \begin{cases} 1 & \text{if } 0 < t < 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

We already computed the expected value of X in example 3.22. Using the law of the unconscious statistician,

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} t^2 f(t)\mathrm{d}t = \int_0^1 t^2 \mathrm{d}t = \frac{1}{3},$$

which yields:

$$\mathbb{V}\mathrm{ar}(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}.$$

Proposition 3.33

Let X be a random variable and $(a, b) \in \mathbb{R}^2$. If the variance of X exists, then:

$$\mathbb{V}\mathrm{ar}(aX + b) = a^2 \mathbb{V}\mathrm{ar}(X).$$

Proof. Again, it's the same proof as the one for the discrete case. □

Definition 3.34. The *standard deviation* of a continuous random variable X, when it exists, is $\sigma(X) = \sqrt{\mathbb{V}\mathrm{ar}(X)}$.

20 *mandatory.* Let X be a random variable with expected value μ and standard deviation σ. Compute the expected value and the variance of:

$$Y = \frac{X - \mu}{\sigma}.$$

V Usual distributions

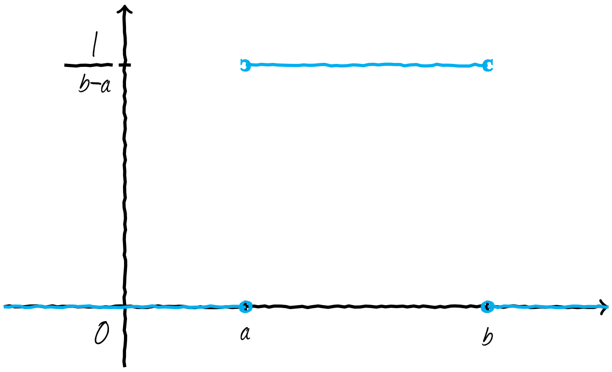
V.1 The uniform distribution

Definition 3.35. Let $a < b$ be two real numbers. A random variable X is *uniformly distributed* on $[a, b]$ iff the following map is a density of X:

$$f : \mathbb{R} \to \mathbb{R}, \quad t \mapsto \begin{cases} \dfrac{1}{b-a} & \text{if } a < t < b \\ 0 & \text{otherwise.} \end{cases}$$

We write $X \sim \mathcal{U}([a, b])$.

Illustration 3.36.



Example 3.37. Suppose that $X \sim \mathcal{U}([0, 1])$. Then:

$$\mathbb{P}(X \leqslant .5) = \int_{-\infty}^{.5} \mathbb{1}_{[0,1]}(t)\,\mathrm{d}t = \int_{0}^{.5} 1\mathrm{d}t = .5$$

Remark 3.38. Every programming language has a `random` function that outputs a *realization* of the uniform distribution on $[0, 1]$. One can show that any probability distribution can be expressed as a function of the uniform distribution on $[0, 1]$.

21 *mandatory.* Let $X \sim \mathcal{U}([5, 7])$. Compute $\mathbb{P}(X < 0)$ and $\mathbb{P}(X \in [5; 5.5])$.

22 *optional.* Let $U \sim \mathcal{U}([0, 1])$ and let $p \in [0, 1]$. Let also:

$$Z = \mathbb{1}_{[0,p]}(U) = \begin{cases} 1 & \text{if } U \leqslant p, \\ 0 & \text{otherwise.} \end{cases}$$

Show that $Z \sim \mathcal{B}(p)$.

23 *optional.* Let $X \sim \mathcal{U}([0, 1])$ and $a < b$ be two real numbers. Compute the cumulative distribution function of $Y = (b - a)X + a$. What is the distribution of Y?
Hint: take a look at the proof of proposition 3.48.

Proposition 3.39

Let $a < b$ be two real numbers and let $X \sim \mathcal{U}([a, b])$. Then:

$$\mathbb{E}(X) = \frac{a+b}{2} \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}(X) = \frac{(b-a)^2}{12}.$$

Proof. Denote by f the probability density of X. Then:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a}\cdot\frac{b^2-a^2}{2} = \frac{a+b}{2}.$$

Moreover:

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^2+ab+a^2}{3}.$$

Finally:

$$\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{(b-a)^2}{12}.$$

□

24 *optional.* Use examples 3.22 and 3.32 and the properties on the mean and the variance to give a simpler proof.

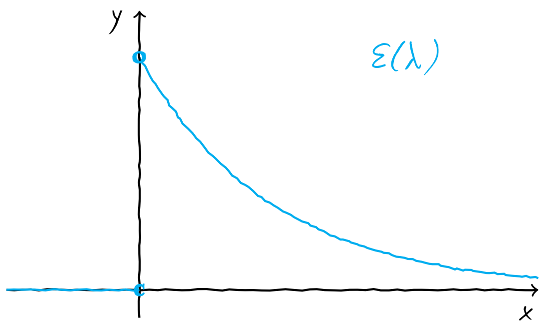
V.2 The exponential distribution

Definition 3.40. Let $\lambda > 0$. A random variable X is *exponentially distributed* with parameter λ if the following map is a density of X:

$$f : \mathbb{R} \to \mathbb{R}, \quad t \mapsto \begin{cases} \lambda e^{-\lambda t} & \text{if } t > 0, \\ 0 & \text{otherwise.} \end{cases}$$

We write $X \sim \mathcal{E}(\lambda)$.

Illustration 3.41.



Remark 3.42. In some books, you might find a slightly different definition for the exponential distribution where $1/\lambda$ is used instead of λ.

Proposition 3.43

Let $\lambda > 0$ and let $X \sim \mathcal{E}(\lambda)$. Then:

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad \text{and} \quad \mathbb{V}\mathrm{ar}(X) = \frac{1}{\lambda^2}.$$

Proof. Exercise *(optional)*. □

Remark 3.44. One can show that the exponential distribution is "memory-less", that is:

$$\forall s, t > 0, \quad \mathbb{P}_{(X \geqslant t)}(X \geqslant t+s) = \mathbb{P}(X \geqslant s).$$

V.3 The normal distribution

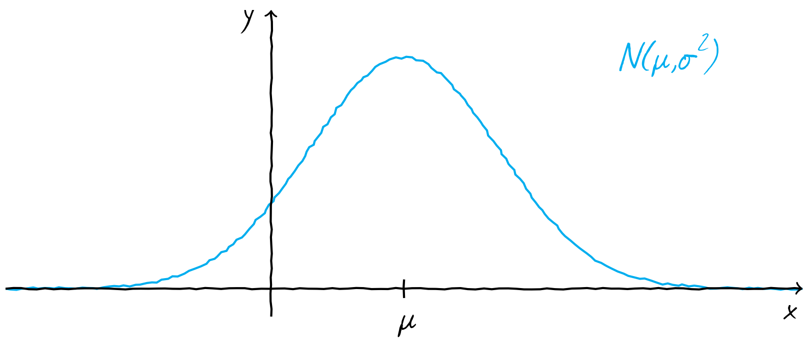
Definition 3.45. Let $(\mu,\sigma) \in \mathbb{R}\times\mathbb{R}_+^\star$. A random variable X has a *normal probability distribution* with parameters (μ,σ^2) if the following map is a density of X:

$$f : \begin{array}{ccc} \mathbb{R} & \to & \mathbb{R} \\ t & \mapsto & \dfrac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\dfrac{(t-\mu)^2}{2\sigma^2}} \end{array}$$

We write $X \sim \mathcal{N}(\mu,\sigma^2)$.

Definition 3.46. When $Z \sim \mathcal{N}(0,1)$, we say that Z has a *standard* normal distribution.

Illustration 3.47.



Proposition 3.48

Let $X \sim \mathcal{N}(0,1)$ and $(\mu,\sigma) \in \mathbb{R}\times\mathbb{R}_+\star$. Then:

$$\sigma X + \mu \sim \mathcal{N}(\mu,\sigma^2).$$

Proof. Denote by F_X the cumulative distribution function of X and by F_Y the cumulative distribution function of $Y = \sigma X + \mu$. Let $x \in \mathbb{R}$. Then:

$$F_Y(x) = \mathbb{P}(Y \leqslant x) = \mathbb{P}(\sigma X + \mu \leqslant x) = \mathbb{P}\left(X \leqslant \frac{1}{\sigma}(x-\mu)\right) = F_X\left(\frac{1}{\sigma}(x-\mu)\right).$$

By definition:

$$F_X\left(\frac{1}{\sigma}(x-\mu)\right) = \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}\mathrm{d}x.$$

A simple integration by substitution ($u = (x-\mu)/\sigma$) yields the result. □

Proposition 3.49

Let $(\mu,\sigma) \in \mathbb{R}\times\mathbb{R}_+^\star$ and let $X \sim \mathcal{N}(\mu,\sigma^2)$. Then:

$$\mathbb{E}(X) = \mu \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}(X) = \sigma^2.$$

Idea of the proof. According to proposition 3.48, one can write $X = \sigma Z + \mu$ where $Z \sim \mathcal{N}(0,1)$. Thus:

$$\mathbb{E}(X) = \sigma\mathbb{E}(Z) + \mu.$$

But, the probability density function f_Z of Z is even:

$$f_Z(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2}.$$

Thus $t \mapsto t f_Z(t)$ is odd and:

$$\mathbb{E}(Z) = \int_{-\infty}^{+\infty} t f_Z(t) = 0.$$

Computing the variance is a little bit trickier: the easiest way is probably to calculate $\mathbb{E}(Z^2)$ by using an integration by parts. □

Remark 3.50. One can show that the antiderivatives of $t \mapsto \mathrm{e}^{-t^2}$ cannot be expressed with "usual functions" such as sin, exp, etc... Thus, to compute $\mathbb{P}(Z \in [a,b])$ where $Z \sim \mathcal{N}(\mu, \sigma^2)$ for some real numbers (a,b), one will use the *standard normal table*, which is a table with the values of the cumulative distribution function of a standard normal distribution.

Example 3.51. Suppose that the size X of a man in centimeters satisfies $X \sim \mathcal{N}(175, 49)$. What is the probability that X is greater than 180? Write:

$$Z = \frac{X - 175}{7} \sim \mathcal{N}(0,1).$$

Then:

$$\mathbb{P}(X \geqslant 180) = \mathbb{P}(7Z + 175 \geqslant 180) = \mathbb{P}(Z \geqslant 5/7) = 1 - \mathbb{P}(Z \leqslant 5/7).$$

Now, looking at the standard normal table:

$$\mathbb{P}(\mathcal{N}(0,1) \leqslant 5/7) \simeq 76.11\%.$$

Thus:

$$\mathbb{P}(X \geqslant 180) \simeq 23.89\%.$$

I Discrete random variables

I.1 Joint distribution of two random variables

Definition 4.1. Let X and Y be two discrete random variables. The *joint probability distribution* of X and Y is the distribution of $Z = (X, Y)$.

Definition 4.2. Let X and Y be two discrete random variables. The *joint probability mass function* of X and Y is:

$$p_{XY} \quad : \quad \begin{array}{ccc} X(\Omega) \times Y(\Omega) & \to & [0,1] \\ (x,y) & \mapsto & \mathbb{P}(X=x, Y=y) \end{array}$$

Example 4.3. You roll two dice, a green one and a blue one. Denote by X and Y the results of the green and the blue one respectively. The joint probability mass function of X and Y is:

$$p_{XY} \quad : \quad \begin{array}{ccc} [\![1,6]\!] \times [\![1,6]\!] & \to & [0,1] \\ (x,y) & \mapsto & \frac{1}{36} \end{array}$$

Example 4.4. Let X and Y be two random variables with joint probability distribution given by:

$$p_{XY} \quad : \quad \begin{array}{ccc} [\![0,2]\!] \times [\![0,2]\!] & \to & [0,1] \\ (x,y) & \mapsto & \frac{x+y}{18} \end{array}$$

Let's write the corresponding *joint probability table*:

X \ Y	$y_1 = 0$	$y_2 = 1$	$y_3 = 2$	$\mathbb{P}(X = x_i)$
$x_1 = 0$	0	$1/18$	$2/18$	$3/18$
$x_2 = 1$	$1/18$	$2/18$	$3/18$	$6/18$
$x_3 = 2$	$2/18$	$3/18$	$4/18$	$9/18$
$\mathbb{P}(Y = y_j)$	$3/18$	$6/18$	$9/18$	

This table contains all the information on the joint probability distribution of X and Y and more. The *marginal* distribution of X, which is simply the probability distribution of X (see next section) can be read in the last column.
The law of total probabilities is hidden in the last row and the last column. Here is an example why:

$$\mathbb{P}(Y=0) = \sum_{i=1}^{3} \mathbb{P}(Y=0, X=x_i) = 0 + \frac{1}{18} + \frac{2}{18} = \frac{3}{18}$$

I.2 Marginal distribution

Definition 4.5. Given two random variables X and Y, the *marginal distribution* of X is the probability distribution of X.

Proposition 4.6

Let X and Y be two random variables. The marginal distribution of X is given by:

$$\mathbb{P}(X=x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X=x, Y=y).$$

Proof. Exercise *(optional)*. Use the law of total probabilities with the partition $(Y = y)_{y \in Y(\Omega)}$. □

I.3 Conditional distribution

Definition 4.7. The *conditional distribution* of X given $Y = y$ is the probability distribution of X given the event $Y = y$. More specifically, it is the:

$$\mathbb{P}(X = x | Y = y), \quad \text{for all } x \in X(\Omega).$$

Example 4.8. You roll a fair die and denote by X the result. You then flip a coin X times and denote by Y the number of heads. The conditional distribution of Y given $(X = n)$ is the binomial distribution with parameters $(n, 1/2)$. This can be written as:

$$(Y | X = n) \sim \mathcal{B}(n, p).$$

25 *mandatory.* Let X and Y be two random variables with probability distribution:

X \ Y	1	2	3	4
1	0.08	0.04	0.16	0.12
2	0.04	0.02	0.08	0.06
3	0.08	0.04	0.16	0.12

1. Compute the marginal distributions of X and Y
2. Compute the conditional distribution of X given $Y = 2$.
3. Compute $\mathbb{E}(X | Y = 2)$.

II Independence

Definition 4.9. Two random variables (discrete or continuous) X and Y are said to be *independent* if for all intervals A and B:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(X \in B).$$

Property 4.10. Two random variables X and Y are independent iff:

$$\forall (x, y) \in \mathbb{R}^2, \ \mathbb{P}(X \leqslant x, Y \leqslant y) = \mathbb{P}(X \leqslant x)\mathbb{P}(Y \leqslant y).$$

Proof. Admitted. □

Proposition 4.11

Let X and Y be two discrete random variables. Then, X and Y are independent iff:

$$\forall (x, y) \in X(\Omega) \times Y(\Omega), \ \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y).$$

Proof. It's a bit technical: admitted. □

Remark 4.12. This proposition doesn't hold for continuous random variables as if X is continuous then $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$.

Proposition 4.13

Let X and Y be two independent random variables and let $\varphi, \psi\colon \mathbb{R} \to \mathbb{R}$. Then, $\varphi(X)$ and $\psi(Y)$ are independent.

Proof. Admitted. □

Definition 4.14. Let $X_1, \dots, X_n$ be n random variables. We say that they are *mutually independent* if for all intervals $A_1, \dots, A_n$:

$$\mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) = \mathbb{P}(X_1 \in A_1) \times \dots \times \mathbb{P}(X_n \in A_n).$$

Property 4.15. Let $X_1, \dots, X_n$ be n random variables. They are mutually independent iff:

$$\forall (t_1, \dots, t_n) \in \mathbb{R}^n,\ \mathbb{P}(X_1 \leqslant t_1, \dots, X_n \leqslant t_n) = \mathbb{P}(X_1 \leqslant t_1) \times \dots \times \mathbb{P}(X_n \leqslant t_n).$$

Remark 4.16. Proposition 4.11 is easily adapted to this case.

III Sum of random variables

III.1 Variance

Proposition 4.17

Let X and Y be two **independent** random variables. Then:

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y),$$

provided that $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ exist.

Proof. Admitted. □

Remark 4.18. The converse is false in general.

Remark 4.19. If the random variables are not independent then the result is false in general.

26 *mandatory.* Let $X \sim \mathcal{B}(p)$ where $p \in (0, 1)$. and let $Y = X$. Compute $\mathbb{E}(XY)$ and $\mathbb{E}(X)\mathbb{E}(Y)$. Are X and Y independent?

27 *optional.* Let $X \sim \mathcal{N}(0, 1)$ and let $Y = X$. Show that $\mathbb{E}(XY) \neq \mathbb{E}(X)\mathbb{E}(Y)$ by using the properties of the normal distribution.

Definition 4.20. The *covariance* of two random variables X and Y is:

$$\mathbb{C}\mathrm{ov}(X, Y) = \mathbb{E}\big((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big),$$

provided that $\mathbb{E}(X)$, $\mathbb{E}(Y)$ and $\mathbb{E}(XY)$ exist.

Proposition 4.21

Let X and Y be two random variables admitting a covariance. Then:

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Proof. exercice (*mandatory*). □

Warning. If X and Y are independent, then $\mathbb{C}\text{ov}(X, Y) = 0$ (see proposition 4.17), but the converse is false.

Remark 4.22. Let X and Y be two random variables admitting a covariance. Then:

$$\mathbb{C}\text{ov}(X, Y) = \mathbb{C}\text{ov}(Y, X).$$

Theorem 4.23

Let X and Y be two random variables admitting second order moments. Then:

$$\mathbb{V}\text{ar}(X + Y) = \mathbb{V}\text{ar}(X) + 2\mathbb{C}\text{ov}(X, Y) + \mathbb{V}\text{ar}(Y).$$

Proof. Exercise (*optional*). Hint: $\mathbb{V}\text{ar}(Z) = \mathbb{E}(Z^2) - (\mathbb{E}(Z))^2$. □

Corollary 4.24. If X and Y are two **independent** random variables, then:

$$\mathbb{V}\text{ar}(X + Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y),$$

provided that the above variances exist.

Example 4.25. Let $Y \sim \mathcal{B}(n, p)$. Then Y can be expressed as a sum of independent Bernoulli variables with parameter p:

$$Y = \sum_{i=1}^{n} X_i.$$

Thus, the variance of Y is given by:

$$\mathbb{V}\text{ar}(Y) = \sum_{i=1}^{n} \mathbb{V}\text{ar}(X_i) = \sum_{i=1}^{n} p(1 - p) = np(1 - p).$$

28 *optional.* You roll a die and denote by X the result of the roll. Let:

$$Y = \begin{cases} -4 & \text{if } X = 6, \\ 5 & \text{otherwise.} \end{cases}$$

1. Write the corresponding joint probability table.
2. Compute $\mathbb{C}\text{ov}(X, Y)$.
3. Are X and Y independent?

III.2 Sum of two continuous random variables

Theorem 4.26

Let X and Y be two independent continuous random variables with respective densities f_X and f_Y. Then, $Z = X + Y$ is continuous and a density of Z is given by:

$$f_Z(t) = \int_{-\infty}^{+\infty} f_X(x) f_Y(t-x) \mathrm{d}x.$$

Proof. Admitted. □

Remark 4.27. Using a substitution in the integral ($u = t - x$) one can show that the result does not depend on whether Z was written as $X + Y$ or $Y + X$.

III.3 Stability of the usual distributions

III.3.1 Binomial distribution

Proposition 4.28

Let $X \sim \mathcal{B}(n, p)$ and $Y \sim \mathcal{B}(m, p)$ be two independent random variables. Then:

$$X + Y \sim \mathcal{B}(n + m, p).$$

Proof. We will do it in class. □

Remark 4.29. X counts the number of successes after n independent Bernoulli trials with probability of success p. Y counts the number of successes after m independent Bernoulli trials with the same probability of success. Thus, $X + Y$ counts the number of successes after $n + m$ independent Bernoulli trials with probability of success p.

III.3.2 Poisson distribution

Proposition 4.30

Let $X \sim \mathcal{P}(\lambda)$ and $Y \sim \mathcal{P}(\mu)$ be two independent random variables. Then:

$$X + Y \sim \mathcal{P}(\lambda + \mu).$$

Proof. We will do it in class. □

III.3.3 Normal distribution

Proposition 4.31

Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \mathcal{N}(\mu', \sigma'^2)$ be two independent random variables. Then:

$$X + Y \sim \mathcal{N}(\mu + \mu', \sigma^2 + \sigma'^2).$$

Proof. Let $X \sim \mathcal{N}(0,1)$ have density f and $Y \sim \mathcal{N}(0,1)$ have density g. Denote by h the convolution of f and g. Then, for any $x \in \mathbb{R}$:

$$h(x) = \int_{-\infty}^{+\infty} \frac{1}{2\pi} \exp\left(\frac{-t^2}{2} - \frac{x^2 - 2tx + t^2}{2}\right) \mathrm{d}t = \frac{1}{2\pi} \mathrm{e}^{-x^2/4} \int_{-\infty}^{+\infty} \mathrm{e}^{-\,(t - x/2)^2} \mathrm{d}t$$

By substitution ($u = \sqrt{2}(t - x/2)$), we get:

$$h(x) = \frac{1}{2\pi} \mathrm{e}^{-x^2/4} \int_{-\infty}^{+\infty} \mathrm{e}^{-u^2/2} \frac{\mathrm{d}u}{\sqrt{2}} = \frac{1}{\sqrt{2\pi}\sqrt{2}} \mathrm{e}^{-x^2/(2\sqrt{2}^2)} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-u^2/2} \mathrm{d}u$$

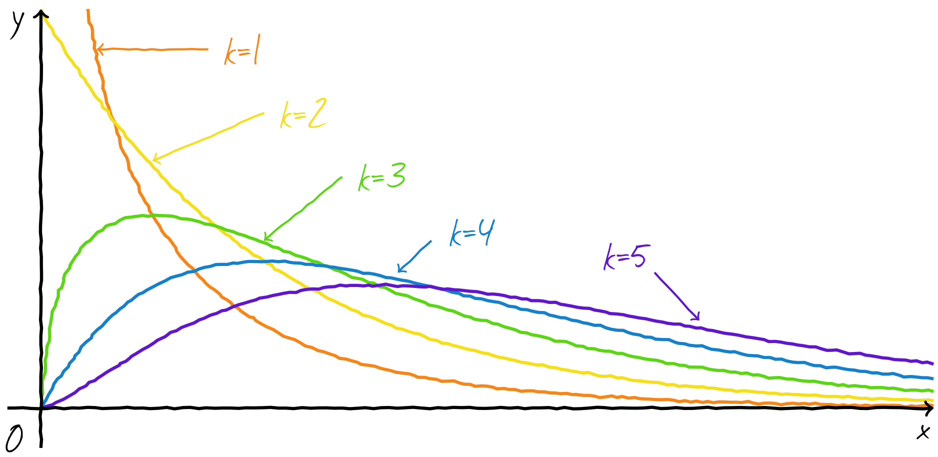
□

III.3.4 Chi-squared distribution

Definition 4.32. Let $X_1, \ldots, X_k$ be k mutually independent random variables with the same distribution $\mathcal{N}(0,1)$. The *chi-squared distribution* with k *degrees of freedom* is the distribution of the sum:

$$\sum_{i=1}^{k} X_i^2 \sim \chi_k^2.$$

Illustration 4.33. Probability density function of the chi-squared distribution with different degrees of freedom.



Proposition 4.34

Let $Z \sim \chi_k^2$. Then:

$$\mathbb{E}(Z) = k \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}(Z) = 2k$$

Proof. For the mean: exercise (*optional*). You can use the properties of the standard normal distribution and the definition of the chi-squared distribution. The result on the variance is admitted. □

I The law of large numbers

I.1 Preliminary results

Proposition 5.1 (Markov's inequality)

Let X be a non-negative random variable with finite expected value. Then, for any $\varepsilon > 0$:

$$\mathbb{P}(X \geqslant \varepsilon) \leqslant \frac{\mathbb{E}(X)}{\varepsilon}.$$

Proof. We will do it in class. □

Proposition 5.2 (Chebyshev's inequality)

Let X be a random variable with finite variance. Then, for any $\varepsilon > 0$:

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant \varepsilon\right) \leqslant \frac{\mathbb{V}\mathrm{ar}(X)}{\varepsilon^2}.$$

Proof. Exercise (*optional*): apply Markov's inequality to the random variable $Y = (X - \mathbb{E}(X))^2$. □

Remark 5.3. There is a small chance that X will be far away from its mean.

I.2 The law of large numbers

Definition 5.4. Let $(X_n)_n$ be a sequence of random variables and let X be a random variable.. We will say that $(X_n)_n$ *converges in probability* towards X if:

$$\forall \varepsilon > 0, \quad \mathbb{P}\left(|X_n - X| \geqslant \varepsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

We write:

$$X_n \xrightarrow{\mathbb{P}} X.$$

Theorem 5.5 (Law of large numbers)

Let $X_1, X_2, \ldots$ be a sequence of mutually independent[1] random variables with the same expectation μ and the same variance. Then:

$$\overline{X}_n \xrightarrow{\mathbb{P}} \mu,$$

where:

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Proof. We will do it in class. □

[1] $X_1, X_2, \ldots$ is a sequence of *mutually independent* random variables if for any finite $J \subset \mathbb{N}$, the X_j for $j \in J$ are independent

Remark 5.6. If a "large number" of dice are rolled, the average of their values will be close to 3.5 according to the law of large numbers.

29 *mandatory.* Suppose someone rolls a die and denote by:

$$X_i = \begin{cases} 1 & \text{if it lands on a 6 on the i-th roll,} \\ 0 & \text{otherwise.} \end{cases}$$

Show that, as n goes to infinity:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \frac{1}{6}.$$

30 *optional.* Let $\varphi \in \mathcal{C}([0,1])$ and let $(U_n)_n$ be a sequence of mutually independent $\mathcal{U}([0,1])$ random variables. Write:

$$\forall n \in \mathbb{N}^\star, \quad X_n = \varphi(U_n).$$

Then, the X_n's are mutually independent and have the same distribution.

1. Using the law of the unconscious statistician, show that:

$$\mathbb{E}(X_n) = \int_0^1 \varphi(t)\mathrm{d}t.$$

2. Apply the law of large numbers to the sequence $(X_n)_n$.

II Convergence in distribution

II.1 Definition

Definition 5.7. A sequence $(X_n)_n$ of random variables is said to *converge in distribution* to a random variable X if:

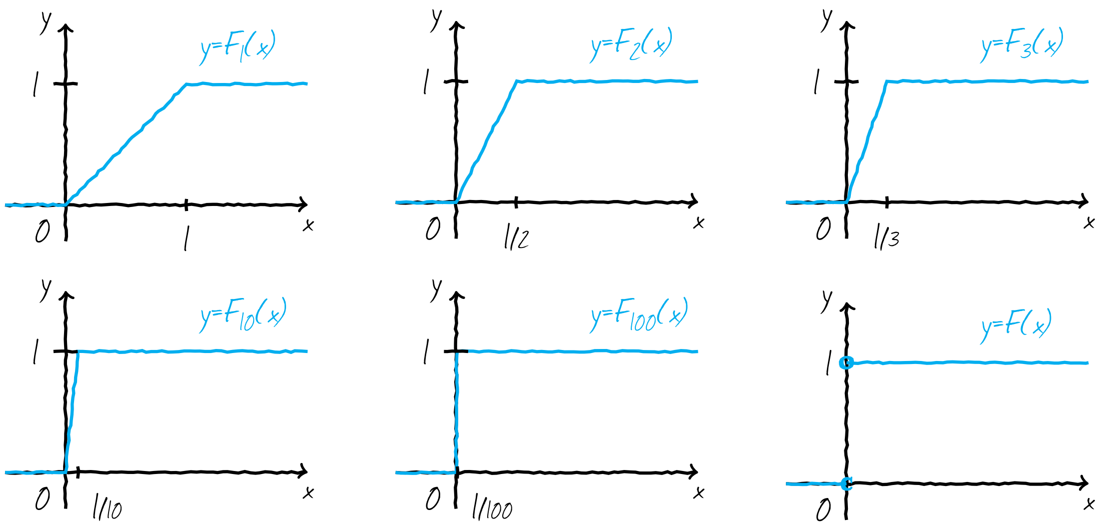
$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every $x \in \mathbb{R}$ at which F is continuous, where F and F_n are the cdfs of X and X_n, respectively. We write:

$$X_n \xrightarrow{\mathcal{D}} X.$$

Example 5.8. Let (X_n) be a sequence of random variables with $X_n \sim \mathcal{U}([0, 1/n])$. Then, (X_n) converges in distribution to X where X satisfies $\mathbb{P}(X = 0) = 1$.

Illustration 5.9. Graphs of the corresponding cdfs.



31 *optional.* Let $(X_n)_n$ be a sequence of random variables such that:

$$\forall n \in \mathbb{N}^\star, \quad X_n \sim \mathcal{U}\left(\left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\right\}\right),$$

i.e.:

$$\forall k \in [\![0, n-1]\!], \quad \mathbb{P}\left(X_n = \frac{k}{n}\right) = \frac{1}{n}.$$

1. Draw the cdf of X_2, X_5 and X_n for n "large".
2. What seems to be the limiting distribution of $(X_n)_n$?
3. Prove it.

Remark 5.10. If $(X_n)_n$ is a sequence of discrete random variables and if X is also discrete, then $(X_n)_n$ converges in distribution to X iff:

$$\forall k \in X(\Omega), \quad \mathbb{P}(X_n = k) \underset{n\to\infty}{\longrightarrow} \mathbb{P}(X = k).$$

Example 5.11. Let $(X_n)_n$ be a sequence of $\mathcal{B}(n, \lambda/n)$ variables. Then:

$$\forall k \in \mathbb{N}, \quad \mathbb{P}(X_n = k) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}.$$

Thus:

$$X_n \xrightarrow{\mathcal{D}} \mathcal{P}(\lambda).$$

This result corresponds to the Poisson approximation of the Binomial distribution.

32 *mandatory.* Let $(X_n)_n$ be a sequence of random variables with cdfs:

$$\forall n \in \mathbb{N}^\star, \quad F_{X_n}(x) = \begin{cases} 0 & \text{if } x \leqslant 0, \\ 1 - \left(1 - \frac{x}{n}\right)^n & \text{if } 0 < x \leqslant n, \\ 1 & \text{otherwise.} \end{cases}$$

Show that:

$$X_n \xrightarrow{\mathcal{D}} \mathcal{E}(1).$$

Hint:

$$\forall z \in \mathbb{C}, \quad \left(1 + \frac{z}{n}\right)^n \underset{n\to\infty}{\longrightarrow} \mathrm{e}^z.$$

II.2 The central limit theorem

Theorem 5.12 (central limit theorem)

Let $(X_n)_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with expected value μ and finite variance σ^2. Then:

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where:

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n}.$$

Proof. Admitted. □

Remark 5.13. The limiting distribution ($\mathcal{N}(0,1)$) **does not** depend on the distribution of the X_i's. When n is "large enough"[2], one can make the approximation:

$$\mathbb{P}\left(\sqrt{n}\,\frac{\overline{X}_n-\mu}{\sigma}\in A\right)\simeq\mathbb{P}(\mathcal{N}(0,1)\in A),$$

for any subset A of $\mathbb{R}$.

Example 5.14. Someone tosses a coin 100 times. What is the probability that he gets at least 50 heads? Denote by:

$$X_i=\begin{cases}1 & \text{if it lands on heads on the i-th toss,}\\ 0 & \text{otherwise.}\end{cases}$$

Then, $(X_i)_i$ is a sequence of i.i.d. random variables with expected value 0.5 and variance 0.25 (why?). Thus:

$$\mathbb{P}\left(\sum_{i=1}^{100}X_i\geqslant 50\right)=\mathbb{P}\left(\overline{X}_{100}-0.5\geqslant 0\right)=\mathbb{P}\left(\sqrt{100}\cdot\frac{\overline{X}_{100}-0.5}{\sqrt{0.25}}\geqslant 0\right)\simeq\mathbb{P}(\mathcal{N}(0,1)\geqslant 0).$$

Hence, the probability is approximately equal to 0.5.

Example 5.15 (continuity correction)**.** Someone tosses a coin 100 times. What is the probability that he gets exactly 50 heads? Using the same notations as in the previous example, one gets:

$$\mathbb{P}\left(\sum_{i=1}^{100}X_i=50\right)\simeq\mathbb{P}(\mathcal{N}(0,1)=0)=0.$$

To tackle this problem, we remark that:

$$\mathbb{P}\left(\sum_{i=1}^{100}X_i=50\right)=\mathbb{P}\left(49.5\leqslant\sum_{i=1}^{100}X_i\leqslant 50.5\right),$$

and then we use the normal approximation to get

$$\mathbb{P}\left(49.5\leqslant\sum_{i=1}^{100}X_i\leqslant 50.5\right)\simeq\mathbb{P}\big(-0.1\leqslant\mathcal{N}(0,1)\leqslant 0.1\big).$$

Example 5.16. Show that:

$$\mathrm{e}^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}\underset{n\to\infty}{\longrightarrow}\frac{1}{2}.$$

Let $(X_n)_n$ be a sequence of independent random variables with probability distribution $\mathcal{P}(1)$. Recall that, in this case:

$$X_1+\ldots+X_n\sim\mathcal{P}(1+\ldots+1)=\mathcal{P}(n),$$

and that $\mathbb{E}(X_1)=\mathbb{V}\mathrm{ar}(X_1)=1$. Then:

$$\mathrm{e}^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}=\mathbb{P}\left(\mathcal{P}(n)\leqslant n\right)=\mathbb{P}\left(\sum_{k=1}^{n}X_k\leqslant n\right)=\mathbb{P}\left(\overline{X}_n-1\leqslant 0\right)=\mathbb{P}\left(\sqrt{n}\cdot\frac{\overline{X}_n-1}{1}\leqslant 0\right).$$

But, by the central limit theorem:

$$\mathbb{P}\left(\sqrt{n}\cdot\frac{\overline{X}_n-1}{1}\leqslant 0\right)\underset{n\to\infty}{\longrightarrow}\mathbb{P}(\mathcal{N}(0,1)\leqslant 0)=\frac{1}{2}.$$

[2] while it is unclear why, n is usually supposed to be greater than 30.

33 *optional.* A freight elevator can transport a maximum of 9800 pounds. Suppose a load of cargo containing 49 boxes must be transported via the elevator. Experience has shown that the weight of boxes of this type of cargo has a distribution with mean $\mu = 205$ pounds and standard deviation $\sigma = 15$ pounds. Based on this information, what is the probability that all 49 boxes can be safely loaded onto the freight elevator and transported?
Hint: write $X_1, \dots, X_{49}$ the weights of the boxes and try to express the corresponding probability with the sum $\sum X_i$.

III Classical approximations

III.1 Binomial approximation to the hypergeometric

Suppose one draws without replacement n items in a population of size N with m possible successes. The distribution corresponding to this experiment is the hypergeometric one.
If N, the population size, is "large enough" and if n, the number of draws, is "small enough", then the experiment is similar to a binomial one (i.e. n draws with replacement and a probability of success $p = m/N$).

Proposition 5.17

Suppose that $(m_N)_N$ is a sequence of integers satisfying $m_N/N \longrightarrow p$ as $N \to \infty$. Then, as N goes to infinity:

$$\mathcal{H}(N, n, m_N) \xrightarrow{\mathcal{D}} \mathcal{B}(n, p).$$

Remark 5.18. We will approximate the hypergeometric distribution $\mathcal{H}(N, n, m)$ with the binomial $\mathcal{B}(n, m/N)$ when $n < 0.1N$.

III.2 Normal approximation to the binomial

Proposition 5.19

Let $(Y_n)_n$ be a sequence of random variables with $Y_n \sim \mathcal{B}(n, p)$. Then:

$$\sqrt{n}\,\frac{\frac{1}{n}Y_n - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Proof. We will do it in class. □

Remark 5.20. We will approximate the binomial distribution $\mathcal{B}(n, p)$ with the normal $\mathcal{N}(np, p(1-p))$ when $n \geqslant 30$, $np \geqslant 15$ and $np(1-p) > 5$.

Remark 5.21 (continuity correction)**.** As we already noticed in example 5.15, we can't approximate $\mathbb{P}(Y_n = k)$ with $\mathbb{P}(\mathcal{N}(np, p(1-p)) = k)$. Instead, we will write:

$$\mathbb{P}(Y_n = k) = \mathbb{P}(k - .5 \leqslant Y_n \leqslant k + .5) \simeq \mathbb{P}\big(k - .5 \leqslant \mathcal{N}(np, p(1-p)) \leqslant k + .5\big).$$

34 *mandatory.* Sixty two percent of 12th graders attend school in a particular urban school district. If a sample of 500 12th grade children are selected, find the probability that at least 290 are actually enrolled in school.

Hint: suppose that the variables X_i are independent Bernoulli variables defined by:

$$X_i = \begin{cases} 1 & \text{if the i-th child is in school,} \\ 0 & \text{otherwise,} \end{cases}$$

and write the corresponding probability using the sum $\sum X_i$.

I Introduction

Example 6.1. Suppose you work for a pharmaceutical company and developed a new drug: how can you *estimate* its efficiency without testing it on the whole population?

Example 6.2. Suppose one wants to know the average height of a population of 10,000 men. He lacks time and can't measure everyone's height. An *estimate* of the average height is the average height of 100 randomly selected men in the population. Is this solution legitimate?

I.1 First definitions

Definition 6.3. Let X be a random variable. A *(random) sample of size n* (of X) is a vector $(X_1, \dots, X_n)$ of i.i.d. (independent and identically distributed) random variables with the same distribution as X.

Definition 6.4. An *observation* of a sample $(X_1, \dots, X_n)$ is a *realization* of the sample, usually denoted by lower case letters $(x_1, \dots, x_n)$.

Example 6.5. Let X be uniformly distributed on $[\![1, 6]\!]$. To get an observation of size 10 of X, one could roll 10 times a fair die:

$$(x_1, \dots, x_{10}) = (1, 5, 2, 6, 4, 4, 1, 2, 3, 4).$$

Definition 6.6. Let θ be a parameter that depends on the distribution of some random variable X[1]. An *estimator* of θ is a function T_n of a sample $(X_1, \dots, X_n)$ of X that is used to approximate θ.

Example 6.7. Suppose you have an unfaire die and want to approximate the probability θ that it lands on a 6. A solution to this problem is to roll a "large" number of times the die and to compute the average number of sixes. Why? Denote by:

$$X_i = \begin{cases} 1 & \text{if you get a 6 on the } i\text{-th roll,} \\ 0 & \text{otherwise.} \end{cases}$$

Then, by the law of large numbers:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \frac{1}{6}.$$

Thus, $\frac{1}{n}\sum_{i=1}^{n} X_i$ is an estimator of θ.

35 *optional.* Let $(X_1, \dots, X_n)$ be a sample of an unknown variable X with mean μ and variance σ^2. Give an estimator of μ and σ^2.

I.2 Properties of an estimator

Definition 6.8. Let T_n be an estimator of θ such that $\mathbb{E}(T_n)$ exists. The *bias* of T_n is:

$$b(T_n) = \mathbb{E}(T_n) - \theta.$$

Definition 6.9. An estimator T_n of θ is *unbiased* if $b(T_n) = 0$.

Definition 6.10. An estimator T_n of θ is *asymptotically unbiased* if $b(T_n) \underset{n\to\infty}{\to} 0$.

36 *mandatory.* Show that an unbiased estimator of the mean of a sample $(X_1, \ldots, X_n)$ is:

$$T_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Remark 6.11. An estimator can be unbiased and still "not good". Consider a unfair die roll for example and suppose you want to estimate θ, the probability to get a 6. Denote by:

$$X_i = \begin{cases} 1 & \text{if you get a 6 on the } i\text{-th roll,} \\ 0 & \text{otherwise.} \end{cases}$$

Then, an unbiased estimator of θ is:

$$T_n(X_1, \ldots, X_n) = X_1,$$

as $\mathbb{E}(X_1) = \theta$.

Definition 6.12. A sequence of estimators $(T_n)_n$ of θ is *consistent* if:

$$T_n \xrightarrow{\mathbb{P}} \theta.$$

37 *mandatory.* Show that a consistent estimator of the mean of a sample $(X_1, \ldots, X_n)$ is:

$$T_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Definition 6.13. The mean squared error of an estimator T_n of θ is:

$$\mathbb{E}((T_n - \theta)^2).$$

Remark 6.14. One could consider the quantity $\mathbb{E}(|T_n - \theta|)$ to assess the quality of T_n, but computations are quite easier with the mean squared error.

Lemma 6.15. Let (T_n) be a sequence of estimators of θ such that $\mathbb{V}\mathrm{ar}(T_n)$ exists. Then:

$$\mathbb{E}((T_n - \theta)^2) = b(T_n)^2 + \mathbb{V}\mathrm{ar}(T_n).$$

Proof. Exercise (*mandatory*). Expand the square and use the linearity of the mean. Hint:

$$\mathbb{E}(T_n^2) = \mathbb{V}\mathrm{ar}(T_n) + (\mathbb{E}(T_n))^2.$$

□

Proposition 6.16

Let $(T_n)_n$ be a sequence of estimators of θ such that:

(i) T_n is unbiased;

(ii) $\mathbb{V}\mathrm{ar}(T_n) \underset{n\to\infty}{\to} 0$.

Then, $(T_n)_n$ is consistent.

Proof. Exercise (*optional*). Hint: use Chebyshev's inequality. □

II Classical estimators

II.1 Estimator of the mean

Proposition 6.17

Let X be a random variable with expected value μ and let $(X_1, \ldots, X_n)$ be a random sample. Then,

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is a consistent and unbiased estimator of μ.

Proof. See exercises 36 and 37. □

II.2 Estimator of the variance

Proposition 6.18

Let X be a random variable with expected value μ and variance σ^2. Suppose μ is known and let $(X_1, \ldots, X_n)$ be a random sample. Then,

$$\Sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

is a consistent and unbiased estimator of σ^2.

Proof. We will do it in class. □

Proposition 6.19

Let X be a random variable with unknown expected value μ and variance σ^2. Let also $(X_1, \ldots, X_n)$ be a random sample ($n > 1$). Then,

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

is a consistent and unbiased estimator of σ^2.

Proof. Admitted. □

I Introduction

In the previous chapter, we gave estimators for the mean and the variance of a sample. We also tried to give a definition of a "good" estimator (unbiased and/or consistent).
In this chapter, instead of giving an approximation of a parameter, we will try to give an interval in which we are «pretty confident» the parameter is.

In this chapter we will consider:

- X a random variable with unknown parameter θ;
- $(X_1, \ldots, X_n)$ a random sample from X;
- $T_n(X_1, \ldots, X_n)$ an estimator of θ;
- $\alpha \in]0,1[$ a *risk* factor.

II Confidence interval for the mean

II.1 Normally distributed variable

II.1.1 First case: known variance

Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ where μ is unknown and $\sigma^2 > 0$ is known. We know that

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

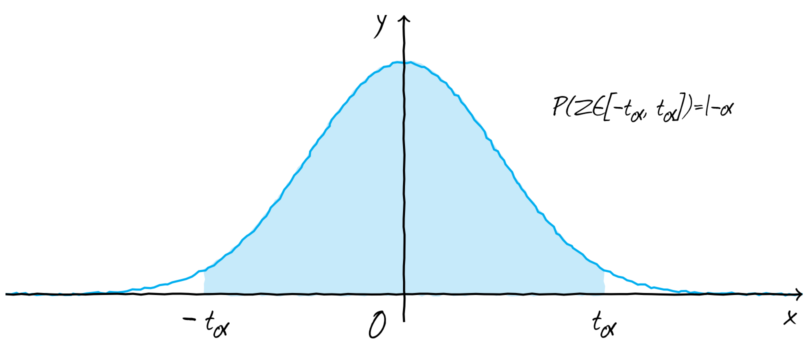
is a consistent and unbiased estimator of the mean μ. Moreover, using the properties of the normal distribution:

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Let:

$$Z = \sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \sim \mathcal{N}(0,1).$$

The next step is to find a *confidence interval* for Z, that is, an interval I_α such that $\mathbb{P}(Z \in J_\alpha) = 1-\alpha$. There are a lot of choices for the set I_α and, unless there is a good reason not to do so, we will choose the smallest one.



Let t_α be the real number satisfying:

$$\mathbb{P}(Z \in [-t_\alpha, t_\alpha]) = 1 - \alpha$$

Then, we can write:

$$1-\alpha = \mathbb{P}\left(-t_\alpha \leqslant \sqrt{n}\,\frac{\overline{X}_n-\mu}{\sigma} \leqslant t_\alpha\right) = \mathbb{P}\left(\overline{X}_n - t_\alpha\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \overline{X}_n + t_\alpha\frac{\sigma}{\sqrt{n}}\right).$$

Hence, a confidence interval for the mean of the sample is:

$$I_\alpha = \left[\overline{X}_n - t_\alpha\frac{\sigma}{\sqrt{n}}, \overline{X}_n + t_\alpha\frac{\sigma}{\sqrt{n}}\right].$$

Remark 7.1. The size of the confidence interval depends on the variance, the risk factor α and the size of the sample: the larger the sample, the more accurate the confidence interval.

Example 7.2 (from Wikipedia)**.** A machine fills cups with a liquid, and is supposed to be adjusted so that the content of the cups is 250g of liquid. As the machine cannot fill every cup with exactly 250g, the content added to individual cups shows some variation, and is considered a random variable X. This variation is assumed to be normally distributed with unknown mean μ and variance $\sigma^2 = 6.25$. A sample of $n = 25$ cups of liquid are chosen at random and the cups are weighed. The resulting measured masses of liquid are $X_1, \ldots, X_{25}$, a random sample from X. An estimator of the mean is:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The sample showed actual weights $x_1, \ldots, x_{25}$ with mean:

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 250.2g$$

Thus, an estimator of the mean is $\overline{x} = 250.2g$. To construct a 95%-confidence interval for the mean μ, we remark that:

$$Z = \sqrt{n}\,\frac{\overline{X}_n-\mu}{\sigma} \sim \mathcal{N}(0,1).$$

Looking at the standard normal table, we have:

$$\mathbb{P}\left(-1.96 \leqslant \mathcal{N}(0,1) \leqslant 1.96\right) = 95\%.$$

Thus:

$$95\% = \mathbb{P}\left(-1.96 \leqslant \sqrt{n}\,\frac{\overline{X}_n-\mu}{\sigma} \leqslant 1.96\right) = \mathbb{P}\left(\overline{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \leqslant \mu \leqslant \overline{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right).$$

Which yields:

$$\mathbb{P}\left(\mu \in \left[\overline{X}_n - 0.98, \overline{X}_n + 0.98\right]\right) = 95\%.$$

We can then give a 95%-confidence interval for the mean:

$$I = [249.22; 251.18].$$

38 *mandatory.* Change the size of the sample in the preceding example to $n = 50$ (keep the same standard deviation) and compute the new confidence interval: is this result surprising?

II.1.2 Second case: unknown variance

Suppose $X \sim \mathcal{N}(\mu, \sigma^2)$ where μ and $\sigma^2 > 0$ are unknown. We know that

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is a consistent and unbiased estimator of the mean μ. Moreover,

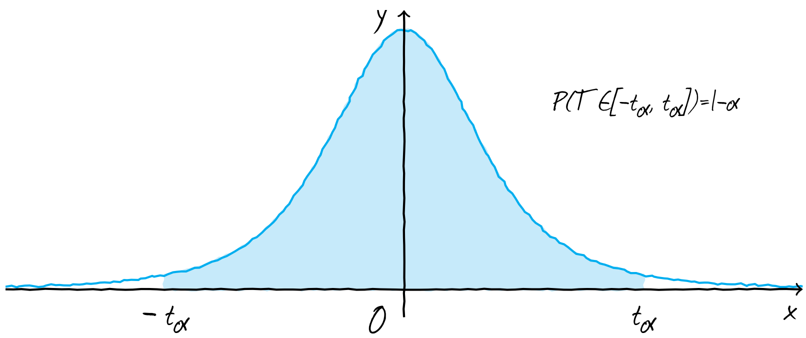
$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

is a consistent and unbiased estimator of σ^2. One can show that in this situation:

$$T = \sqrt{n}\,\frac{\overline{X}_n - \mu}{S_n} \sim \mathcal{T}_{n-1},$$

where $\mathcal{T}_{n-1}$ is the Student's t-distribution[1] (or simply t-distribution) with $n-1$ degrees of freedom.

Illustration 7.3. Confidence interval for a t-distributed random variable.



To find a confidence interval for the mean, we will use the same method as in the previous section. The only difference is that the number t_α will come from the Student distribution table instead of the standard normal one.

Remark 7.4. For large n ($n > 100$), the Student distribution can be approximated with the standard normal one.

Example 7.5. Let's go back to example 7.2, but suppose now that the variance and the mean are both unknown. To construct a confidence interval for the mean, we compute:

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 250.2g$$

and

$$s_n^2 = \frac{1}{25-1}\sum_{i=1}^{25}(x_i - \overline{x})^2 = 6.4g^2.$$

Those quantities were computed using a calculator and the observed weights $x_1, \ldots, x_n$. Looking at the t-distribution table, we have:

$$\mathbb{P}\left(-2.06 \leqslant \mathcal{T}_{24} \leqslant 2.06\right) = 95\%.$$

[1]The t-distribution is symmetric and bell-shaped and looks like the normal distribution (especially for n "large").

Thus:

$$95\% = \mathbb{P}\left(-2.064 \leqslant \sqrt{n}\,\frac{\overline{X}_n - \mu}{S_n} \leqslant 2.064\right) = \mathbb{P}\left(\overline{X}_n - 2.064\frac{S_n}{\sqrt{n}} \leqslant \mu \leqslant \overline{X}_n + 2.064\frac{S_n}{\sqrt{n}}\right),$$

which yields the following 95%-confidence interval for the mean μ of the sample

$$I = \left[\overline{x}_n - 2.064\frac{s_n}{\sqrt{25}}, \overline{x}_n + 2.064\frac{s_n}{\sqrt{25}}\right] = [249.16; 251.24]\,.$$

39 *mandatory.* Change the estimator of the variance in the preceding example to $s_n^2 = 3$ and compute the new confidence interval: is this result surprising?

40 *optional.* Construct a 99%-confidence interval for the mean height of the students in our school given that the mean height of 100 randomly selected student is $\bar{x} = 1.75$ with observed standard deviation $s = 0.06$.

II.2 Non-normally distributed variables

Suppose X be a random variable with unknown expected value μ and unknown variance σ^2. By the central limit theorem, we know that

$$\sqrt{n}\,\frac{\overline{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

As σ is unknown, we can't use this result as is. But, one can show that the consistency of S_n^2 implies that:

$$Z = \sqrt{n}\,\frac{\overline{X}_n - \mu}{S_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

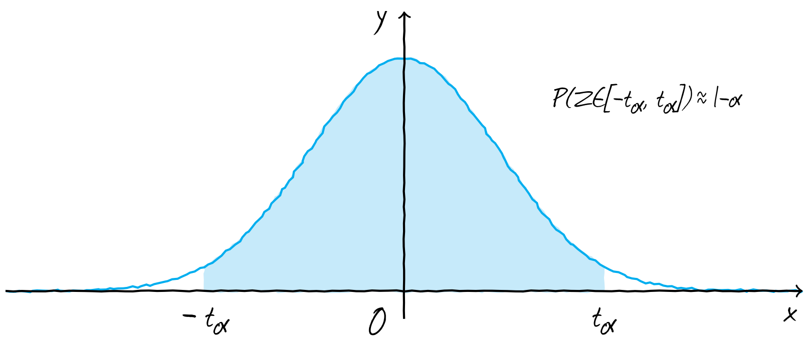
where:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

We will then use a confidence interval for the standard normal distribution as an approximation of a confidence interval for the variable Z:

$$1 - \alpha \simeq \mathbb{P}(-t_\alpha \leqslant Z \leqslant t_\alpha),$$

where t_α is found in the standard normal table.



Thus, we can write:

$$1 - \alpha \simeq \mathbb{P}\left(-t_\alpha \leqslant \sqrt{n}\,\frac{\overline{X}_n - \mu}{S_n} \leqslant t_\alpha\right) = \mathbb{P}\left(\overline{X}_n - t_\alpha\frac{S_n}{\sqrt{n}} \leqslant \mu \leqslant \overline{X}_n + t_\alpha\frac{S_n}{\sqrt{n}}\right).$$

Hence, an *asymptotic* confidence interval for the mean of the sample is:

$$I_\alpha = \left[\overline{X}_n - t_\alpha\frac{S_n}{\sqrt{n}}, \overline{X}_n + t_\alpha\frac{S_n}{\sqrt{n}}\right].$$

II.3 Confidence interval for a proportion

Example 7.6. Suppose you have an unfair die and denote by X the corresponding distribution. You can estimate $\mathbb{P}(X = 6)$ by using the law of large numbers (as seen in chapter 6): you roll the die a "large number" of times and count the proportion of sixes. But what if you want to construct a confidence interval for $\mathbb{P}(X = 6)$?

In this section, we assume that $X \sim \mathcal{B}(p)$ and that $(X_1, \dots, X_n)$ is a random sample of X. We also suppose that $n \geqslant 30$, $np \geqslant 5$ and $np(1-p) \geqslant 5$. As $p = \mathbb{E}(X)$, an estimator of p is given by:

$$\overline{X}_n = \frac{1}{n} \sum_{k=1}^{n} X_k.$$

Moreover, by the central limit theorem:

$$\sqrt{n}\,\frac{\overline{X} - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

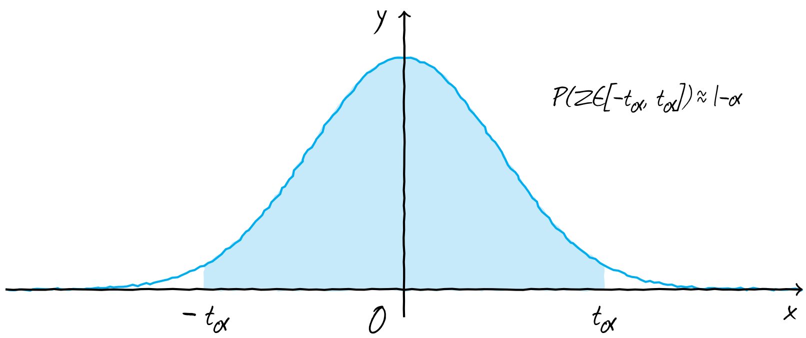
One can show that the above convergence still holds when the variance $p(1-p)$ is approximated by $\overline{X}(1-\overline{X})$, that is:

$$Z = \sqrt{n}\,\frac{\overline{X} - p}{\sqrt{\overline{X}(1-\overline{X})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Thus, we can construct a confidence interval using the same technique as in section II.2: we use a confidence interval for the standard normal distribution as an approximation of a confidence interval for the variable Z:

$$1 - \alpha \simeq \mathbb{P}(-t_\alpha \leqslant Z \leqslant t_\alpha),$$

where t_α is found in the standard normal table.



We can then write:

$$\begin{aligned} 1 - \alpha &\simeq \mathbb{P}\left(-t_\alpha \leqslant \sqrt{n}\,\frac{\overline{X} - p}{\sqrt{\overline{X}(1-\overline{X})}} \leqslant t_\alpha \right) \\ &= \mathbb{P}\left(\overline{X} - t_\alpha \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \leqslant p \leqslant \overline{X} + t_\alpha \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \right). \end{aligned}$$

Hence, an *asymptotic* confidence interval for p is:

$$I_\alpha = \left[\overline{X} - t_\alpha \sqrt{\frac{\overline{X}(1-\overline{X})}{n}}, \overline{X} + t_\alpha \sqrt{\frac{\overline{X}(1-\overline{X})}{n}} \right].$$

III Confidence interval for the variance of a normal variable

In this section we assume that $X \sim \mathcal{N}(\mu, \sigma^2)$ and that $(X_1, \ldots, X_n)$ is a random sample of X. We want to give a confidence interval for σ^2.

III.1 First case: known expected value

In this section, we suppose that we know the value of the parameter μ and we construct a confidence interval for σ^2.

41 *mandatory.* Let $k \in [\![1, n]\!]$. Compute $\mathbb{E}(Z_k)$ and $\mathbb{V}\text{ar}(Z_k)$ where:

$$Z_k = \frac{X_k - \mu}{\sigma},$$

and where $X_k \sim \mathcal{N}(\mu, \sigma^2)$. Using proposition 4.31, deduce the probability distribution of Z_k.

Recall that if $Z_1, \ldots, Z_n$ are mutually independent random variables with the same distribution $\mathcal{N}(0, 1)$, then:

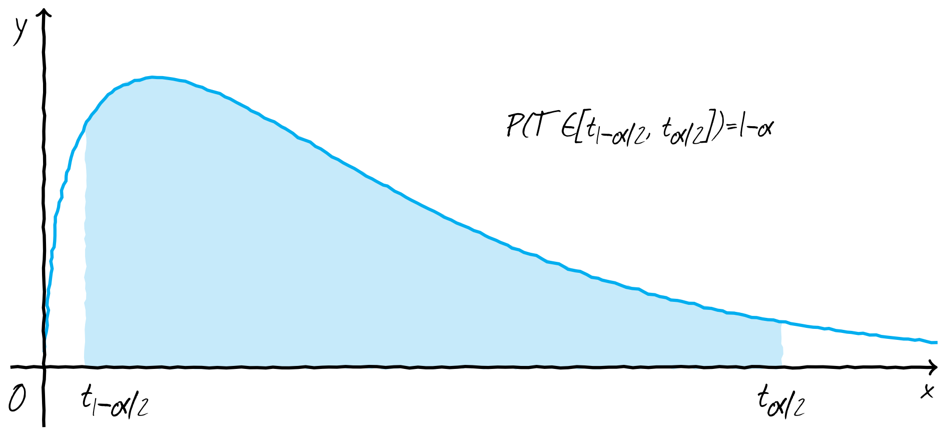
$$\sum_{k=1}^{n} Z_k^2 \sim \chi_n^2.$$

From exercise 41, we can deduce that:

$$T = \sum_{k=1}^{n} \left(\frac{X_k - \mu}{\sigma}\right)^2 \sim \chi_n^2.$$

Now, looking at the χ^2 table, we find the real numbers $t_{\alpha/2}$ and $t_{1-\alpha/2}$ such that:

$$\mathbb{P}\left(\chi_n^2 \leqslant t_{\alpha/2}\right) = 1 - \alpha/2$$
$$\mathbb{P}\left(\chi_n^2 \leqslant t_{1-\alpha/2}\right) = \alpha/2.$$



Then,

$$1 - \alpha = \mathbb{P}\left(t_{1-\alpha/2} \leqslant \frac{1}{\sigma^2}\sum_{k=1}^{n}(X_k - \mu)^2 \leqslant t_{\alpha/2}\right),$$

which yields a confidence interval for σ^2:

$$I_\alpha = \left[\frac{1}{t_{\alpha/2}}\sum_{k=1}^{n}(X_k - \mu)^2, \frac{1}{t_{1-\alpha/2}}\sum_{k=1}^{n}(X_k - \mu)^2\right].$$

III.2 Second case: unknown expected value

In this section, we suppose that we don't know the value of the parameter μ and we construct a confidence interval for σ^2. Let:

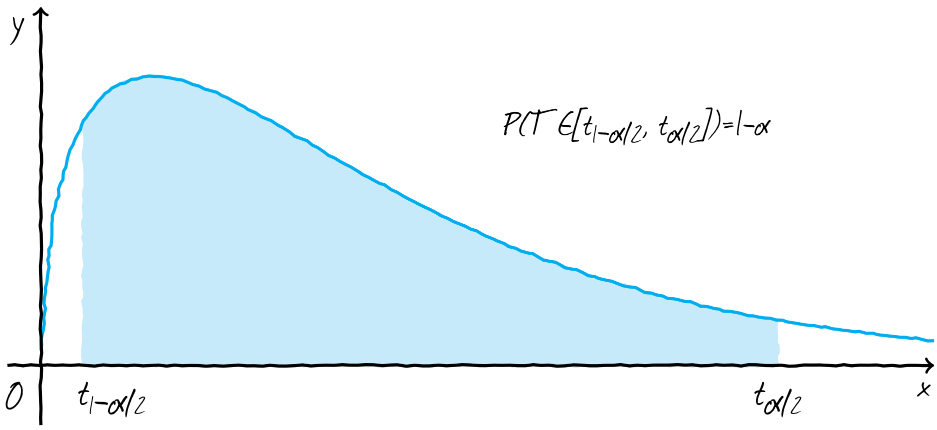
$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Then S_n^2 is a consistent and unbiased estimator of σ^2. Moreover, it can be shown that:

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}.$$

We can then construct a confidence interval using the same technique as in the previous section. Looking at the χ^2 table, we find the real numbers $t_{\alpha/2}$ and $t_{1-\alpha/2}$ such that:

$$\mathbb{P}\left(\chi^2_{n-1} \leqslant t_{\alpha/2}\right) = 1-\alpha/2$$
$$\mathbb{P}\left(\chi^2_{n-1} \leqslant t_{1-\alpha/2}\right) = \alpha/2.$$



Then,

$$1-\alpha = \mathbb{P}\left(t_{1-\alpha/2} \leqslant \frac{(n-1)S_n^2}{\sigma^2} \leqslant t_{\alpha/2}\right),$$

which yields a confidence interval for σ^2:

$$I_\alpha = \left[\frac{(n-1)S_n^2}{t_{\alpha/2}}, \frac{(n-1)S_n^2}{t_{1-\alpha/2}}\right].$$

I Introduction

In the previous chapters, we dealt with estimating parameters and constructing confidence intervals. Another problematic in statistics is what we call "hypothesis testing".
Suppose you have a possibly unfair die: how can you test whether or not it is fair? A chocolate bar factory advertises a 10% sugar content: how can you check that fact?

In this chapter we will consider:

- X a random variable with uknown parameter θ;
- $(X_1, \ldots, X_n)$ a random sample from X;
- $\alpha \in]0,1[$ a *risk* factor.

II Testing a proportion

In this section, we give a method to test whether a proportion is equal to a given value.
Suppose $X \sim \mathcal{B}(p)$ where p is unknown, let $(X_1, \ldots, X_n)$ be a random sample of X and let $p_0 \in (0,1)$. We want to test:

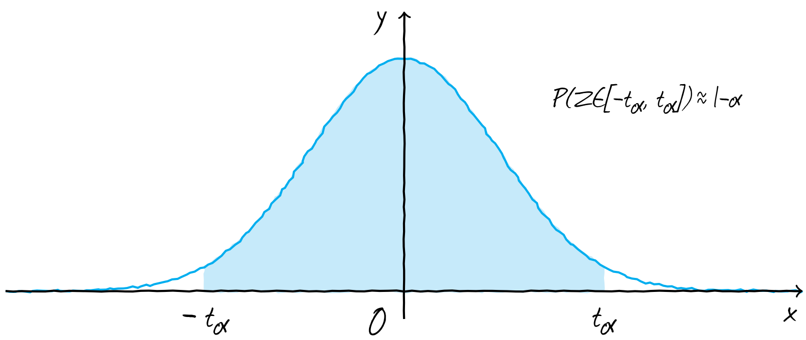
$$\mathcal{H}_0 \ : \ p = p_0 \qquad \text{against} \qquad \mathcal{H}_1 \ : \ p \neq p_0.$$

In order to do so, we choose a *risk factor* $\alpha \in (0,1)$ and we suppose that $\mathcal{H}_0$ is true. Under this assumption, we know that:

$$Z = \sqrt{n}\,\frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)}} \sim_\infty \mathcal{N}(0,1).$$

Let $t_\alpha \in \mathbb{R}$ satisfy:

$$\mathbb{P}(-t_\alpha \leqslant \mathcal{N}(0,1) \leqslant t_\alpha) = 1 - \alpha.$$



Then, under $\mathcal{H}_0$:

$$1 - \alpha \simeq \mathbb{P}(-t_\alpha \leqslant Z \leqslant t_\alpha) = \mathbb{P}\left(p_0 - t_\alpha\sqrt{\frac{p_0(1-p_0)}{n}} \leqslant \overline{X}_n \leqslant p_0 + t_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\right).$$

Thus, if $\mathcal{H}_0$ is true, there is a probability $1-\alpha$ that $\overline{X}_n$ will be in the interval:

$$I_\alpha = \left[p_0 - t_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}, p_0 + t_\alpha\sqrt{\frac{p_0(1-p_0)}{n}}\right].$$

The converse of this implication tells us that if $\overline{X}_n$ is not in I_α then there is a small chance (α) that $\mathcal{H}_0$ is true.
Given an observed sample $(x_1, \ldots, x_n)$, we will compute the observed mean $\overline{x}_n$ and check wheter it is in I_α or not:

- if $\overline{x}_n \notin I_\alpha$, then we will reject $\mathcal{H}_0$.
- if $\overline{x}_n \in I_\alpha$, then we will not reject $\mathcal{H}_0$

Remark 8.1. Not rejecting $\mathcal{H}_0$ **is not** satisfying because in this set-up, we can't compute the *type II error* which is:

$$\beta = \mathbb{P}\left(\text{accept } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ false}\right).$$

That's why we say we "don't reject $\mathcal{H}_0$" instead of we "accept $\mathcal{H}_0$". On the other hand, the risk of rejecting $\mathcal{H}_0$ while it is true is controlled by the factor α:

$$\alpha = \mathbb{P}\left(\text{reject } \mathcal{H}_0 \mid \mathcal{H}_0 \text{ true}\right),$$

which is called the *type I error*.

Example 8.2. You have a possibly unfair coin. Let $\alpha = 5\%$ and write:

$$X_i = \begin{cases} 1 & \text{if it lands on Heads on the } i\text{th trial,} \\ 0 & \text{otherwise.} \end{cases}$$

We want to test:

$$\mathcal{H}_0 \ : \ p = 0.5 \qquad \text{against} \qquad \mathcal{H}_1 \ : \ p \neq 0.5,$$

at a risk level of 5%, and we only have the time to toss our coin 50 times. Under the assumption that $\mathcal{H}_0$ is true, we know that the statistics Z is approximately normally distributed with parameters $(0, 1)$ where:

$$Z = \sqrt{n}\,\frac{\overline{X}_n - p_0}{\sqrt{p_0(1-p_0)}}.$$

Thus (see above), there is a 95% chance that $\overline{X}_{50}$ is in the interval:

$$\left[0.5 - 1.96\sqrt{\frac{0.5 \times 0.5}{50}}, 0.5 + 1.96\sqrt{\frac{0.5 \times 0.5}{50}}\right] \simeq [0.36; 0.64]$$

Suppose you tossed your coin 50 times and got 30 Heads. Then, $\overline{x} = 0.6$: we cannot reject $\mathcal{H}_0$. Suppose you tossed another coin 50 times and got 17 Heads. Then, $\overline{x} = 0.34$ and we can reject $\mathcal{H}_0$.

III Testing the mean of a normal sample

III.1 Testing an equality

Suppose $(X_1, \dots, X_n)$ is a random sample of the $\mathcal{N}(\mu, \sigma^2)$ distribution. Let $\mu_0 \in \mathbb{R}$: how can one test whether $\mu = \mu_0$ or not? Denote by:

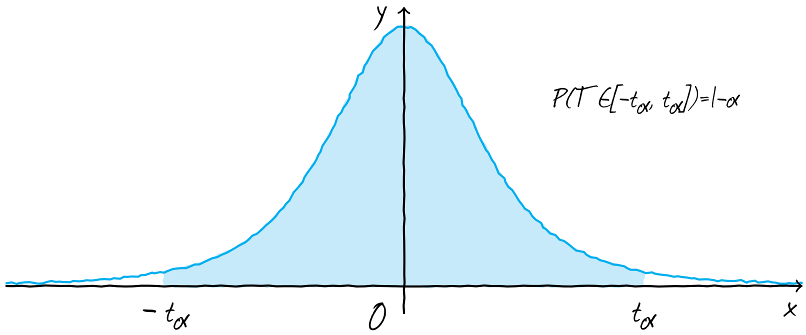
$$\mathcal{H}_0 \ : \ \mu = \mu_0 \qquad \text{and} \qquad \mathcal{H}_1 \ : \ \mu \neq \mu_0.$$

Let $\alpha \in (0, 1)$ be a *risk factor* and suppose that $\mathcal{H}_0$ is true. Then, one can show that:

$$T = \sqrt{n}\,\frac{\overline{X}_n - \mu_0}{S_n} \sim \mathcal{T}_{n-1}.$$

Looking at the Student's table, one can find t_α such that:

$$\mathbb{P}\left(-t_\alpha \leqslant Z \leqslant t_\alpha\right) = 1 - \alpha.$$



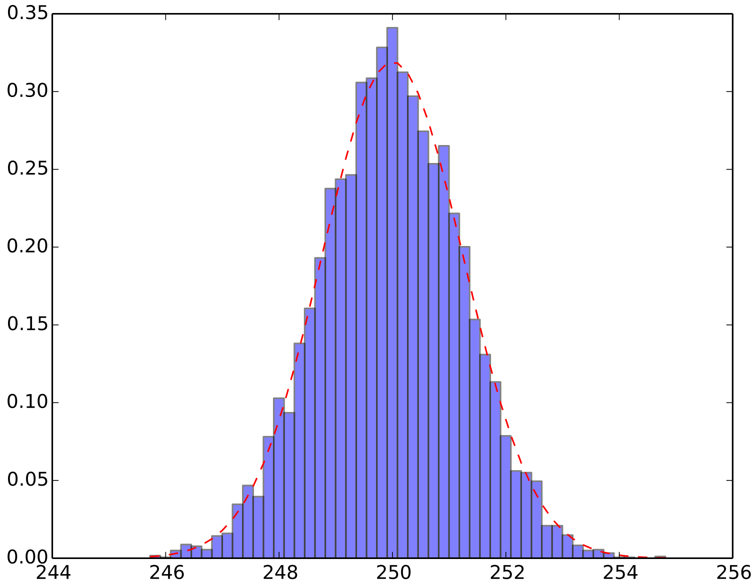
Thus, **if $\mathcal{H}_0$ is true**, then:

$$\mathbb{P}\left(\overline{X} \in \left[\mu_0 - t_\alpha \frac{S_n}{\sqrt{n}}; \mu_0 + t_\alpha \frac{S_n}{\sqrt{n}}\right]\right) = 1 - \alpha.$$

Given an observation $(x_1, \ldots, x_n)$ of $(X_1, \cdots, X_n)$, the correponding decision rule is:

- if $\overline{x} \in [\mu_0 - t_\alpha s_n/\sqrt{n}; \mu_0 + t_\alpha s_n/\sqrt{n}]$ then don't reject $\mathcal{H}_0$ (with an unknown risk),
- otherwise, reject $\mathcal{H}_0$ (with a risk α).

Example 8.3. Coffee beans bags are weighted at the end of a production line. The average weight is $250.2g$ and the standard deviation of the sample is 1.25. The results of the $n = 70$ weights are shown in the following histogram



Looking at the data, one can assume that the underlying distribution is a normal one. Denote by $(X_1, \ldots, X_n)$ a sample of $\mathcal{N}(\mu, \sigma^2)$, the distribution of the weight of the coffee bean bags. Suppose the company wants to test with a risk of 1%:

$$\mathcal{H}_0 \ : \ \mu = 250 \qquad \text{against} \qquad \mathcal{H}_1 \ : \ \mu \neq 250.$$

Then, **if $\mathcal{H}_0$ is true**:

$$\mathbb{P}\left(-2.632 \leqslant T \leqslant 2.632\right) = 99\%,$$

where

$$T = \sqrt{n}\,\frac{\overline{X}_n - \mu_0}{S_n} \sim T_{69}.$$

Hence:

$$"\mathcal{H}_0 \overset{99\%}{\Longrightarrow} \overline{X} \in \left[\mu_0 - t_\alpha \frac{S_n}{\sqrt{n}}; \mu_0 + t_\alpha \frac{S_n}{\sqrt{n}}\right]"$$

Here,

$$\overline{x} = 250.2 \in \left[250 - 2.632 \times 1.25/\sqrt{70}; 250 + 2.632 \times 1.25/\sqrt{70}\right] = [249.60; 250.39]$$

We can not reject $\mathcal{H}_0$ but we can not accept $\mathcal{H}_1$ either as there is no way of knowing the risk we would take by accepting $\mathcal{H}_1$.

III.2 Testing an inequality

Suppose $(X_1, \ldots, X_n)$ is a random sample of the $\mathcal{N}(\mu, \sigma^2)$ distribution. Let $\mu_0 \in \mathbb{R}$: how can one test whether $\mu \leqslant \mu_0$ or not? Denote by:

$$\mathcal{H}_0 \ : \ \mu \leqslant \mu_0 \qquad \text{and} \qquad \mathcal{H}_1 \ : \ \mu > \mu_0.$$

Let $\alpha \in (0,1)$ be a *risk factor* and suppose that $\mathcal{H}_0$ is true. Then, one can show that:

$$Z = \sqrt{n}\,\frac{\overline{X}_n - \mu_0}{S_n} = \sqrt{n}\,\frac{\overline{X}_n - \mathbb{E}(X_1)}{S_n} + \frac{\sqrt{n}}{S_n}\left(\mathbb{E}(X_1) - \mu_0\right) \sim \mathcal{T}_{n-1} + \frac{\sqrt{n}}{S_n}\left(\mathbb{E}(X_1) - \mu_0\right).$$

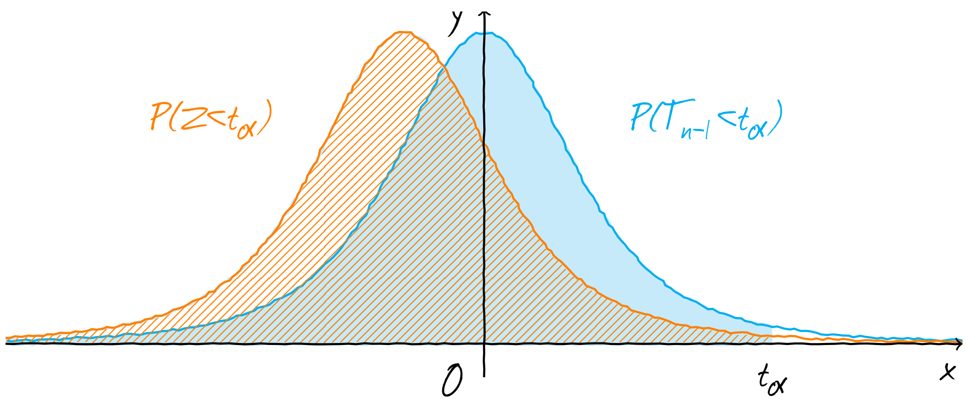
Moreover, as $\mathbb{E}(X_1) - \mu_0 \leqslant 0$, if $t_\alpha \in \mathbb{R}$ is such that

$$\mathbb{P}\left(\mathcal{T}_{n-1} \leqslant t_\alpha\right) = 1 - \alpha$$

then

$$\mathbb{P}\left(Z \leqslant t_\alpha\right) \geqslant 1 - \alpha.$$

Illustration 8.4. The probability density function of Z (in orange) is shifted to the left compared to the probability density function of the Student distribution (in blue).



Hence, **if $\mathcal{H}_0$ is true**, then:

$$\mathbb{P}\left(\overline{X} \leqslant \mu_0 + t_\alpha \frac{S_n}{\sqrt{n}}\right) \geqslant 1 - \alpha,$$

which yields the corresponding decision rule for an observed sample $(x_1, \ldots, x_n)$:

- if $\overline{x} \leqslant \mu_0 + t_\alpha s_n/\sqrt{n}$ then don't reject $\mathcal{H}_0$ (with an unknown risk),
- otherwise, reject $\mathcal{H}_0$ (with a risk α).

Example 8.5. Let's go back to example 8.3 and suppose now that we want to test with a risk of 1%:

$$\mathcal{H}_0 \ : \ \mu \leqslant 249 \qquad \text{against} \qquad \mathcal{H}_1 \ : \ \mu > 249.$$

Looking at the Student's table, we know that:

$$\mathbb{P}\left(\mathcal{T}_{69} \geqslant 2.381\right) = 99\%$$

Hence, **if $\mathcal{H}_0$ is true**, then:

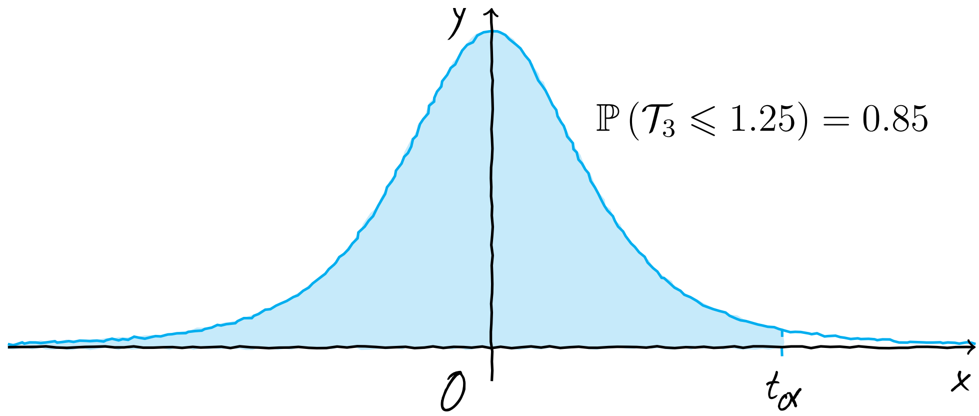
$$\mathbb{P}\left(\overline{X} \leqslant 249 + 2.381\frac{S_n}{\sqrt{70}}\right) \geqslant 99\%,$$

Here, the observed data is $\overline{x} = 250.2$ and $s_n = 1.25$ so:

$$\overline{x} \nleqslant 249 + 2.381 \times 1.25/\sqrt{70} = 249.35$$

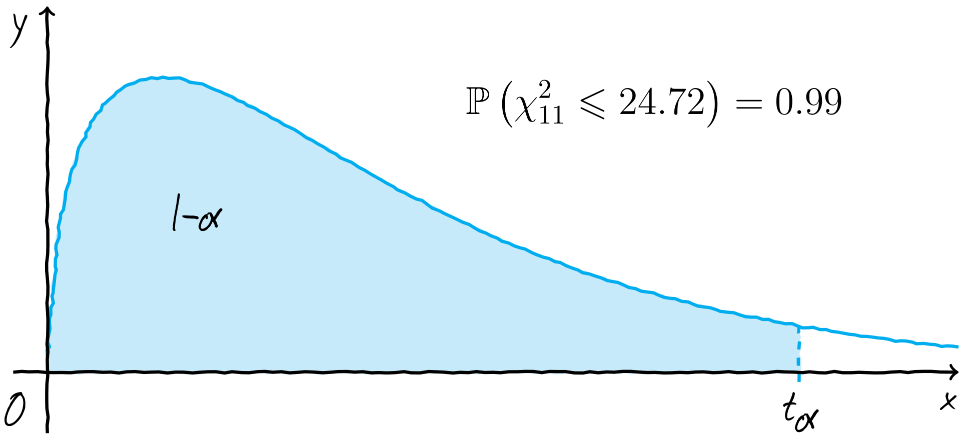
Thus, we can reject the hypothesis $\mathcal{H}_0$ with a risk of 1%.

Student table



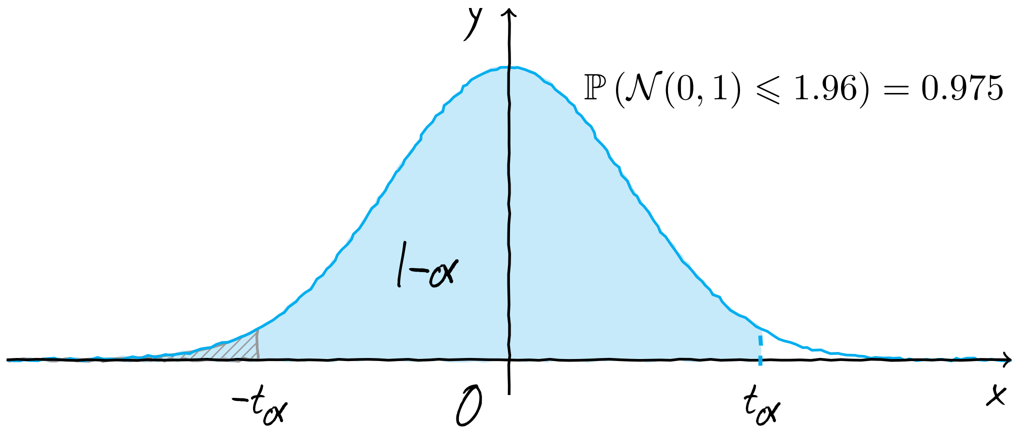
k	0.6	0.7	0.8	0.85	0.9	0.95	0.975	0.99	0.995	0.999
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706	31.821	63.657	318.309
2	0.289	0.617	1.061	1.386	1.886	2.92	4.303	6.965	9.925	22.327
3	0.277	0.584	0.978	1.25	1.638	2.353	3.182	4.541	5.841	10.215
4	0.271	0.569	0.941	1.19	1.533	2.132	2.776	3.747	4.604	7.173
5	0.267	0.559	0.92	1.156	1.476	2.015	2.571	3.365	4.032	5.893
6	0.265	0.553	0.906	1.134	1.44	1.943	2.447	3.143	3.707	5.208
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785
8	0.262	0.546	0.889	1.108	1.397	1.86	2.306	2.896	3.355	4.501
9	0.261	0.543	0.883	1.1	1.383	1.833	2.262	2.821	3.25	4.297
10	0.26	0.542	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144
11	0.26	0.54	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.93
13	0.259	0.538	0.87	1.079	1.35	1.771	2.16	2.65	3.012	3.852
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733
16	0.258	0.535	0.865	1.071	1.337	1.746	2.12	2.583	2.921	3.686
17	0.257	0.534	0.863	1.069	1.333	1.74	2.11	2.567	2.898	3.646
18	0.257	0.534	0.862	1.067	1.33	1.734	2.101	2.552	2.878	3.61
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579
20	0.257	0.533	0.86	1.064	1.325	1.725	2.086	2.528	2.845	3.552
21	0.257	0.532	0.859	1.063	1.323	1.721	2.08	2.518	2.831	3.527
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505
23	0.256	0.532	0.858	1.06	1.319	1.714	2.069	2.5	2.807	3.485
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467
25	0.256	0.531	0.856	1.058	1.316	1.708	2.06	2.485	2.787	3.45
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421
28	0.256	0.53	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408
29	0.256	0.53	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396
30	0.256	0.53	0.854	1.055	1.31	1.697	2.042	2.457	2.75	3.385
40	0.255	0.529	0.851	1.05	1.303	1.684	2.021	2.423	2.704	3.307
50	0.255	0.528	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.261
60	0.254	0.527	0.848	1.045	1.296	1.671	2.0	2.39	2.66	3.232
70	0.254	0.527	0.847	1.044	1.294	1.667	1.994	2.381	2.648	3.211
80	0.254	0.526	0.846	1.043	1.292	1.664	1.99	2.374	2.639	3.195
90	0.254	0.526	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.183
100	0.254	0.526	0.845	1.042	1.29	1.66	1.984	2.364	2.626	3.174

Chi-squared table



k	0.005	0.01	0.025	0.05	0.1	0.9	0.95	0.975	0.99	0.995
1	0.0	0.0	0.0	0.0	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.1	0.21	4.61	5.99	7.38	9.21	10.6
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.3	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.2	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.7	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.6	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.4	5.23	6.3	18.55	21.03	23.34	26.22	28.3
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.6	5.23	6.26	7.26	8.55	22.31	25.0	27.49	30.58	32.8
16	5.14	5.81	6.91	7.96	9.31	23.54	26.3	28.85	32.0	34.27
17	5.7	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.2	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.0
21	8.03	8.9	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.4
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.8
23	9.26	10.2	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.4	13.85	15.66	33.2	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.2	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.6	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.5	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.4	79.08	83.3	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43	104.21
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
90	59.2	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.34	129.56	135.81	140.17

Standard normal table



$$\mathbb{P}(\mathcal{N}(0,1) \leqslant -t_\alpha) = \mathbb{P}(\mathcal{N}(0,1) \geqslant t_\alpha) = 1 - \mathbb{P}(\mathcal{N}(0,1) \leqslant t_\alpha)$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Credits

A large fraction of these lecture notes was adapted from Frédéric Holweck and André Turbergue's courses, which may (or may not) be found on their respective websites.

Great online video courses are available at the Khan Academy, where one can also find a lot of exercises (with detailed solutions).

One might also find Wikipedia to be helpful when looking for information *on a specific subject.*

Frédéric Holweck's website: `http://utbmfh.pagesperso-orange.fr/`

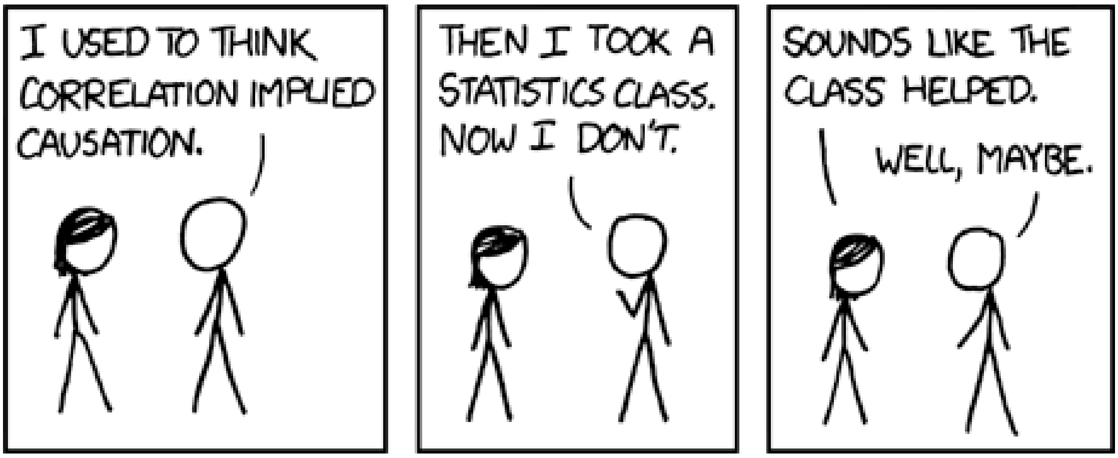
Khan Academy: `https://www.khanacademy.org/math/probability`

André Turbergue's website: `http://andre.turbergue.free.fr`

Virtual lab: `http://www.math.uah.edu/stat/`

Wikipedia: `http://en.wikipedia.org/wiki/Portal:Statistics`

Other resources:
`http://www.statisticshowto.com/`
`http://www.stat.ucla.edu/`
`http://stattrek.com/`



`http://www.xkcd.com/552/`