# Probabilité et Simulation

### PolyTech INFO4 (Grenoble) – 2024-2025 – Practical Sessions

Last updated: 2024-11-05. Info: davide.legacci@univ-grenoble-alpes.fr

## Contents

# 1. Theory recap 11.9.24 - Probability on finite set

- *Jet set* $\Omega =$ finite set of possible outcomes $\omega$
- *Probability* on $\Omega =$ set of weights $P(\omega) \in \mathbb{R}$ on each $\omega \in \Omega$ such that
  - $P(\omega) > 0 \forall \omega \in \Omega$
  - $\sum_{\omega \in \Omega} P(\omega) = 1$
- *Event* $A \subseteq \Omega =$ subset of the jet set
- *Complementary event* $A^c = \Omega / A$
- The cardinality of a set $S$ is denoted by $|S|$
- *Uniform probability of the event* $A$

$$P(A) = \sum_{\omega \in A} P(\omega) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|} \tag{1}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2}$$

- *Binomial theorem*

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} y^j \tag{3}$$

## 1.1. Counting

- Number of *permutations* of $k$ elements:
  - Number of ways to *order* $k$ elements
  - **Only order matters**

$$P_k = k! \tag{4}$$

- Number of *dispositions* of $k$ elements out of $n$ $(k \leq n)$:
  - Number of ways to *choose and order* $k$ elements out of $n$
  - **Order and elements** matter
  - Number of injections $f : \{1, ..., k\} \to \{1, ..., n\}$

$$D_{n,k} = \underbrace{n(n-1)...}_{k \text{ times}} = n(n-1)...(n-k+1) = \frac{n!}{(n-k)!} \tag{5}$$

- Number of *combinations* of $k$ elements out of $n$ $(k \leq n)$:
  - Number of ways to *choose* $k$ elements out of $n$
  - **Only elements** matter
  - Number of subsets of cardinality $k$ of a set of cardinality $n$
  - Number of dispositions modulo number of permutations

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n!}{k!(n-k)!} = \binom{n}{k} = \text{choose}(n,k) \tag{6}$$

## Exercises

**Exercise 1.1** *(Handshakes and kisses)*

There are $f$ girls and $g$ boys in a room. Boys exchange handshakes, girs exchange kisses, boys and girls exchange kisses. How many kisses in total?

The number of kisses exchanegs among girls is the number of subsets of cardinality 2 of a set of cardinality $f$, that is $\binom{f}{2} = \frac{f(f-1)}{2}$. Or, think that each girl gives $f-1$ kisses, and one needs a factor of one half to avoid double counting.

For the number of kisses exchanged between boys and girls: the first girl gives $g$ kisses, the second girl gives $g$ kisses, and so on, so we have $fg$ in total.

$$\text{number of kisses} = \frac{f(f-1)}{2} + fg \tag{7}$$

**Exercise 1.2** *(Throwing a dice)* Throw a fair dice with $f$ faces $n$ times. What's the prob to never get the same result twice?

Let $\mathcal{N} = \{1, ..., n\}$ and $\mathcal{F} = \{1, ..., f\}$. The jet set is

$$\Omega = \{\omega = (\omega_1, ..., \omega_n) : \omega_i \in \mathcal{F} \text{ for all } i \in \mathcal{N}\} = \mathcal{F}^n \tag{8}$$

with cardinality

$$|\Omega| = |\mathcal{F}^n| = |\mathcal{F}|^n = f^n \tag{9}$$

The event we're looking at is

$$A = \left\{\omega \in \Omega : \omega_i \neq \omega_j \text{ for all } i \neq j \in \mathcal{N}\right\} \tag{10}$$

Clearly if $n > f$ then $P(A) = 0$. Let $n \leq f$. The (uniform) probability of the event $A$ is $P(A) = \frac{|A|}{|\Omega|}$, with

$$\begin{aligned} |A| &= \# \text{ of ways to choose and order } n \text{ elements out of } f \\ &= \underbrace{f(f-1)...}_{n} = f(f-1)...(f-n+1) = \frac{f!}{(f-n)!} \end{aligned} \tag{11}$$

$$P(A) = \frac{f!}{f^n(f-n)!} \tag{12}$$

**Exercise 1.3** *(Birthday paradox)* What is the probability that at least 2 people out of $n$ have the same birthday? (Assume: uniform birth probability and year with $y$ number of days).

**Quick solution**

$$\begin{aligned} P(A) &= 1 - P\left(\underbrace{\text{no two people have the same birthday}}_{\text{Ex. 2}}\right) \\ &= 1 - \frac{y!}{y^n(y-n)!} \end{aligned} \tag{13}$$

**Formal solution** Let $\mathcal{N} = \{1, ..., n\}$ and $\mathcal{Y} = \{1, ..., y\}$ with $n \leq y$. The jet set is

$$\begin{aligned}\Omega &= \text{distributions of possible birthdays of } n \text{ people} \\ &= \{\omega = (\omega_1, \ldots, \omega_n) : \omega_i \in \mathcal{Y} \text{ for all } i \in \mathcal{N}\} = \mathcal{Y}^n\end{aligned} \tag{14}$$

where $\omega_i$ is the birthday of the $i$-th person. The cardinality of the jet set is

$$|\Omega| = |\mathcal{Y}^n| = |\mathcal{Y}|^n = y^n \tag{15}$$

The event we're looking at is

$$A = \{\omega \in \Omega : \exists i \neq j \in \mathcal{N} : \omega_i = \omega_j\} \tag{16}$$

Note that this is the complementary event to the event defined in Equation 10 of Exercise 2. Thus we can compute its probability as

$$P(A) = 1 - P(A^c) \tag{17}$$

in agreement with Equation 13.

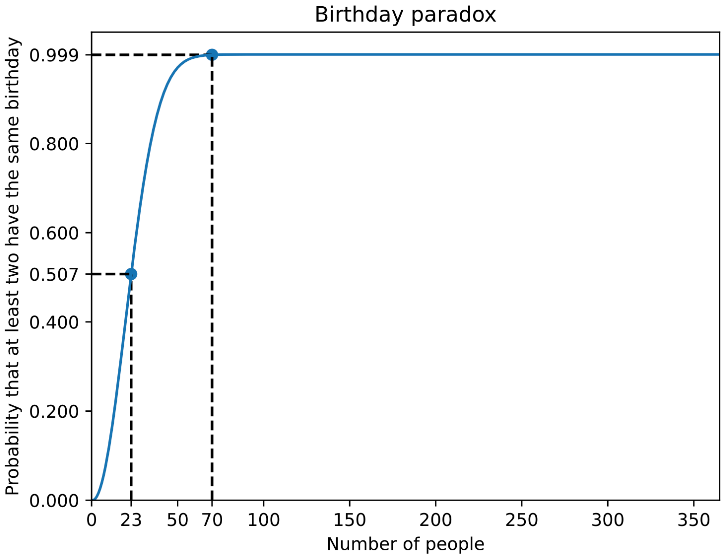


Figure 1: Birthday paradox probability. Code available.

**Exercise 1.4** *(Same birthday as the prof)* What is the probability that at least 1 student out of $n$ has the same birthday of the prof? (Assume: uniform birth probability and year with $y$ number of days).

**Quick solution**

$$\begin{aligned}P(A) &= 1 - P\left(\underbrace{\text{nobody has the prescribed birth date}}\right) \\ &= 1 - \left(\frac{y-1}{y}\right)^n\end{aligned} \tag{18}$$

**Formal solution 1** As above $\mathcal{N} = \{1, \ldots, n\}$ and $\mathcal{Y} = \{1, \ldots, y\}$ with $n \leq y$. The jet set is $\Omega = \mathcal{Y}^{n+1}$, that is the set of possible birthdays of $n+1$ people, the $(n+1)$-th being the prof. Its cardinality is $|\Omega| = y^{n+1}$. The event we're looking at is

$$A = \{\omega \in \Omega : \exists i \in \mathcal{N} : \omega_i = \omega_{n+1}\} \tag{19}$$

with complementary event

$$A^c = \{\omega \in \Omega : \omega_i \neq \omega_{n+1} \forall i \in \mathcal{N}\} \tag{20}$$

As usual $P(A) = 1 - P(A^c) = 1 - \frac{|A^c|}{|\Omega|}$, with

$$|A^c| = \underbrace{y}_{\text{prof}} \cdot \underbrace{(y-1)^n}_{\text{students}} \tag{21}$$

So, $P(A) = 1 - \frac{y(y-1)^n}{y^{n+1}} = 1 - \left(\frac{y-1}{y}\right)^n$, in agreement with Equation 18.

**Formal solution 2** Using the probability of the complementary event is often the smartest way to proceed, but for the sake of completeness let's see how to get the same result directly. Let $A_j$ be the event "exactly $j$ students out of $n$ have the same birthday as the prof". The event we look at then is

$$A = \sqcup_{j \in \mathcal{N}} A_j \tag{22}$$

with probability (cf Equation 2)

$$P(A) = \sum_{j \in \mathcal{N}} P(A_j) = \frac{\sum_{j \in \mathcal{N}} |A_j|}{|\Omega|} \tag{23}$$

The cardinality of $A_j$ is

$$\begin{aligned} |A_j| &= \underbrace{1...1}_{j \text{ times}} \cdot \underbrace{(y-1)...(y-1)}_{n-j \text{ times}} \cdot \underbrace{y}_{\text{prof}} \cdot \underbrace{\binom{n}{j}}_{\text{number of ways to choose } j \text{ elements out of } n} \\ &= y(y-1)^{n-j}\binom{n}{j} \end{aligned} \tag{24}$$

By an application of the binomial theorem (Equation 3) and a short manipulation,

$$\sum_{j=1}^{n} |A_j| = y(y^n - (y-1)^n) \tag{25}$$

which leads back to Equation 18.

# 2. Theory recap 18.9.24 - Conditional probability and independence

## 2.1. Conditional probability

- *Conditional probability*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \text{ if } P(B) \neq 0 \tag{26}$$

- not reallty defined if $P(B) = 0$, cf [1] pag. 427.
- often used as

$$P(A \cap B) = P(A \mid B)P(B) \tag{27}$$

- Conditional probability and complementary event (proof: simple exercise.)

$$P(A \mid B) + P(A^c \mid B) = 1 \tag{28}$$

- *Total probability theorem*

$$P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c) \tag{29}$$

- *Bayes theorem*

$$P(A \cap B) = P(B \cap A) \Rightarrow P(A \mid B)P(B) = P(B \mid A)P(A) \tag{30}$$

See this notebook for an example of Bayes theorem in action.

### 2.2. Independent events

Let $\Omega$ be equipped with a probability $P$.

- two events $A, B \subseteq \Omega$ are said *independent* if

$$P(A \cap B) = P(A)P(B) \tag{31}$$

- $n$ events $A_1, \ldots, A_n$ are said *independent* if

$$P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i) \text{ for all } I \subseteq \{1, \ldots, n\} \tag{32}$$

- pairwise independence does not imply independence of all events!

## Exercises

**Exercise 2.1** *(Pile ou Face)* Jet de 2 pieces, $\Omega = \{\mathrm{PP}, \mathrm{PF}, \mathrm{FP}, \mathrm{FF}\}$. Cet espace est muni de la probabilite uniforme. Soient les evenements:

- $A = \text{1ere piece donne P}$
- $B = \text{2eme piece donne F}$
- $C = \text{le deux pieces donnent le meme resultat}$

Questions:

- $A$ et $B$ sont indépendantes?
- $A, B$ et $C$ sont indépendants?

$$\begin{aligned}
A &= \{PP, PF\} & \mathbb{P}(A) &= 1/2 \\
B &= \{PF, FF\} & \mathbb{P}(B) &= 1/2 \\
C &= \{PP, FF\} & \mathbb{P}(C) &= 1/2 \\
A \cap B &= \{PF\} & \mathbb{P}(A \cap B) &= 1/4 & &= \mathbb{P}(A)\mathbb{P}(B) \\
A \cap C &= \{PP\} & \mathbb{P}(A \cap C) &= 1/4 & &= \mathbb{P}(A)\mathbb{P}(C) \\
B \cap C &= \{FF\} & \mathbb{P}(B \cap C) &= 1/4 & &= \mathbb{P}(B)\mathbb{P}(C) \\
A \cap B \cap C &= \emptyset & \mathbb{P}(A \cap B \cap C) &= 0 & &\neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).
\end{aligned}$$

Ainsi les événements $A, B$ et $C$ sont 2 à 2 indépendants mais pas indépendants.

Figure 2: Pairwise independence does not imply independence of all events!

**Exercise 2.2** *(Pieces mecaniques defectueuses)* Parmi 10 pièces mécaniques, 4 sont déefectueuses. On prend successivement deux pièces au hasard dans le lot (sans remise). Quelle est la probabilité pour que les deux pièces soient correctes?

**Solution 1** Let $A_i$ be the event *the i-th drawn piece is good*, with $i \in \{1, 2\}$. We need the probability of the event $A_2 \cap A_1$. By definition of conditional probavility,

$$P(A_2 \cap A_1) = \underbrace{P(A_2 \mid A_1)}_{\frac{5}{9}} \underbrace{P(A_1)}_{\frac{6}{10}} = \frac{1}{3}. \tag{33}$$

**Solution 2** The jet set is the set of subsets of cardinality 2 of a set of cardinality 10, so $|\Omega| = \binom{10}{2}$. The event we consider is the set of subsets of cardinality 2 of a set of cardinality 6, so $|A| = \binom{6}{2}$. Then

$$P(A) = \frac{\binom{6}{2}}{\binom{10}{2}} = \frac{6 \cdot 5}{10 \cdot 9} = \frac{1}{3}. \tag{34}$$

**Exercise 2.3** *(Betting on cards)* We have three cards:

- a *red* card with both faces red;
- a *white* card with both faces white;
- a *mixed* card with a red face and a white face.

One of the three cards is drawn at random and one of the faces of this card (also chosen at random) is exposed. This face is red. You are asked to bet on the color of the hidden face. Do you choose red or white?

**Intuitive solution** The cards are $RR, RW, WW$ with $W$ for white and $R$ for red. Call $RR$ the "red" card, $WW$ the "white" card, and $WR$ the "mixed" card. Since we observe a red face, the white card cannot be on the table. There are three possibilities left: 1. we're observing a face of the red card (in which case the hidden face is red); 2. we are observing the other face of the red card (in which case the hidden face is red); 3. we are observing the red face of the mixed card (in which case the hidden face is white). So the hidden face is red 2 out of 3 times.

**Formal solution** The jet set contains the possible outcomes of a sequence of two events: 1. draw a card (out of three), and 2. observe a face (out of two). Denote by $R$ a red face and by $W$ a white face, and denote by a subscript $o$ the observed face, and by a subscript $h$ the hidden face. The possible outcomes then are

$$\Omega = \{R_h \cap R_o, R_h \cap W_o, W_h \cap R_o, W_h \cap W_o\} \tag{35}$$

where the first entry indicates the hidden face, and the second entry indicates the observed face. For example, $W_h \cap R_o$ is the event "*the hidden face is white and the observed face is red*", and similarly for the others.

In this formulation, every element in the jet set is the intersection of two (dependent) events of the type 1. a face is hidden, and 2. a face is observed. Note that the event $W_h \cap R_o$ is equivalent to the event "*the mixed card is drawn, and the red face is observed.*" Under this second point of view, each outcome in $\Omega$ is the intersection of two (dependent) events of the type 1. a card is drawn, and 2. a face is observed. Denoting the event *"draw the red card"* by $r$, the event *"draw the white card"* by $w$, and the event *"draw the mixed card"* by $m$, the jet set is equivalently

$$\Omega = \{r \cap R_o, m \cap W_o, m \cap R_o, w \cap W_o\} \tag{36}$$

This formulation helps to understand that the probability on $\Omega$ **is not uniform**. The probabilities of the events in $\Omega$ are computed by Equation 27:

$$P(R_h \cap R_o) = P(r \cap R_o) = \frac{P(r \mid R_o)}{R_o} \tag{37}$$

However, we do not know the probabilities on the right hand side. As a simple trick, remember that $P(A \cap B) = P(B \cap A)$, so we can turn this around:

$$\begin{aligned} P(R_h \cap R_o) &= P(R_o \cap r) \\ &= \underbrace{P(R_o \mid r)}_{1} \underbrace{P(r)}_{\frac{1}{3}} = \frac{2}{6} \end{aligned} \tag{38}$$

$$P(R_h \cap W_o) = P(W_o \cap m)$$
$$= \underbrace{P(W_o \mid m)}_{\frac{1}{2}} \underbrace{P(m)}_{\frac{1}{3}} = \frac{1}{6} \tag{39}$$

$$P(W_h \cap R_o) = P(R_o \cap m)$$
$$= \underbrace{P(R_o \mid m)}_{\frac{1}{2}} \underbrace{P(m)}_{\frac{1}{3}} = \frac{1}{6} \tag{40}$$

$$P(W_h \cap W_o) = P(W_o \cap w)$$
$$= \underbrace{P(W_o \mid w)}_{1} \underbrace{P(w)}_{\frac{1}{3}} = \frac{2}{6} \tag{41}$$

Now by Equation 26 and using these probabilities,

$$\begin{aligned} P(W_h \mid R_o) &= \frac{P(W_h \cap R_o)}{P(R_o)} \\ &= \frac{P(W_h \cap R_o)}{P(R_h \cap R_o) + P(W_h \cap R_o)} = \frac{1}{3} \end{aligned} \tag{42}$$

$$\begin{aligned} P(R_h \mid R_o) &= \frac{P(R_h \cap R_o)}{P(R_o)} \\ &= \frac{P(R_h \cap R_o)}{P(R_h \cap R_o) + P(W_h \cap R_o)} = \frac{2}{3} \\ &= 1 - P(W_h \mid R_o) \end{aligned} \tag{43}$$

where the last line follows from Equation 28 and gives directly the answer. So in conclusion, given the fact that we observe a red face, the hidden face is also red with probability $2/3$.

**Exercise 2.4** *(Russian roulette)* You are playing two-person Russian roulette with a revolver featuring a rotating cylinder with six bullet slots. Each time the gun is triggered, the cylinder rotates by one slot. Two bullets are inserted one next to the other into the cylinder, which is then randomly positioned. Your opponent is the first to place the revolver against her temple. She presses the trigger and... she stays alive. With great display of honor, she offers you to rotate the barrel again at random before firing in turn. What do you decide?

The bullets occupy the positions $x$ and $x + 1 \bmod 6$:

$$\Omega = \{12, 23, 34, 45, 56, 61\} \tag{44}$$

Say the revolver shots from position 1. The event "*the first player dies*" is

$$\text{die}_1 = \{12, 61\} \tag{45}$$

so $P(\text{die}_1) = \frac{1}{3}$ and $P(\text{live}_1) = \frac{2}{3}$. We need to compute

$$P(\text{die}_2 \mid \text{live}_1) = \frac{P(\text{die}_2 \cap \text{live}_1)}{P(\text{live}_1)} \tag{46}$$

Since the cylynder rotates after being triggered we have $\text{die}_2 = \{56, 61\}$ and $\text{die}_2 \cap \text{live}_1 = \{56\}$, so $P(\text{die}_2 \mid \text{live}_1) = \frac{1}{6} / \frac{2}{3} = \frac{1}{4} < P(\text{die}_1)$. So you don't shuffle the barrel.

# 3. Theory recap 26.9.24 - Probability space, dicrete random variables, and distributions

## 3.1. Probability measure

The relevant references are [2] pag. 11 and [1], pag. 22 and 160.

**Definition 3.1.1** *(sigma-field)* Let $\Omega$ be any set. A *$\sigma$-field* $\mathcal{A}$ on $\Omega$ is a collection of subsets of $\Omega$ that[1]

1. is closed under complement: if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$;
2. contains the whole set: $\Omega \in \mathcal{A}$;
3. is closed under countable union: if $A_1, A_2, \ldots$ is a countable family of sets of $\mathcal{A}$ then their union $\cup_{i=1}^{\infty} A_i$ is also in $\mathcal{A}$.

A subset of $\Omega$ that is in $\mathcal{A}$ is called *event*.

**Definition 3.1.2** *(Measure)* Given a set $\Omega$ and a $\sigma$-algebra $\mathcal{A}$ on $\Omega$, a *measure* $\mu$ is a function

$$\mu : \mathcal{A} \to \mathbb{R}_{\geq 0} \tag{47}$$

such that

1. $\mu(\emptyset) = 0$
2. *countable additivity* (also called *$\sigma$-additivity*) is fulfilled, namely the measure of a *disjoint* countable union of sets in $\mathcal{A}$ is the sum of their measures:

$$\mu\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{48}$$

**Definition 3.1.3** *(Probability measure)* Given a set $\Omega$ and a $\sigma$-algebra $\mathcal{A}$ on $\Omega$, a *probaility measure* $P$ is a measure (in the sense above) with the additional requirement that

$$P(\Omega) = 1. \tag{49}$$

- Note that this implies that $P(A) \leq 1$ for all events $A \in \mathcal{A}$.
- A triple $(\Omega, \mathcal{A}, P)$ where $\mathcal{A}$ is a $\sigma$-algebra on $\Omega$ and $P$ is a probability measure is called *probability space*.

> **Tip** Putting all together, a probability measure $P : \mathcal{A} \to [0, 1]$ on a space $\Omega$ is a function from a "well-behaved" collection of subsets of $\Omega$ (the $\sigma$-field) to $[0, 1]$, such that $P(\emptyset) = 0$, $P(\Omega) = 1$, and fulfilling countable additivity.

## 3.2. Discrete random variables

**Definition 3.2.1** *(Discrete random variable)* Given a probability space $(\Omega, \mathcal{A}, P)$, a *discrete random variable $X$* is a function $X : \Omega \to F$ such that

1. $F$ is a *countable* set;
2. *the level sets of $X$ are events*, that is

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{A} \text{ for all } x \in F \tag{50}$$

[1] In french, this set is called *tribu* on $\Omega$. The term *$\sigma$-algebra* is also used – and is more common in the context of pure analysis, c.f. [3] – whereas the term *$\sigma$-field* is more common in the context of probabioloty theory, c.f. [1].

- clearly, $\{X = x\} = \emptyset \in \mathcal{A}$ for all $x \in F \setminus \mathrm{Im}(X)$
- the second property guarantees that $P\{X = x\}$ is well-defined for all $x \in F$, which allows for the following definition:

**Definition 3.2.2** *(Distribution of a discrete random variable)* The *distribution* (or *law*) of a random variable $X$ is the function $\mu : F \to [0, 1]$ defined by

$$\mu(x) = P\{X = x\} \text{ for all } x \in F. \tag{51}$$

- two discrete random variables $X$ and $Y$ taking values resp. in $F$ and $G$ are *independent* if

$$P\{X = x, Y = y\} = P\{X = x\}P\{Y = y\} \text{ for all } x \in F, y \in G \tag{52}$$

- it is understood that $\{X = x, Y = y\}$ is a shorthand for the event

$$\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\} \in \mathcal{A}. \tag{53}$$

- the definition generalises to collections of DRVs, see Section 2.2.3 in [2].

> **Tip** A discrete random variable is a *function on* $\Omega$ with *countable range*. Think of it as an experiment with a random outcome. Its *law, or distribution*, gives the probability to observe each of the possible (countable) *values* of the random variable.

> **Tip**
> - $(\Omega, \mathcal{A}, P)$ with $P : \mathcal{A} \to [0, 1]$ and $P(\Omega) = 1$
> - $X : \Omega \to F$ countable, with $\{X = x\} \in \mathcal{A}$ for all $x \in F$
> - $\mu : F \to [0, 1]$ such that $\mu(x) = P\{X = x\}$

## 3.3. Standard discrete distributions

### 3.3.1. Bernoulli $\mathcal{B}(p)$

- The Bernoulli distribution models a random experiment which has two possible outcomes.
- More precisely, the Bernoulli distribution is the distribution of a discrete random variable $X$ that can assume only values in $F = \{0, 1\}$.
- The distribution is parametrized by $p \in [0, 1]$, and we write $X \sim \mathcal{B}(p)$.

$$\begin{aligned} \mu : F &\to [0, 1] \\ 1 &\longmapsto p \\ 0 &\longmapsto 1 - p \\ x &\longmapsto p^x (1 - p)^{1-x} \end{aligned} \tag{54}$$

### 3.3.2. Binomial $\mathcal{B}(n, p)$

- Distribution of the discrete random variable $X = X_1 + ... + X_n$, where the $X_i$-s are independent Bernoulli variables of parameter $p \in [0, 1]$.
- $F = \{0, ..., n\}$; $k \in F$ is value of the sum

$$\begin{aligned} \mu : F &\to [0, 1] \\ k &\longmapsto \binom{n}{k} p^k (1 - p)^{n-k} \end{aligned} \tag{55}$$

### 3.3.3. Poisson $\mathcal{P}(\lambda)$

- probability of observing a given number of independent events occurring at constant rate $\lambda > 0$
- $F = \mathbb{N}$; $n \in F$ is number of observed events

$$\begin{aligned} \mu : F &\to [0,1] \\ n &\longmapsto e^{-\lambda}\frac{\lambda^n}{n!} \end{aligned} \tag{56}$$

#### 3.3.4. Geometric $\mathcal{G}(p)$

- First successful event from a sequence of independent $p$-Bernoulli events.
- $F = \mathbb{N}^*$; $k \in F$ is first succesful event

$$\begin{aligned} \mu : F &\to [0,1] \\ k &\longmapsto p(1-p)^{k-1} \end{aligned} \tag{57}$$

### 3.4. Useful stuff

- Vandermonde's identity

$$\sum_{k_1=0}^{k} \binom{n_1}{k_1}\binom{n_2}{k-k_1} = \binom{n_1+n_2}{k} \tag{58}$$

## Exercises

**Exercise 3.1** *(Sum of independent binomial distributions)* Let $X_i \sim \mathcal{B}(n_i, p)$ with $i \in \{1,2\}$ be independent discrete random variables following the Bernoulli law. Find the law of $X_1 + X_2$.

*Hint: c.f. derivation of binomial distribution [2] pag. 16.*

The laws $\mu_i : F_i = \{0, ..., n_i\} \to [0,1]$ of the two variables are given by

$$\mu_i(k_i) = P(X_i = k_i) = \binom{n_i}{k_i} p^{k_i}(1-p)^{n_i-k_i} \tag{59}$$

The law of $X_1 + X_2$ takes value in $F = \{0, ..., n_1 + n_2\}$ and for all $k \in F$ is given by

$$\begin{aligned} \mu(k) &= P(X_1 + X_2 = k) \\ &= P\left( \bigsqcup_{\substack{k_i \in F_i \\ k_1+k_2=k}} \{X_1 = k_1, X_2 = k_2\} \right) \\ &= \sum_{\substack{k_i \in F_i \\ k_1+k_2=k}} P(X_1 = k_1)P(X_2 = k_2) && \text{by c. add and indep.} \\ &= \sum_{\substack{k_i \in F_i \\ k_1+k_2=k}} \mu(k_1)\mu(k_2) && \text{by def of law} \\ &= \sum_{\substack{k_i \in F_i \\ k_1+k_2=k}} \binom{n_1}{k_1}\binom{n_2}{k_2} p^{k_1+k_2}(1-p)^{n_1+n_2-k_1-k_2} \\ &= p^k(1-p)^{n_1+n_2-k} \sum_{\substack{k_i \in F_i \\ k_1+k_2=k}} \binom{n_1}{k_1}\binom{n_2}{k_2} \end{aligned} \tag{60}$$

Let's focus on the sum. For each fixed $k_1 \in F_1$, $k_2$ is constrained to be $k - k_1$. Furthermore, in order for $k_2$ to be $\geq 0$, $k_1$ can be at most equal to $k$. So the constraints

$$
\begin{aligned}
&k_1 \in \{0, ..., n_1\} \\
&k_2 \in \{0, ..., n_2\} \\
&k_1 + k_2 = k
\end{aligned}
\tag{61}
$$

can be replaced by the constraints

$$
\begin{aligned}
&k_1 \in \{0, ..., k\} \\
&k_2 = k - k_1
\end{aligned}
\tag{62}
$$

namely

$$
\sum_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \binom{n_1}{k_1}\binom{n_2}{k_2} = \sum_{k_1=0}^{k} \binom{n_1}{k_1}\binom{n_2}{k - k_1} = \binom{n_1 + n_2}{k}
\tag{63}
$$

where the second step follows by by Vandermonde's identity.

**Remark 3.4.1** Note that it is correct to have $k_1$ running from 0 to $k$:
- If $k \leq n_1$, $k_1$ can be at most $k$ so that $k_2 = k - k_1 \geq 0$, so we have directly Equation 63.
- If $k > n_1$, we have

$$
\begin{aligned}
\sum_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \binom{n_1}{k_1}\binom{n_2}{k_2} &= \sum_{k_1=0}^{n_1} \binom{n_1}{k_1}\binom{n_2}{k - k_1} \\
&= \sum_{k_1=0}^{n_1} \binom{n_1}{k_1}\binom{n_2}{k - k_1} + \sum_{k_1=n_1+1}^{k} \binom{n_1}{k_1}\binom{n_2}{k - k_1}
\end{aligned}
\tag{64}
$$

since each summand in the the second sum is zero[2], and we get again Equation 63.

So in conclusion

$$
\mu(k) = \binom{n_1 + n_2}{k} p^k (1 - p)^{n_1 + n_2 - k}
\tag{65}
$$

namely $X_1 + X_2 \sim \mathcal{B}(n_1 + n_2, p)$.

**Exercise 3.2** *(Sum of independent Poisson distributions)* Let $X_i \sim \mathcal{P}(\lambda_i)$ with $i \in \{1, 2\}$ be independent discrete random variables following the Poisson law. Find the law of $X_1 + X_2$.

*Hint: c.f. previous exercise and binomial theorem.*

Analogously to before, with $i \in \{1, 2\}$, we look for the law $\mu : \mathbb{N} \to [0, 1]$ given by

$$
\begin{aligned}
\mu(n) = P(X_1 + X_2 = n) &= \sum_{\substack{n_i \in \mathbb{N} \\ n_1 + n_2 = n}} \mu_1(n_1)\mu_2(n_2) \\
&= \sum_{\substack{n_i \in \mathbb{N} \\ n_1 + n_2 = n}} e^{-\lambda_1} \frac{\lambda_1^{n_1}}{n_1!} e^{-\lambda_2} \frac{\lambda_2^{n_2}}{n_2!}
\end{aligned}
\tag{66}
$$

As before we replace the consraint by $n_1 \in \{0, ..., n\}$ and $n_2 = n - n_1$, so

[2]Recall that $\binom{a}{b} = 0$ if $b > a$.

$$\begin{aligned} \mu(n) &= e^{-(\lambda_1+\lambda_2)} \sum_{n_1=0}^{n} \frac{\lambda_1^{n_1}}{n_1!} \frac{\lambda_2^{n-n_1}}{n-n_1!} \cdot \frac{n!}{n!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{n_1=0}^{n} \binom{n}{n_1} \lambda_1^{n_1} \lambda_2^{n-n_1} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!} \end{aligned} \tag{67}$$

So $X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

**Exercise 3.3** *(Min of independent geometric distributions)* Let $X_i \sim \mathcal{G}(p_i)$ with $i \in \{1, 2\}$ be independent DRVs following the geometric law. Find the law of $\min\{X_1 + X_2\}$.

*Hint: set up the problem in terms of inequalities.*

Let $Z = \min\{X_1 + X_2\}$. We look for the law $\mu : \mathbb{N}^* \to [0, 1]$ such that

$$\begin{aligned} \mu(k) &= P(Z = k) \\ &= P(Z \geq k) - P(Z \geq k+1) \\ &= P(X_1 \geq k, X_2 \geq k) - P(X_1 \geq k+1, X_2 \geq k+1) \\ &= P(X_1 \geq k)P(X_2 \geq k) - P(X_1 \geq k+1)P(X_2 \geq k+1) \end{aligned} \tag{68}$$

Let's drop the subscript for a moment. For a DRV $X \sim \mathcal{G}(p)$ and for $k \in \mathbb{N}^*$ we need

$$\begin{aligned} P(X \geq k) &= P\left(\bigsqcup_{x \geq k} (X = i)\right) \\ &= \sum_{i \geq k} P(X = i) \\ &= \sum_{i \geq k} p(1-p)^{i-1} \\ &= p(1-p)^{k-1} + p(1-p)^k + p(1-p)^{k+1} + \ldots \\ &= p(1-p)^{k-1}\left(1 + (1-p) + (1-p)^2 + \ldots\right) \\ &= p(1-p)^{k-1} \sum_{j=0}^{\infty} (1-p)^j \\ &= (1-p)^{k-1} \end{aligned} \tag{69}$$

Plugging in Equation 68 we get

$$\begin{aligned} \mu(k) &= P(X_1 \geq k)P(X_2 \geq k) - P(X_1 \geq k+1)P(X_2 \geq k+1) \\ &= (1-p_1)^{k-1}(1-p_2)^{k-1} - (1-p_1)^k(1-p_2)^k \\ &= (1-p_1)^{k-1}(1-p_2)^{k-1}[1 - (1-p_1)(1-p_2)] \end{aligned} \tag{70}$$

Let $\alpha = (1-p_1)(1-p_2)$ and $\beta = 1 - \alpha = p_1 + p_2 - p_1 p_2$, so $\mu(k) = \beta(1-\beta)^{k-1}$, i.e. $Z \sim \mathcal{G}(\beta)$.

# 4. Theory recap 3.10.24 - Expectation and variance

Recall from last week that

> **Tip**
> - $(\Omega, \mathcal{A}, P)$ with $P : \mathcal{A} \to [0,1]$ and $P(\Omega) = 1$
> - $X : \Omega \to F$ countable, with $\{X = x\} \in \mathcal{A}$ for all $x \in F$
> - $\mu : F \to [0,1]$ such that $\mu(x) = P\{X = x\}$

## 4.1. Expected value of a discrete random variable

In this section when we say *"$X$ is a RV"* we mean "*$X : \Omega \to F \subset \mathbb{R}$ is a discrete random variable with real values.*"

**Definition 4.1.1** *(Expected value)* A RV is *integrable* if $\sum_{x\in F} |x| P(X = x) < +\infty$, and in this case its *expected value* $\mathbb{E}(X)$ is the real number

$$\mathbb{E}(X) := \sum_{x\in F} x P(X = x). \tag{71}$$

**Proposition 4.1.2** *(Linearity of expectation)*

$$\mathbb{E}(X + aY) = \mathbb{E}(X) + a\mathbb{E}(Y) \tag{72}$$

Some **properties** of the expected value of a RV:

- $\mathbb{E}(\text{constant}) = \text{constant}$
- Sufficient condition, positivity, monotonicity: see [2] pag. 20.

Some **examples**:
- $X \sim \mathcal{B}(p) \Rightarrow \mathbb{E}(X) = p$
- $X \sim \mathcal{B}(n,p) \Rightarrow \mathbb{E}(X) = np$ (immediate by linearity from the above)
- $X \sim \mathcal{P}(\lambda) \Rightarrow \mathbb{E}(X) = \lambda$
- $X \sim \mathcal{G}(p) \Rightarrow \mathbb{E}(X) = \frac{1}{p}$

**Theorem 4.1.3** *(Expectation of function)* Let $X : \Omega \to F \subset \mathbb{R}$ be RV, and consider some function $f : F \to \mathbb{R}$. Then

$$\mathbb{E}(f(X)) = \sum_{x\in F} f(x) P(X = x) \tag{73}$$

whenever defined (see [2] Th. 2.3.6 for details).

**Proposition 4.1.4** *(Expectation and independence)* Let $X, Y$ be RV and $f, g$ two functions on their values such that all the expectations are well-defined (i.e. al the random variables are integrable). Then

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)) \tag{74}$$

## 4.2. Variance of a discrete random variable

**Definition 4.2.1** *(Variance)* A RV $X$ is called *square integrable* if $X^2$ is integrable, that is if $\sum_{x\in F} x^2 P(X = x) < +\infty$, and in this case

$$\operatorname{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \tag{75}$$

- If $X$ is square integrable then the variance is well defined, cf [2] Remark 2.3.11

- The variance is a measure of the spreading, dispersion, of a random variable around its expected value

The two following properties of the variance are very useful for concrete calculations:

**Lemma 4.2.2** *(Variance as difference of expectations)*

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \tag{76}$$

The variance is in general **not linear**:

**Lemma 4.2.3** *(Variance after scaling and shifting)*

$$\mathrm{Var}(aX + b) = a^2 \ \mathrm{Var}(X) \text{ for all } a, b \in \mathbb{R} \tag{77}$$

**Proof** Exercise; both lemmas follow by linearity of the expectation. □

**Proposition 4.2.4** *(Variance and independence)* Let $(X_i)_{i \in \{1,...,n\}}$ be a family of square integrable random variables. Then their sum is square integrable, and **if the $X_i$ are independent** then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) \tag{78}$$

Some **examples**:
- $X \sim \mathcal{B}(p) \Rightarrow \mathrm{Var}(X) = p(1-p)$, see Exercise 4.3 and proof.
- $X \sim \mathcal{B}(n,p) \Rightarrow \mathrm{Var}(X) = np(1-p)$ (immediate by Proposition 4.2.4)
- $X \sim \mathcal{P}(\lambda) \Rightarrow \mathrm{Var}(X) = \lambda$, see proof.

## Exercises

**Exercise 4.1** *(Lost messages)* On a telecommunication channel, it has been estimated that in $T$ time units there arrives a number of messages that can be estimated by a DRV $\sim \mathcal{P}(\lambda T)$. Each message has a loss probability equal to $p$, independent of the other messages. Find the probability that the number of lost message in $T$ units of time is equal to $l$.

Without loss of generality rescale $\lambda \leftarrow \lambda T$. We need to find the discrete random variable $L$ whose range $\{0, 1, 2, ...\} \ni l$ contains the possible numbers $l$ of lost messages in one time unit. The probability $P(L = l)$ to lose $l$ message is then by definition by the law of $L$.

Let $X_i$ be the DRV for the event "*the $i$-th message is lost*". Since each message is lost with probability $p$, $X_i \sim \mathcal{B}(p)$ for all $i \in \{1, 2, ...\}$.

Let $L_a = \sum_{i=1}^{a} X_i$ be the DRV whose range $\mathrm{Im}(L_a) = \{0, 1, ..., a\} \ni l$ contains the numbers $l$ of possible lost messages out of $a$ arrived. Since $L_a$ is the sum of $a$ independent $p$-Bernoulli DRVs, $L_a$ follows the binomial distribution:

$$L_a \sim \mathcal{B}(a, p). \tag{79}$$

Finally, let $A$ be the DRV estimating the number of arrived messages $a \in \{0, 1, ..., \}$ in one time unit; we are given that $A \sim \mathcal{P}(\lambda)$.

The law of $L$ is given by

$$P(L = l) = P\left[\bigsqcup_{a=l}^{\infty} \{L_a = l \cap A = a\}\right], \tag{80}$$

that is, we look at the disjoint union of all the events in which, given $a$ arrived messages, $l$ are lost. By countable additivity and independence,

$$P(L = l) = \sum_{a=l}^{\infty} P(L_a = l)P(A = a) \tag{81}$$

Now $L_a$ follows a binomial distribution and $A$ follows a Poisson distribution, so

$$\begin{aligned} P(L = l) &= \sum_{a=l}^{\infty} \binom{a}{l} p^l (1-p)^{a-l} e^{-\lambda} \frac{\lambda^a}{a!} \cdot \frac{\lambda^l}{\lambda^l} \\ &= \lambda^l p^l \frac{1}{l!} e^{-\lambda} \sum_{a=l}^{\infty} \frac{1}{(a-l)!} (1-p)^{a-l} \lambda^{a-l} \\ &= \lambda^l p^l \frac{1}{l!} e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda - \lambda p)^j}{j!} \\ &= \lambda^l p^l \frac{1}{l!} e^{-\lambda} e^{\lambda - \lambda p} \\ &= \frac{(\lambda p)^l}{l!} e^{-\lambda p}. \end{aligned} \tag{82}$$

So $V \sim \mathcal{P}(\lambda p)$.

**Exercise 4.2** *(Poisson expectation)* Let $N \sim \mathcal{P}(\lambda)$. Find $\mathbb{E}\left(X := \frac{1}{N+1}\right)$.

By Theorem 4.1.3 we have

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \frac{1}{n+1} e^{-\lambda} \frac{\lambda^n}{n!} \tag{83}$$

Multiply and divide by $\lambda$ and shift the running index to get $\mathbb{E}(X) = \frac{1-e^{-\lambda}}{\lambda}$.

**Exercise 4.3** *(Archery)* An archer shots $n$ arrows at a target. The shots are independent, and each shot hits the target with probability $p$. Let $X$ be the random variable "*number of times the target is hit*".
1. What is the law of $X$?
2. What is the expectation of $X$?
3. What is the value of $p$ that maximises the variance of $X$?

The archer bets on his result. He gets $g$ euros when he hits the target, and loses $l$ euros when he misses it. Let $Y$ be the random variable that represent the net gain of the archer at the end of the $n$ shots.

4. What is the expectation of $Y$?
5. What is the relation between $g$ and $l$ that guarantees the archer an expected gain of zero?

1. $X$ is the sum of $n$ independent $p$-Bernoullli variables, hence $X \sim \mathcal{B}(n, p)$ (binomial distribution).
2. We have to compute the expectation of a binomial random variable $X = X_1 + \ldots + X_n$, where each $X_i$ is a Bernoulli variable. Since expectations are linear we can compute the expectation of the Bernoulli variables, and sum them:

$$\mathbb{E}(X_i) = 1 \cdot p + 0 \cdot (1-p) = p \tag{84}$$

$$\mathbb{E}(X_1 + ... + X_n) = \mathbb{E}(X_1) + ... + \mathbb{E}(X_n) = np \tag{85}$$

For example, if $p = 0.5$ and $n = 10$, this means that the archer expects to hit the target 5 times.

3. Let's compute the variance of a Bernoulli and a binomial variable by Equation 76:

$$\mathbb{E}(X_i^2) = 1^2 \cdot p + 0^2 \cdot (1-p) = p \tag{86}$$

$$\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = p(1-p) \tag{87}$$

By independence, $\mathrm{Var}(X_1 + ... + X_n) = \mathrm{Var}(X_1) + ... + \mathrm{Var}(X_n) = np(1-p)$. To find the value of $p$ that maximises the variance differentiate: $n(1-2p) = 0 \Rightarrow p = 0.5$.

4. Let $Y_i =$ gain of $i$-th shot. Then $\mathbb{E}(Y_i) = gp - l(1-p)$, and

$$\mathbb{E}(Y = Y_1 + ... + Y_n) = n[gp - l(1-p)] \tag{88}$$

For example if $n = 10$, $g = 1$, $l = 2$, we have $\mathbb{E}(Y) = 30p - 20$; and if furthermore $p = 0.5$ then $\mathbb{E}(Y) = -5$.

5. To find the value of relation between $g$ and $l$ required to have an expected gain of zero solve the equation $\mathbb{E}(Y) = 0$ to get

$$\frac{g}{l} = \frac{1-p}{p}. \tag{89}$$

Thus as the probability $p$ to hit the target goes to zero, a very big $\frac{g}{l}$ is required to guarantee an expected gain of zero; viceversa $\frac{g}{l}$ becomes infinitely small as $p \to 1$. At $p = 0.5$, as one would expect, $g = l$.

**Exercise 4.4** *(Martingale doubling betting system)* See Exercise 5.1.

# 5. Theory recap 10.10.24 - Conditional law and conditional expectation

**Tip** All the random variables (RVs) we consider in the following are supposed to be integrable discrete random variables with real values. Recall that
- the *expected value* of a random variable $X : \Omega \to F$ is $\mathbb{E}[X] = \sum_{x \in F} xP(X = x)$;
- the *law* of $X$ is the function $F \to [0, 1]$ defined by $x \longmapsto P(X = x)$.

## 5.1. Marginal law

Let $X : \Omega \to F$ and $Y : \Omega \to G$ be random variables. Knowing the law of $X$ and the law of $Y$ is not enough to obtain the law of $(X, Y)$, unless we have further information on how $X$ and $Y$ are related. If they are independent, then $P(X = x, Y = y) = P(X = x)P(Y = y)$.

- Conversely, knowing the law of $(X, Y)$ we can recover the laws of $X$ and $Y$ by the *marginal law*:

$$P(X = x) = \sum_{y \in G} P(X = x, Y = y). \tag{90}$$

- If the law of $(X, Y)$ **factors** as a product,

$$P(X = x, Y = y) = c\mu(x)\nu(y) \text{ for all } x \in F, y \in G, \tag{91}$$

then $X$ and $Y$ are **independent** (cf. [2] Rem. 2.2.12.)

## 5.2. Conditional law

Let $X$ and $Y$ be random variables.

**Definition 5.2.1** The *conditional law of $X$ given* $(Y = y)$ is the function from $F$ to $[0, 1]$ defined by

$$x \longmapsto P(X = x \mid Y = y). \tag{92}$$

- By definition of conditional probability (Equation 26),

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} \tag{93}$$

- Recall that $(X = x, Y = y) \equiv \{X = x\} \cap \{Y = y\}$ is a short-hand for the intersection of two events.

## 5.3. Conditional expectation

- Let $X : \Omega \to F$ and $Y : \Omega \to G$ be random variables.
- Consider a function $f : F \times G \to \mathbb{R}$.
- $f(X, Y)$ is a random variable (assume it is integrable).

**Definition 5.3.1** The *conditional expectation of $f(X, Y)$ given $Y$* is a **discrete random variable**; it is denoted by $\mathbb{E}[f(X, Y) \mid Y]$, and it is defined by

$$\mathbb{E}[f(X, Y) \mid Y] := \psi(Y) \tag{94}$$

with $\psi : G \to \mathbb{R}$ defined by

$$\psi(y) = \sum_{x \in F} f(x, y) P(X = x \mid Y = y) \tag{95}$$

Note that the probability appearing in the sum is the conditional law of $X$ given $Y = y$.

**Special cases**

- If $f(x, y) = x$ the conditional expectation of $X$ given $Y$ is the discrete random variable

$$\mathbb{E}[X \mid Y] = \psi(Y) \text{ with } \psi(y) = \sum_{x \in F} x P(X = x \mid Y = y). \tag{96}$$

- If $X$ and $Y$ are **independent** we have

$$\psi(y) = \sum_{x \in F} f(x, y) P(X = x) = \mathbb{E}[f(X, y)]. \tag{97}$$

- Combining the previous two cases, if $f(x, y) = x$ and $X, Y$ are independent we have

$$\mathbb{E}[X \mid Y] = \psi(Y) \text{ with } \psi(y) = \mathbb{E}[X]. \tag{98}$$

### Expected value

Now since $\psi(Y) = \mathbb{E}[f(X, Y) \mid Y]$ is a RV it makes sense to look at its expected value $\mathbb{E}[\psi(Y)]$.

- Let $X : \Omega \to F$ and $Y : \Omega \to G$ be random variables
- consider functions $f : F \times G \to \mathbb{R}$ and $g : G \to \mathbb{R}$
- let $\mathbb{E}[f(X, Y) \mid Y] = \psi(Y)$ as above
- $\psi(Y)$ and $g(Y)$ are random variables, and so is their product $\psi(Y)g(Y)$
- similarly, $f(X, Y)$ and $g(Y)$ are random variables, and so is their product $f(X, Y)g(Y)$
- the expected values of these RVs are related by the following

**Proposition 5.3.2**

$$\mathbb{E}[\psi(Y)g(Y)] = \mathbb{E}[f(X,Y)g(Y)] \tag{99}$$

whenever all the RVs involved are integrable.

By setting $g \equiv 1$:

**Corollary 5.3.3**

$$\mathbb{E}[\psi(Y)] = \mathbb{E}[f(X,Y)] \tag{100}$$

whenever all the RVs involved are integrable.

> **Tip**
> - Let $X : \Omega \to F$ and $Y : \Omega \to G$ be random variables.
> - The law of $X$ is $x \longmapsto P(X = x)$
> - The expectation of $X$ is $\mathbb{E}[X] = \sum_{x \in F} xP(X = x)$
> - The conditional law of $X$ given $(Y = y)$ is $x \longmapsto P(X = x \mid Y = y)$
> - The conditional expectation of $X$ given $Y$ is $\mathbb{E}[X \mid Y] = \psi(Y)$ with
> - $\psi(y) = \sum_{x \in F} xP(X = x \mid Y = y)$

**!to check** ——————————————————————

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X] \tag{101}$$

## Exercises

> **Exercise 5.1** *(Martingale doubling betting system)* Let $X_n \sim \mathcal{B}(p)$, $n = 1, 2, 3, 4...$ be a sequence of independent Bernoulli variables (set $P(X = 0) = q$ and $P(X = 1) = p$ with $q + p = 1$, $q > 0$, $p > 0$). Fix a number $b_1 > 0$. Before each $X_n$ is drawn, you bet on the outcome $X_n = 1$ the amount $b_n$ defined recursively by $b_n = 2b_{n-1}$. If you win the bet, you receive $g_n = 2b_n$ and the game ends; else, you keep going.
> 1. What is your final gain?
> 2. What is the expected amount of money bet?

First, note that the amount placed on the $n$-th bet can be written as $b_n = 2^{n-1}b_1$. Let's then write down explicitely the process:

---

0. You bet $b_1$ on the event $X_1 = 1$;
1. $X_1$ is drawn:
   - if $X_1 = 1$, you get $g_1$ and the game ends;
   - else, you bet $b_2$ on the event $X_2 = 1$.
2. $X_2$ is drawn:
   - if $X_2 = 1$, you get $g_2$ and the game ends;
   - else, you bet $b_3$ on the event $X_3 = 1$.

...

n. $X_n$ is drawn:
- if $X_n = 1$, you get $g_n$ and the game ends;
- else, you bet $b_{n+1}$ on the event $X_{n+1} = 1$.

...

### Final gain

After placing $n$ bets the invested amount is

$$B_n = \sum_{k=1}^{n} b_k = \sum_{k=1}^{n} b_1 2^{k-1} = b_1 \sum_{l=0}^{n-1} 2^l = b_1 \frac{1-2^n}{1-2} = b_1(2^n - 1). \tag{102}$$

If the $n$-th placed bet is the winning one you receive $g_n = 2b_n$ and the game ends, so your final gain is always

$$2b_n - B_n = 2^n b_1 - (2^n - 1)b_1 = b_1. \tag{103}$$

This result is surprising: with this strategy (doubling your bet until you win), no matter how big $n$ is (that is, how late you win) and how small the probability $p$ to win is, *at the end of the day you go home with a net gain equal to your initial bet.* But before running to the closest casino, beware! If the probability to win is sufficiently small this well known strategy[3] can lead you to invest colossal sums which may well bankrupt you before you get your win.

**Expected bet amount** The first occurrence of a winning bet is described by a random variable $X \sim \mathcal{G}(p)$ with values $F = \{1, 2, 3, ...\} \ni n$. After placing $n$ bets, the amount of invested money is the function $B : F \to \mathbb{R}$ with $B_n = b_1(2^n - 1)$ as per Equation 102. By Theorem 4.1.3 and Equation 57 the expected amount of money bet is

$$\mathbb{E}(B(X)) = \sum_{n=1}^{\infty} B_n P(X = n) = \sum_{n=1}^{\infty} B_n (1-p)^{n-1} p \tag{104}$$

For clarity let's write down more explicitely what is going on, recalling that $P(X_i = 0) = q$ is the probabiloty to lose each bet and $P(X_i = 1) = p$ is the probability to win each bet, with $q + p = 1$.

$$\begin{aligned} \mathbb{E}(B(X)) = b_1 \quad & P(X_1 = 1) + \\ +(b_1 + b_2) \quad & P(X_1 = 0, X_2 = 1) + \\ +(b_1 + b_2 + b_3) & P(X_1 = 0, X_2 = 0, X_3 = 1) + \\ +... + & \\ +(b_1 + ... + b_n) & P(X_1 = 0, ..., X_{n-1} = 0, X_n = 1) + ... \end{aligned} \tag{105}$$

Then

$$\begin{aligned} \mathbb{E}(B(X)) &= b_1 \sum_{n=1}^{\infty} (2^n - 1) q^{n-1} p \\ &= b_1 \frac{p}{q} \sum_{n=1}^{\infty} [(2q)^n - q^n] \end{aligned} \tag{106}$$

Recall that the geometric series $\sum_{i=0}^{\infty} r^i$ with $r \geq 0$ converges to $\frac{1}{1-r}$ iff $r < 1$, else it diverges to $+\infty$; and analogously for $\sum_{i=1}^{\infty} r^i = \frac{r}{1-r}$. Since $q < 1$ the second series converges, whereas the first series converges iff $2q < 1$, that is if the probability $p$ to win is strictly bigger than 0.5:

$$\text{Finite expected bet amount} \Leftrightarrow \text{prob. to win } p > \frac{1}{2}. \tag{107}$$

[3] Martingale betting system and some random discussions on the matter.

*Thus, if $p \leq \frac{1}{2}$, the expected invested amount diverges to infinity.* On the other hand, if $p > \frac{1}{2}$:

$$\begin{aligned} \mathbb{E}(B(X)) &= b_1 \frac{p}{q} \sum_{n=1}^{\infty} [(2q)^n - q^n] \\ &= b_1 \frac{p}{q} \left[ \frac{2q}{1-2q} - \frac{q}{1-q} \right] \\ &= \frac{b_1}{2p-1}. \end{aligned} \tag{108}$$

Reasonably, the expected invested amount diverges to $+\infty$ as $p \to \frac{1}{2}$ from the right, and is equal to $b_1$ if there is certainty to win ($p = 1$). The graph of $\left(\frac{1}{2}, 1\right] \ni p \longmapsto \frac{1}{2p-1}$ is shown below:

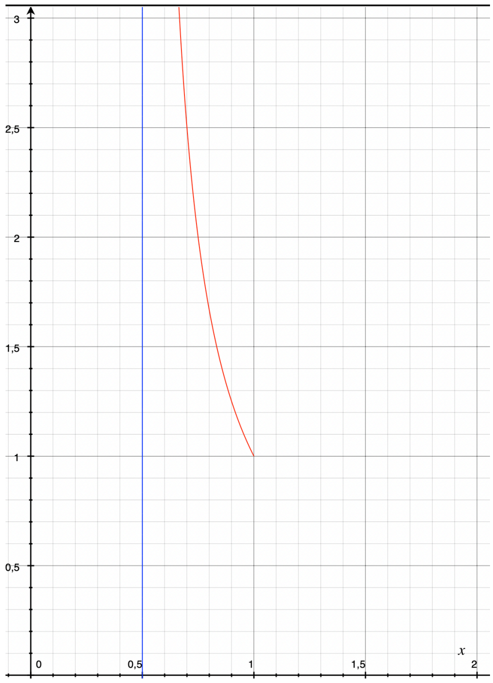


Figure 3: Horizontal: probability $p$ of each winning Bernoulli event. Vertical: expected bet amount.

> **Tip** Whenever a "special" number appears in your results, ask yourself whether it arises from the theory or whether it is imposed by the model at hand. Why is the probability $\frac{1}{2}$ special? Is it a consequence of our modeling choices, like the "doubling" betting strategy, or is it intrinsic to this problem (if there were an outcome more likely than the other, then you would bet on that one)? Try to generalize this scenario to different betting and reward policies, like $b_n = \alpha b_{n-1}$ and $g_n = \beta b_n$ for some $\alpha, \beta > 0$.

> **Exercise 5.2** *(Conditional law and expectation 1)* Let $X_1, \ldots, X_n$ be independent and identically distributed (I.I.D.) random variables following $\mathcal{B}(p)$ with $p \in (0, 1)$, and let $S = X_1 + \ldots + X_n$. For $s \in [0, n]$,
> 1. find the conditional law of $X_1$ given $S = s$, and
> 2. find the conditional expectation of $X_1$ given $S$.

**Hint**: use Definition 5.2.1 to find the conditional law, and Equation 96 to compute the conditional expectation.

1. By Definition 5.2.1 the conditional law of $X_1$ given $S = s$ is

$$
\begin{aligned}
x \longmapsto & P(X_1 = x \mid S = s) \\
&= \frac{P(X_1 = x, S = s)}{P(S = s)} \\
&= \frac{P(X_1 = x, X_2 + \ldots + X_n = s - x)}{P(S = s)} \\
&= P(X_1 = x)\frac{P(X_2 + \ldots + X_n = s - x)}{P(S = s)}
\end{aligned} \tag{109}
$$

We have that $X_1 \sim \mathcal{B}(p)$, $X_2 + \ldots + X_n \sim \mathcal{B}(n-1,p)$, and $S \sim \mathcal{B}(n,p)$, so

$$
P(X_1 = x \mid S = s) = \frac{p^x(1-p)^{1-x}\binom{n-1}{s-x}p^{s-x}(1-p)^{n-1-s+x}}{\binom{n}{s}p^s(1-p)^{n-s}} = \frac{\binom{n-1}{s-x}}{\binom{n}{s}} \tag{110}
$$

Let's evaluate this for $x \in \{0,1\}$. For $x = 0$ the binomial coefficients simplify to

$$
P(X_1 = 0 \mid S = s) = \frac{n-s}{n}, \tag{111}
$$

while for $x = 1$ we get

$$
P(X_1 = 1 \mid S = s) = \frac{s}{n}. \tag{112}
$$

For $x \in \{0,1\}$ these can be packed into

$$
P(X_1 = x \mid S = s) = \left(\frac{s}{n}\right)^x \left(\frac{n-s}{n}\right)^{1-x} \tag{113}
$$

meaning that the law of $X_1$ given $S = s$ is $\mathcal{B}\left(\frac{s}{n}\right)$; note that this is independent of $p$!

2. For the second part, by Equation 96 the conditional expectation of $X_1$ given $S$ is

$$
\mathbb{E}[X_1 \mid S] = \psi(S) \text{ with } \psi(s) = \sum_{x \in \{0,1\}} xP(X_1 = x \mid S = s) = \frac{s}{n} \tag{114}
$$

Hence, the conditional expectation of $X_1$ given $S$ is the random variable

$$
\mathbb{E}[X_1 \mid S] = \frac{S}{n}. \tag{115}
$$

**Exercise 5.3** *(Conditional law and expectation 2)* Let $X \sim \mathcal{P}(\lambda)$ and $Y \sim \mathcal{P}(\mu)$ be independent random variables.
1. Find the conditional law of $X$ given $X + Y = s$, and
2. Find the conditional expectation of $X$ given $X + Y$.

1. Let $S = X + Y$. The conditional law of $X$ given $S = s$ is

$$
\{0, 1, \ldots, s\} \ni x \longmapsto P(X = x \mid S = s) \tag{116}
$$

Following the steps of the previous exercise we readily arrive to

$$
P(X = x \mid S = s) = \frac{P(X = x)P(Y = s - x)}{P(S = s)}. \tag{117}
$$

Recall by Exercise 3.2 that $S \sim \mathcal{P}(\mu + \lambda)$, so

$$P(X = x \mid S = s) = \frac{e^{-\lambda}\lambda^x}{x!} \frac{e^{-\mu}\mu^{s-x}}{(s-x)!} \frac{s!}{e^{-(\lambda+\mu)}(\lambda+\mu)^s} \tag{118}$$

which can be manipulated into

$$P(X = x \mid S = s) = \binom{s}{x} p^x (1-p)^{s-x} \text{ with } p = \frac{\lambda}{\lambda+\mu}. \tag{119}$$

Hence the law of $X$ given $S = s$ is binomial with $s$ iterations and probability $p$.

2. To find the conditional expectation of $X$ given $S$, solve

$$\mathbb{E}[X \mid S] = \psi(S) \text{ with } \psi(s) = \sum_{x=0}^{s} x \binom{s}{x} p^x (1-p)^{s-x} = sp \tag{120}$$

The series can be solved directly (see for example this page), but it's much more convenient to recognise this as $\mathbb{E}[\mathcal{B}(s,p)]$, the expected value of the binomial distribution, which by linearity (c.f. Section 4.1) is immediately concluded to be $sp$ (where as above $p = \frac{\lambda}{\lambda+\mu}$). So in conclusion the conditional expectation of $X$ given $X+Y$ is the random variable

$$\mathbb{E}[X \mid X+Y] = \frac{\lambda}{\lambda+\mu}(X+Y). \tag{121}$$

**Exercise 5.4** *(An electronic component has a lifetime...)* Exercice corrigé 2.6.4. in [2].

**Exercice corrigé 2.6.4.** Un composant électronique a une durée de vie $X$ qu'on mesure en nombre entier d'unités de temps. On fait l'hypothèse que, à chaque unité de temps, ce composant a une probabilité $p \in ]0,1[$ de tomber en panne, de sorte que $X \sim \mathcal{G}eo(p)$. On considère un autre composant dont la durée de vie $Y$ est indépendante de $X$ et de même loi. On pose

$$S = \min(X, Y) \text{ et } T = |X - Y|.$$

1. Que représentent $S$ et $T$ ?
2. Calculer $\mathbb{P}(S = s \text{ et } T = t)$ pour $s \geq 1$ et $t \geq 0$ (distinguer $t = 0$ de $t \geq 1$).
3. En déduire les lois de $S$ et $T$ puis $\mathbb{E}(T)$. Quel est le nom de la loi de $S$ ?
4. Les variables aléatoires $S$ et $T$ sont-elles indépendantes ?

1. **Que représentent $S$ et $T$ ?**
S représente le premier temps de panne et $T$ la durée qui sépare les deux temps de panne.
2. **Calculer $\mathbb{P}(S = s$ et $T = t)$ pour $s \geq 1$ et $t \geq 0$ (distinguer $t = 0$ et $t \geq 1$).**
Pour $s \geq 1$ et $t \geq 0$, on a toujours

$$\{S = s, T = t\} = \{X = s, Y = s + t\} \cup \{X = s + t, Y = s\}$$

mais l'union n'est disjointe que si $t \geq 1$ (sinon les deux événements sont égaux à $\{X = Y = s\}$). Donc pour $s, t \geq 1$, on a, en utilisant l'indépendance de $X$ et $Y$,

$$\begin{aligned}\mathbb{P}(S = s \text{ et } T = t) &= \mathbb{P}(X = s \text{ et } Y = s + t) + \mathbb{P}(Y = s \text{ et } X = s + t)\\ &= \mathbb{P}(X = s)\mathbb{P}(Y = s + t) + \mathbb{P}(Y = s)\mathbb{P}(X = s + t)\\ &= 2p^2(1 - p)^{2s+t-2}.\end{aligned}$$

Pour $s \geq 1$ et $t = 0$, on a, toujours par indépendance de $X$ et $Y$,

$$\begin{aligned}\mathbb{P}(S = s \text{ et } T = 0) &= \mathbb{P}(X = s \text{ et } Y = s)\\ &= \mathbb{P}(X = s)\mathbb{P}(Y = s) = p^2(1 - p)^{2s-2}.\end{aligned}$$

On conclut que

$$\forall s \geq 1,\ \forall t \geq 0,\ \mathbb{P}(S = s, T = t) = p^2\left(1 + 1_{\{t>0\}}\right)(1 - p)^t(1 - p)^{2(s-1)}. \tag{10.1}$$

3. **En déduire les lois de $S$ et $T$ puis $\mathbb{E}(T)$. Quel est le nom de la loi de $S$ ?**
Pour $s \geq 1$, on a par la formule des lois marginales

$$\begin{aligned}\mathbb{P}(S = s) &= \sum_{t=0}^{\infty} \mathbb{P}(S = s \text{ et } T = t)\\ &= p^2(1 - p)^{2s-2} + \sum_{t=1}^{\infty} 2p^2(1 - p)^{2s+t-2}\\ &= (1 - p)^{2(s-1)}\left(p^2 + 2p^2(1 - p)\sum_{u=0}^{\infty}(1 - p)^u\right) \text{ où } u = t - 1\\ &= \left(p^2 + 2p^2(1 - p)\frac{1}{p}\right)(1 - p(2 - p))^{s-1} = p(2 - p)(1 - p(2 - p))^{s-1}.\end{aligned}$$

On reconnaît la loi géométrique de paramètre $p(2 - p)$.
Maintenant, toujours par la formule des lois marginales, pour $t \geq 0$,

$$\begin{aligned}\mathbb{P}(T = t) &= \sum_{s=1}^{\infty} \mathbb{P}(S = s \text{ et } T = t) = \sum_{s=1}^{\infty}\left(1 + 1_{\{t>0\}}\right)p^2(1 - p)^{2(s-1)+t}\\ &= \left(1 + 1_{\{t>0\}}\right)p^2(1 - p)^t\sum_{u=0}^{\infty}(1 - p(2 - p))^u = \left(1 + 1_{\{t>0\}}\right)\frac{p(1 - p)^t}{2 - p}.\end{aligned}$$

D'où

$$\begin{aligned}\mathbb{E}(T) &= \sum_{t=0}^{\infty} t\mathbb{P}(T = t) = \sum_{t=1}^{\infty} t\frac{2p(1 - p)^t}{2 - p}\\ &= \frac{2p(1 - p)}{2 - p}\sum_{t=1}^{\infty} t(1 - p)^{t-1} = \frac{2p(1 - p)}{2 - p}f'(1 - p),\end{aligned}$$

où $f(x) = \sum_{t=0}^{\infty} x^t = \frac{1}{1-x}$. Comme $f'(x) = \frac{1}{(1-x)^2}$, $f'(1-p) = 1/p^2$ et $\mathbb{E}(T) = \frac{2(1-p)}{p(2-p)}$.
4. **Les variables aléatoires $S$ et $T$ sont-elles indépendantes ?**
On peut montrer l'indépendance en utilisant la définition. En effet, on a pour $s \geq 1$ et $t \geq 0$,

$$\begin{aligned}\mathbb{P}(S = s)\mathbb{P}(T = t) &= p(2 - p)[(1 - p)^2]^{s-1}\left(1 + 1_{\{t>0\}}\right)\frac{p(1 - p)^t}{2 - p}\\ &= p^2\left(1 + 1_{\{t>0\}}\right)(1 - p)^{t+2(s-1)} = \mathbb{P}(S = s \text{ et } T = t),\end{aligned}$$

d'après (10.1). On peut aussi remarquer sur (10.1) que la loi du couple $(S, T)$ se met sous forme produit $c\mu(s)\nu(t)$ et conclure que ces variables sont indépendantes en utilisant la remarque 2.2.12.

# 6. Theory recap (17 + 24).10.24 - Continuous random variables

## 6.1. PDF, CDF, expected value, variance for scalar random variables

**Notation** Consider a probability space $(\Omega, \mathcal{A}, P)$, a function $X : \Omega \to \mathbb{R}$, and two points $a < b \in \hat{\mathbb{R}}$, with $\hat{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. We use the following notation:

$$\begin{aligned} \{a < X \leq b\} &:= \{\omega \in \Omega : X(\omega) \in (a, b]\} \\ &\equiv \{\omega \in \Omega : X(\omega) \leq b\} \cap \{\omega \in \Omega : X(\omega) > a\} \\ &\equiv \{\omega \in \Omega : X(\omega) \leq b\} \setminus \{\omega \in \Omega : X(\omega) \leq a\}. \end{aligned} \tag{122}$$

Note that $\{a < X \leq b\}$ is a subset of $\Omega$, and recall that a subset of $\Omega$ is an *event* if it belongs to $\mathcal{A}$.

**Definition 6.1.1** *(Continuous random variable)* Given a probability space $(\Omega, \mathcal{A}, P)$, a *continuous random variable* $X$ is a function $X : \Omega \to \mathbb{R}$ such that $\{a < X \leq b\}$ is an event for all $a < b \in \hat{\mathbb{R}}$.

**Definition 6.1.2** *(Probability density function - scalar)* The *probability density function (PDF)* of a continuous random variable $X : \Omega \to \mathbb{R}$ is the function $\rho_X : \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that, for all $a < b \in \hat{\mathbb{R}}$,

$$P\{a < X \leq b\} = \int_a^b \rho_X(x) dx. \tag{123}$$

- $\int_{\mathbb{R}} \rho(x) dx = 1$ for any PDF
- $P(X = x) = 0$ for all $x \in R$

**Definition 6.1.3** *(Cumulative distribution function)* The *cumulative distribution function (CDF)* of a continuous random variable $X : \Omega \to \mathbb{R}$ is the function $F_X : \mathbb{R} \to [0, 1]$ defined by

$$F_X(x) = P(X \leq x). \tag{124}$$

**Proposition 6.1.4** *(PDF = CDF')* The probability density function $\rho_X$ of a continuous random variable $X : \Omega \to \mathbb{R}$ can be determined from the cumulative distribution function $F_X$ by differentiating (as long as the derivative exists):

$$\rho_X = F'_X \tag{125}$$

**Proof** By Equation 122, $\sigma$-additivity, and the fundamental theorem of integral calculus,

$$\begin{aligned} P\{a < X \leq b\} &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \\ &= \int_a^b F'_X(x) dx. \end{aligned} \tag{126}$$

□

**Definition 6.1.5** The *expected value* and *variance* of a RV $X : \Omega \to \mathbb{R}$ with PDF $\rho_X$ are

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \rho_X(x) dx. \tag{127}$$

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \tag{128}$$

**Tip**
- integral of PDF gives probability to observe RV in an interval
- CDF gives probability to observe RV below a value
- $\mathrm{PDF} = \mathrm{CDF}'$
- $\mathbb{E}[X] = \int_{\mathbb{R}} x \ \mathrm{PDF}_X(x) dx.$

## 6.2. Vector-valued random variable

Given a probability space $(\Omega, \mathcal{A}, P)$ let now $X$ be a **vector-valued random variable**, that is a map $X : \Omega \to \mathbb{R}^n$ such that $\{X \in D\} := \{\omega \in \Omega : X(\omega) \in D\}$ is an event for all open $D \subseteq \mathbb{R}^n$.

**Definition 6.2.1** *(Probability density function - vector)* The *probability density function (PDF)* of a vector-valued RV $X : \Omega \to \mathbb{R}^n$ is the function $\rho_X : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ s.t., for all open $D \subseteq \mathbb{R}^n$,

$$P\{X \in D\} = \int_D \rho_X(x) dx = \int_{\mathbb{R}^n} \mathbb{1}_D(x) \rho_X(x) dx. \tag{129}$$

### 6.2.1. Real-valued RVs as function of vector-valued RVs

- Let $X : \Omega \to \mathbb{R}^n$ be a vector-valued random variable
- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function.
- By composition one can build the **real-valued** random variable $f(X) : \Omega \to \mathbb{R}$

**Example**

- Let $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ be continuous real-valued random variables.
- Together they make the vector-valued random variable $(X, Y) : \Omega \to \mathbb{R}^2$.
- Let now $f : \mathbb{R}^2 \to \mathbb{R}$ be a real-valued function, say $f(a, b) = a + b$
- Then $f(X, Y) = X + Y : \Omega \to \mathbb{R}$ is a real-valued random variable.

### 6.2.2. Théorème de transfert / Méthode de la fonction muette

We have the following generalization of the "Théorème de transfert" concerning the expected value of a real-valued random variable obtained as the composition of a vector-valued random variable with a real-valued function:

**Theorem 6.2.2.2** *(Théorème de transfert vectoriel)* The vector-valued random variable $X : \Omega \to \mathbb{R}^n$ has the probability density function $\rho_X$ if and only if, for any bounded function $f : \mathbb{R}^n \to \mathbb{R}$, the real-valued random variable $f(X) : \Omega \to \mathbb{R}$ has expected value

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^n} f(x) \rho_X(x) dx. \tag{130}$$

**Tip** Theorem 6.2.2.2 is very useful to find the pdf of a function of a real-valued RV $X : \Omega \to \mathbb{R}$. Say we have $X \sim \rho_X$, and we are asked to find $\rho_{f(X)}$ for some $f : \mathbb{R} \to \mathbb{R}$. Let $h : \mathbb{R} \to \mathbb{R}$ be a generic test function; then by the transfert theorem, thinking of the function $h \circ f$ acting on the RV $X$,

$$\mathbb{E}[h(f(X))] = \int_{\mathbb{R}} h \circ f(x) \rho_X(x) dx. \tag{131}$$

Thinking of the function $h$ acting on the RV $f(X)$, $\mathbb{E}[h(f(X))]$ is also equal to

$$\mathbb{E}[h(f(X))] = \int_{\mathbb{R}} h(y) \rho_{f(X)}(y) dy \tag{132}$$

So if we can cast the first integral into the second, usually by the change of variables $y = f(x)$, we can read off $\rho_{f(X)}$:

$$\begin{aligned} \mathbb{E}[h(f(X))] &= \int_{\mathbb{R}} h \circ f(x) \rho_X(x) dx \qquad \text{change variable: } y = f(x) \\ &= \int_{\mathrm{Im}(f)} h(y) \frac{\rho_X(x)}{f'(x)} dy \qquad \text{with } x = f^{-1}(y), \end{aligned} \tag{133}$$

meaning by the transfert theorem that the PDF of $f(X)$ is

$$\rho_{f(X)}(y) = \frac{\rho_X(x)}{f'(x)} \mathbb{1}_{\{\mathrm{Im}(f)\}}(y) \qquad \text{with } x = f^{-1}(y). \tag{134}$$

**Example 6.2.2.3** Let $X : \Omega \to \mathbb{R} \sim \rho_X$ with $\rho_X(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ (cf Section 6.5.3). We want to find the pdf of $f(X) = e^X$. Let $h : \mathbb{R} \to \mathbb{R}$ be a test function; then

- $y = f(x) = e^x \Leftrightarrow x = \ln(y)$
- $f'(x) = e^x$
- $\mathrm{Im}(f) = (0, \infty)$

So

$$\rho_{e^X}(y) = \frac{e^{-\frac{(\ln y)^2}{2}}}{\sqrt{2\pi}} \cdot \frac{1}{y} \cdot \mathbb{1}_{(0,\infty)}(y) \tag{135}$$

We can double-check by performing the changhe of variable in the integral:

Soit $h$ continue bornée.

— On a que

$$\begin{aligned} \mathbb{E}[h(\mathrm{e}^X)] &= \int_{\mathbb{R}} h(\mathrm{e}^x) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}} \, \mathrm{d}x \\ &= \int_{\mathbb{R}_+^*} h(y) \frac{1}{y\sqrt{2\pi}} \mathrm{e}^{-\frac{(\ln y)^2}{2}} \, \mathrm{d}y \text{ par le changement de variable } y = \mathrm{e}^x. \end{aligned}$$

Donc $\mathrm{e}^X$ a pour densité $f_{\mathrm{e}^X} : y \mapsto \frac{1}{y\sqrt{2\pi}} \exp\left(-\frac{(\ln y)^2}{2}\right) \mathbb{1}_{]0;+\infty[}(y)$.

## 6.3. Indicator technique to find the CDF - scalar case

There exists a very useful technique to find the CDF of a RV (and hence its PDF, by differentiating) by expressing the CDF as the expectation of an indicator function, that can be computed by Theorem 6.2.2.2.

**Proposition 6.3.1** *(Indicator technique to find the CDF)* Let $X : \Omega \to \mathbb{R}$ be a real-valued random variable. Then its CDF $F_X(x) = P(X \leq x)$ is given by the expected value of the indicator function of $\{X \leq x\}$, that is

$$F_X(x) = \mathbb{E}\left[\mathbb{1}_{\{X \leq x\}}\right]. \tag{136}$$

Before the proof, let's see make sure this is well posed. Consider the RV $X : \Omega \to \mathbb{R}$, and fix some $x \in \mathbb{R}$. Consider the real-valued indicator function $\mathbb{1}_{\{\leq x\}} : \mathbb{R} \to \mathbb{R}$, defined by $\mathbb{1}_{\{\leq x\}}(z) = 1$ if $z \leq x$ and 0 else. As usual, composing the real-valued function $\mathbb{1}_{\{\leq x\}}$ with the real-valued RV $X$ is itself a random variable, and we denote this random variable by $\mathbb{1}_{\{X \leq x\}} := \mathbb{1}_{\{\leq x\}}(X) : \Omega \to \mathbb{R}$. The expected value of this random variable can then be found by Theorem 6.2.2.2.

**Proof** Let $x \in \mathbb{R}$ and consider the function

$$\begin{aligned} x \longmapsto \mathbb{E}\left[\mathbb{1}_{\{X \leq x\}}\right] &= \int_{\mathbb{R}} \mathbb{1}_{\{\leq x\}}(y)\rho_X(y)dy && \text{by transfert theorem} \\ &= \int_{-\infty}^{x} 1 \cdot \rho_X(y)dy && \text{by def of indicator function} \\ &= P(-\infty \leq X \leq x) && \text{by def of pdf} \\ &= F_X(x) && \text{by def of cdf,} \end{aligned} \tag{137}$$

which is what we need. □

> **Tip** To find the PDF of a *real-valued* random variable, try
> - first finding the CDF by the indicator technique (Proposition 6.3.1),
> - then differentiating (Proposition 6.1.4).

## 6.4. Indicator technique to find the CDF - vector case

The technique of the previous secion generalizes to real-valued functions of vector-valued random variables. We present the idea for 2 RVs, but it readily generalises to $n$.

As in Section 6.2.1,
- Let $X : \Omega \to \mathbb{R}^n$ be a vector-valued random variable
- Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function.
- By composition one can build the **real-valued** random variable $f(X) : \Omega \to \mathbb{R}$
- We can use the indicator technique (Proposition 6.3.1) to find the CDF of $f(X, Y)$, and
- differentiate it (Proposition 6.1.4) to find the PDF of $f(X, Y)$.

A key step in this procedure relies on the *independence* of the RVs:

**Definition 6.4.1** Two continuous random variables $X, Y$ are *independent* if the density of $(X, Y)$ is the product of the densities of $X$ and $Y$:

$$\rho_{(X,Y)}(x, y) = \rho_X(x)\rho_Y(y). \tag{138}$$

**Tip** Given the idependent RVs $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ and the function $f : \mathbb{R}^2 \to \mathbb{R}$, find the CDF of $f(X, Y)$.

$$\begin{aligned} z \longmapsto P(f(X,Y) \le z) &= \mathbb{E}\left[\mathbb{1}_{\{f(X,Y)\le z\}}\right] && \text{by the indicator technique} \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{f(x,y)\le z\}}(x,y)\rho_{X,Y}(x,y)dxdy && \text{by transfert theorem} \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{f(x,y)\le z\}}(x,y)\rho_X(x)\rho_Y(y)dxdy && \text{by independence} \end{aligned} \tag{139}$$

Now the problem is solved up to the resolution of an integral, which usually involves taking care of the integration domain, changes of variables, etc.

See Equation 143, Exercise 6.1, Exercise 6.2, Exercise 6.3 for some examples.

**Tip** To find the PDF of a *function* of *independent* random variables, try
- first finding the CDF by the indicator technique (Proposition 6.3.1),
- then differentiating (Proposition 6.1.4).

### 6.4.1. PDF of sum of RVs: Convolution

**Tip** c.f. [1] pag. 266 and this Wikipedia page plus the references therein for further details.

In the special case of $f(X, Y) = X + Y$ the pdf $\rho_{X+Y}$ takes an important form.

Let $X : \Omega \to F, Y : \Omega \to G$ be *discrete independent* random variables, and let $Z = X + Y$. Recall by Exercise 3.1 that the law of $Z$ is given by

$$\begin{aligned} z \longmapsto\ & P(X+Y=z) \\ &= \sum_{\substack{x\in F\\ y\in G\\ x+y=z}} P(X=x)P(Y=y) \\ &= \sum_{x\in F} P(X=x)P(Y=z-x) \end{aligned} \tag{140}$$

This generalizes to continuous random variables as follows:

**Definition 6.4.1.2** *(Convolution)* Let $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ be continuous random variables with densities $\rho_X, \rho_Y$ respectively. The *convolution* of $\rho_X$ and $\rho_Y$ is the function $\rho_X * \rho_Y$ defined by

$$\rho_X * \rho_Y(z) := \int_{\mathbb{R}} \rho_X(x)\rho_Y(z-x)dx. \tag{141}$$

**Proposition 6.4.1.3** If $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are **independent** continuous random variables then the PDF of their sum is the convolution of their PDFs:

$$\rho_{X+Y} = \rho_X * \rho_Y. \tag{142}$$

**Proof** We first find the CDF of $X + Y$ following the steps of Equation 139 specializing to $f(X, Y) = X + Y$, then differentiate to find the PDF.

$$
\begin{aligned}
z \longmapsto P(X+Y \leq z) &= \mathbb{E}\left[\mathbb{1}_{\{X+Y\leq z\}}\right] && \text{by the indicator technique} \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\leq z\}}(x,y)\rho_{X,Y}(x,y)dxdy && \text{by transfert theorem} \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\leq z\}}(x,y)\rho_X(x)\rho_Y(y)dxdy && \text{by independence.}
\end{aligned} \tag{143}
$$

Now one needs to set the correct domain of integration: the indicator function $\mathbb{1}_{\{x+y\leq z\}}(x,y)$ is 1 when $x+y\leq z$, and 0 elsewhere. This means $x\in(-\infty,+\infty)$ and $y\in(-\infty,z-x)$ - see Figure 8.

So the CDF of $X+Y$ is

$$
F_{X+Y}(z) = P(X+Y\leq z) = \int_{\mathbb{R}} \rho_X(x)\left(\int_{-\infty}^{z-x}\rho_Y(y)dy\right)dx \tag{144}
$$

The PDF follows by differentiating with respect to $z$:

$$
\begin{aligned}
\rho_{X+Y}(z) = \frac{d}{dz}F_{X+Y}(z) &= \int_{\mathbb{R}} \rho_X(x)\frac{d}{dz}\left(\int_{-\infty}^{z-x}\rho_Y(y)dy\right)dx \\
&= \int_{\mathbb{R}} \rho_X(x)\rho_Y(z-x)dx
\end{aligned} \tag{145}
$$

where the last step follows from the fundamental theorem of integral calculus.[4] □

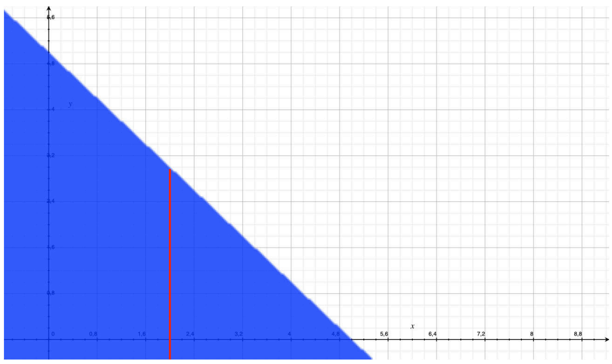

Figure 8: Domain where $\mathbb{1}_{\{x+y\leq z\}}$ is non-zero: $x\in(-\infty,+\infty)$ and $y\in(-\infty,z-x)$.

## 6.5. Standard PDFs

### 6.5.1. Uniform PDF

Given a continuous RV $X:\Omega\to\mathbb{R}$ we write $X\sim\mathcal{U}(a,b)$ and say that $X$ has *uniform* density if

$$
\rho_X(x) = \frac{\mathbb{1}_{[a,b]}(x)}{b-a} \tag{146}
$$

### 6.5.2. Exponential PDF

Given a continuous RV $X:\Omega\to\mathbb{R}$ we write $X\sim\mathcal{E}(\lambda)$ and say that $X$ has *exponential* density if

$$
\rho_X(x) = \begin{cases} 0 & \text{if } x\leq 0 \\ \lambda e^{-\lambda x} & \text{if } x>0 \end{cases} = \lambda e^{-\lambda x}\mathbb{1}_{\{x>0\}} \tag{147}
$$

[4] $\frac{d}{dz}\int_a^{z-b} f(y)dy = \frac{d}{dz}\int_a^{z-b} F'(y)dy$ with $f=F'=\frac{d}{dz}\{F(z-b)-F(a)\}=F'(z-b)=f(z-b)$.

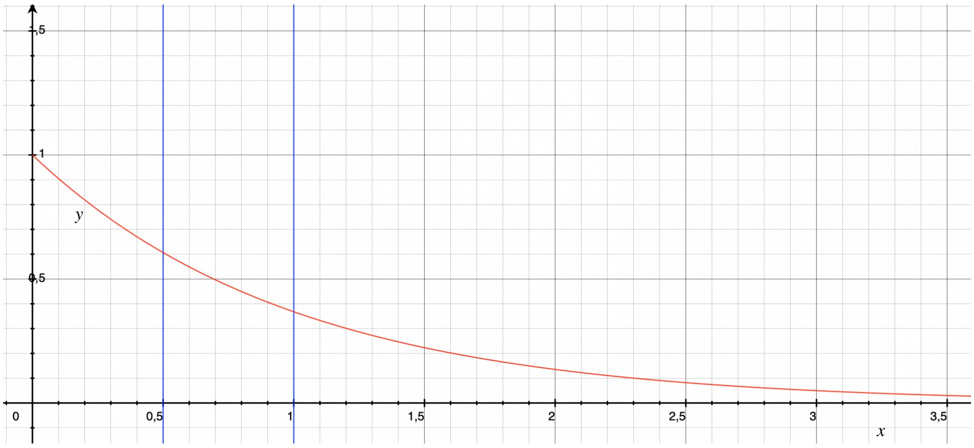


Figure 9: Exponential pdf. $P(0.5 < X < 1) = \int_{0.5}^{1} \rho_X(x)dx$ is the area below the red curve delimited by the blue lines.

#### 6.5.3. Normal PDF

Given a continuous RV $X : \Omega \to \mathbb{R}$ we write $X \sim \mathcal{N}(\mu, \sigma^2)$ and say that $X$ has *normal* density if

$$\rho_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{148}$$

### 6.6. Misc

For the following c.f. Prop. 3.3.6. in [2].

**Proposition 6.6.1** *(Marginal density)* Let $X$ and $Y$ be continuous random variables. If the density $\rho_{(X,Y)}$ of $(X, Y)$ is known then one can obtain the densities of $X$ and $Y$ by

$$\rho_X(x) = \int_{\mathbb{R}} \rho_{X,Y}(x, y)dy \tag{149}$$

and similarly for $\rho_Y$.

## Exercises

**Exercise 6.1** *(Min of exponential variables)* Let $X \sim \mathcal{E}(\lambda)$ and $Y \sim \mathcal{E}(\mu)$. Find the CDF and the PDF of $\min(X, Y)$.

First we find the CDF, then differentiate to get the PDF.

$$\begin{aligned} F_{\min(X,Y)}(z) &= P(\min(X, Y) \leq z) \\ &= P(X \leq z \text{ or } Y \leq z) \\ &= 1 - P(X > z \text{ and } Y > z) \\ &= 1 - P(X > z)P(Y > z) \text{ by independence} \end{aligned} \tag{150}$$

We have $P(X > z) = \int_z^\infty \mu e^{-\mu x} dx = e^{(-\mu z}$, so the CDF is $z \longmapsto 1 - e^{-(\mu+\lambda)z}$, and by taking the derivative with respect to $z$ we find $\rho_{\min(X,Y)}(z) = (\mu + \lambda)e^{-(\mu+\lambda)z}$, so $\min(X, Y) \sim \mathcal{E}(\mu + \lambda)$.

**Exercise 6.2** *(Sum of exponential variables)* Let $X \sim \mathcal{E}(\lambda)$ and $Y \sim \mathcal{E}(\mu)$ with $\mu \neq \lambda$. Find the PDF of $X + Y$.

**Method 1: indicator technique** Let's find the CDF of $X+Y$: by the indicator technique Proposition 6.3.1 and independence,

$$\begin{aligned}
F_{X+Y}(z) &= P(X+Y \le z) = \mathbb{E}\left[\mathbb{1}_{\{X+Y\le z\}}\right] \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\le z\}}(x,y)\rho_X(x)\rho_Y(y)dxdy \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\le z\}}(x,y)\mathbb{1}_{\{x\ge 0\}}(x)\mathbb{1}_{\{y\ge 0\}}(y)\lambda e^{-\lambda x}\mu e^{-\mu y}dxdy \\
&= \int_0^z \lambda e^{-\lambda x}\left(\int_0^{z-x}\mu e^{-\mu y}dy\right)dx \\
&= \text{...some standard calculations...} \\
&= 1-e^{-\lambda z}-\frac{\lambda}{\lambda-\mu}\left[e^{-\mu z}-e^{-\lambda z}\right]
\end{aligned} \tag{151}$$

By taking the derivative with respect to $z$ we get the pdf:

$$\rho_{X+Y}(z) = \frac{\lambda\mu}{\lambda-\mu}\left[e^{-\mu z}-e^{-\lambda z}\right]. \tag{152}$$

This is known as Hypoexponential distribution.

**Method 2: convolution** The indicator technique of the previous section can be used to find the cdf anfd pdf of any $f(X,Y)$. In the particular case $f(X,Y)=X+Y$ we can directly apply Equation 145:

$$\begin{aligned}
\rho_{X+Y}(z) &= \int_{\mathbb{R}} \rho_X(x)\rho_Y(z-x)dx \\
&= \int_0^z \lambda e^{-\lambda x}\mu e^{-\mu(z-x)}dx
\end{aligned} \tag{153}$$

which after some standard steps leads back to Equation 152.

**Exercise 6.3** *(Square of exponential variable)* Let $X \sim \mathcal{E}(\lambda)$ and $Y = X^2$. Find the CDF and the PDF $Y$.

By the indicator technique the cdf is

$$\begin{aligned}
F_{X^2}(z) &= P(X^2 \le z) = \mathbb{E}\left[\mathbb{1}_{\{X^2\le z\}}\right] \\
&= \int_{\mathbb{R}} \mathbb{1}_{\{x^2\le z\}}(x)\rho_X(x)dx \\
&= \int_0^{\sqrt{z}} \rho_X(x)dx = 1-e^{-\lambda\sqrt{z}}
\end{aligned} \tag{154}$$

From which the pdf is

$$\frac{d}{dz}F_{X^2}(z) = \rho_{X^2}(z) = \frac{\lambda}{2}\frac{e^{-\lambda\sqrt{z}}}{\sqrt{z}}. \tag{155}$$

**Exercise 6.4** *(Split a stick into a triangle)* Exercice corrigé 3.5.8. in [2].

**Corrigé de l'exercice 3.5.8.** On coupe un bâton de longueur 1 au hasard en trois morceaux : les abscisses $U$ et $V$ des découpes sont supposées indépendantes et uniformément réparties sur $[0,1]$. On veut calculer la probabilité $p$ pour que l'on puisse faire un triangle avec les trois morceaux (on peut faire un triangle avec trois segments de longueur $l_1$, $l_2$ et $l_3$ si et seulement si $l_1 \leq l_2 + l_3$, $l_2 \leq l_3 + l_1$ et $l_3 \leq l_1 + l_2$).

1. **Exprimer en fonction de $U$ et $V$ les longueurs respectives $L_1$, $L_2$ et $L_3$ du morceau le plus à gauche, du morceau du milieu et du morceau le plus à droite.**

   Il est clair que $L_1 = \inf(U,V)$, $L_2 = \sup(U,V) - \inf(U,V)$ et $L_3 = 1 - \sup(U,V)$.

2. **Montrer que**
$$p = 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$
   En remarquant que lorsque $l_1 + l_2 + l_3 = 1$, la condition $l_1 \leq l_2 + l_3$, $l_2 \leq l_3 + l_1$ et $l_3 \leq l_1 + l_2$ est équivalente à $l_1 \leq \frac{1}{2}$, $l_2 \leq \frac{1}{2}$ et $l_3 \leq \frac{1}{2}$, on en déduit que $p$ est égale à la probabilité de $\left\{L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right\}$. En décomposant cet événement sur la partition $\{U < V\}$, $\{V < U\}$ et $\{U = V\}$ et en utilisant l'égalité $\mathbb{P}(U = V) = 0$, on obtient
$$p = \mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right) + \mathbb{P}\left(V \leq U, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$
   Comme les couples des variables $(U,V)$ et $(V,U)$ ont même loi, les deux probabilités à droite de l'égalité sont égales. Donc
$$p = 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right).$$

3. **Calculer $p$.**
$$\begin{aligned}
p &= 2\mathbb{P}\left(U \leq V, L_1 \leq \frac{1}{2}, L_2 \leq \frac{1}{2}, L_3 \leq \frac{1}{2}\right) \\
&= 2\mathbb{P}\left(U \leq V, U \leq \frac{1}{2}, V - U \leq \frac{1}{2}, V \geq \frac{1}{2}\right) \\
&= 2\int\int 1_{\{x \leq y, x \leq \frac{1}{2}, y - x \leq \frac{1}{2}, y \geq \frac{1}{2}\}} 1_{[0,1]}(x) 1_{[0,1]}(y)\, dx dy \\
&= 2\int_0^{\frac{1}{2}} \left(\int_{\frac{1}{2}}^{x+\frac{1}{2}} dy\right) dx = 2\int_0^{\frac{1}{2}} x dx = \frac{1}{4}.
\end{aligned}$$

# 7. Theory recap - 24.10.24

The theory recap for this session is included in Section 6.

## Exercises

**Exercise 7.1** *(Expected value of functions of normal distribution)* Ler $X \sim \mathcal{N}(0,1)$. Find

1. $\mathbb{E}[|X|]$
2. $\mathbb{E}[e^X]$
3. $\operatorname{Var}(e^X)$

1.

$$\begin{aligned}
\mathbb{E}[|X|] &= \int_{\mathbb{R}} |x| \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\
&= \int_{-\infty}^{0} -x \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx + \int_{0}^{\infty} +x \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\
&= 2 \int_{0}^{\infty} x \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \qquad \text{since } x \longmapsto -xe^{-\frac{x^2}{2}} \text{ is an odd function} \\
&= \sqrt{\frac{2}{\pi}} \qquad \text{by change of variable } u = x^2
\end{aligned} \tag{156}$$

2.

$$\begin{aligned}
\mathbb{E}[e^X] &= \int_{\mathbb{R}} e^x \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\
&= \sqrt{e} \int_{\mathbb{R}} \frac{e^{-\frac{(x-1)^2}{2}}}{\sqrt{2\pi}} dx \text{ by completing the square in the exponential} \\
&= \sqrt{e} \qquad \text{since the integrand is a probability density.}
\end{aligned} \tag{157}$$

3.

$$\begin{aligned}
\operatorname{Var}(e^X) &= \mathbb{E}\left[\left(e^X\right)^2\right] - \mathbb{E}[e^X]^2 \\
&= e^2 - e \qquad \text{as above (completing the square and integral of pdf} = 1)
\end{aligned} \tag{158}$$

$$\begin{aligned}
\mathbb{E}[\mathrm{e}^{2X}] &= \int_{\mathbb{R}} \mathrm{e}^{2x} \frac{\mathrm{e}^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \,\mathrm{d}x = \int_{\mathbb{R}} \frac{\mathrm{e}^{-\frac{1}{2}[x^2-4x]}}{\sqrt{2\pi}} \,\mathrm{d}x \\
&= \int_{\mathbb{R}} \frac{\mathrm{e}^{-\frac{1}{2}[(x-2)^2-4]}}{\sqrt{2\pi}} \,\mathrm{d}x = \mathrm{e}^2 \int_{\mathbb{R}} \frac{\mathrm{e}^{-\frac{(x-2)^2}{2}}}{\sqrt{2\pi}} \,\mathrm{d}x = \mathrm{e}^2.
\end{aligned}$$

**Exercise 7.2** *(Dropping a CD on a parquet)* On laisse tomber au hasard un CD de diamètre $d$ sur un parquet dont les lames ont une largeur $D \geq d$. Calculer la probabilité que le CD tombe à cheval sur deux lames de parquet.

Let $X$ be the RV describing the distance between the center of the CD and the closest floorboard.

Then $X \sim \mathcal{U}\left(\left[0, \frac{D}{2}\right]\right)$. The CD falls between two floorboards iff $X \leq \frac{d}{2}$. The PDF of $X$ is $\rho_X(x) = \frac{2}{D}\mathbb{1}_{\left(0,\frac{D}{2}\right)}$, so

$$P\left(0 \leq X \leq \frac{d}{2}\right) = \int_0^{\frac{d}{2}} \frac{2}{D} dx = \frac{d}{D}. \tag{159}$$

# 8. Theory recap - 7.11.24

## 8.1. Moments

See [1] Section 21. Consider a random variable $X : \Omega \to \mathbb{R}$ with density $\rho_X$.

- The *absolute moments* of $X$ are

$$\mathbb{E}\left[|X|^k\right] = \int_{\mathbb{R}} |x|^k \rho_X(x) dx, \quad k = 1, 2, ... \tag{160}$$

- If $X$ has finite absolute moment of order $k$ then it has finite absolute moments of order $1, 2, ..., k-1$ as well
- For each $k$ for which $\mathbb{E}\left[|X|^k\right] < \infty$, the *$k$-th moment of $X$* is

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k \rho_X(x) dx, \quad k = 1, 2, ... \tag{161}$$

## 8.2. Extra: Conditional density and conditional expectation

- cf [2] (3.3.2 and 3.3.6) for general theory
- cf Section 5 in these notes for comparison with the discrete case
- Given $X : \Omega \to \mathbb{R}^{n_1}$ and $Y : \Omega \to \mathbb{R}^{n_2}$
- Consider $(X, Y) : \Omega \to \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$
- Assume $(X, Y)$ has the pdf $\rho_{X,Y}(x, y) : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}_{\geq 0}$
- Let $\rho_X : \mathbb{R}^{n_1} \to \mathbb{R}_{\geq 0}$ and $\rho_Y : \mathbb{R}^{n_2} \to \mathbb{R}_{\geq 0}$ be the correspinding marginal densities, that is

$$\rho_X(x) = \int_{\mathbb{R}^{n_2}} \rho_{X,Y}(x, y) dy \tag{162}$$

$$\rho_Y(y) = \int_{\mathbb{R}^{n_1}} \rho_{X,Y}(x, y) dx \tag{163}$$

**Definition 8.2.1** *(Conditional density)* For any $y \in \mathbb{R}^{n_2}$, the *conditional density of $X$ given $Y = y$* is the density $\rho_{X,y} : \mathbb{R}^{n_1} \to \mathbb{R}_{\geq 0}$ on $\mathbb{R}^{n_1}$ given by

$$\rho_{X,y}(x) := \frac{\rho_{X,Y}(x, y)}{\rho_Y(y)} \tag{164}$$

**Definition 8.2.2** For any $f : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \to \mathbb{R}$, the *conditional expectation of $f(X, Y)$ given $Y$* is the random variable $\mathbb{E}[f(X, Y) \mid Y] : \Omega \to \mathbb{R}$ given by

$$\mathbb{E}[f(X, Y) \mid Y] := \psi(Y) \tag{165}$$

where $\psi : \mathbb{R}^{n_2} \to \mathbb{R}$ is the function defined by

$$\psi(y) = \int_{\mathbb{R}^{n_1}} f(x, y) \rho_{X,y}(x) dx \tag{166}$$

## Exercises

**Exercise 8.1** *(Uniform distribution)*

**Exercice 37.** Soit $X$ une variable aléatoire de loi uniforme sur le segment $[a, b]$.

1. Tracer le graphe de la fonction de répartition de $X$. Calculer $\mathbb{E}[X]$ et $\mathbb{V}(X)$.
2. Sachant que $\mathbb{E}[X] = 2$ et $\mathbb{V}(X) = 2$, déterminer $a$ et $b$.
3. Soit $a = 1$, $b = 3$. Calculer $\mathbb{E}[|X^2 - 2|]$.

### Exercice 37

1. Un calcul direct donne $\mathbb{E}[X] = \dfrac{a+b}{2}$. Un calcul aussi direct donne que $\mathbb{E}[X^2] = \dfrac{b^3 - a^3}{3(b-a)} = \dfrac{a^2+ab+b^2}{3}$.
   Donc $\mathbb{V}(X) = \dfrac{a^2+ab+b^2}{3} - \dfrac{(a+b)^2}{4} = \dfrac{(b-a)^2}{12}$.

2. Les hypothèses nous amènent à $\begin{cases} a+b=4 \\ (b-a)^2 = 24 \end{cases} \Leftrightarrow \begin{cases} a = 4-b \\ 2b-4 = \sqrt{24} \end{cases} \Leftrightarrow \begin{cases} a = 2-\sqrt{6} \\ b = 2+\sqrt{6} \end{cases}$.

3. D'abord, remarquons que la fonction $x \mapsto x^2 - 2$ est strictement croissante sur $[1;3]$ et s'annule une unique fois en $x = \sqrt{2}$. Donc

$$\begin{aligned}
\mathbb{E}[|X^2-2|] &= \int_1^3 \frac{|x^2-2|}{2}\,\mathrm{d}x \\
&= \int_1^{\sqrt{2}} \frac{2-x^2}{2}\,\mathrm{d}x + \int_{\sqrt{2}}^3 \frac{x^2-2}{2}\,\mathrm{d}x \\
&= \sqrt{2} - 1 - \left[\frac{x^3}{6}\right]_1^{\sqrt{2}} + \left[\frac{x^3}{6}\right]_{\sqrt{2}}^3 - (3-\sqrt{2}) \\
&= \frac{4\sqrt{2}+2}{3}.
\end{aligned}$$

**Exercise 8.2** *(Moments, fonction muette, indicator method)*

**Exercice 38.** Pour $\alpha > 0$ et $x_0 > 0$, on considère la fonction $f$ suivante :

$$f(x) = \frac{\lambda}{x^{\alpha+1}} \mathbb{1}_{[x_0;+\infty[}(x).$$

1. Trouver $\lambda \in \mathbb{R}$ pour que $f$ soit une densité.
2. Quels moments pourra-t-on calculer pour cette variable aléatoire ?

On prend à partir de maintenant $\alpha = 3$ et $x_0 = 1$.

3. Calculer $\mathbb{E}[X]$ et $\mathbb{V}(X)$.
4. Calculer la fonction de répartition de $X$.
5. Calculer pour $a, x \in \mathbb{R}$, $\mathbb{P}\{X < x \mid X \geq a\}$.
6. On pose $Y = -b(X+1)$ où $b \in \mathbb{R}_+^*$. Calculer la fonction de répartition et la densité de $Y$.

### Exercice 38

1. Pour que $f$ soit une densité, il faut que $\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$. Donc

$$\begin{aligned}\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1 &\Leftrightarrow \lambda \int_{x_0}^{+\infty} \frac{1}{x^{\alpha+1}}\,\mathrm{d}x = 1 \\ &\Leftrightarrow \frac{\lambda}{\alpha x_0^{\alpha}} = 1 \\ &\Leftrightarrow \lambda = \alpha x_0^{\alpha}.\end{aligned}$$

2. On a que $\mathbb{E}[|X|^n] = \lambda \int_{x_0}^{+\infty} x^{n-\alpha-1}\,\mathrm{d}x$. Si $n = \alpha$, alors $\mathbb{E}[|X|^n] = +\infty$. Supposons donc que $n \neq \alpha$. Alors
$$\mathbb{E}[|X|^n] = \frac{\lambda}{n-\alpha}\left[x^{n-\alpha}\right]_{x_0}^{+\infty} = \begin{cases} +\infty \text{ si } n > \alpha \\ -\dfrac{\lambda x_0^{n-\alpha}}{n-\alpha} < +\infty \text{ si } n < \alpha \end{cases}.$$
Donc on peut calculer les moments d'ordre strictement inférieur à $\alpha$ de $X$.

3. Si $\alpha = 3$ et $x_0 = 1$, alors $\lambda = 3$. On a que $\mathbb{E}[X] = 3\int_1^{+\infty} \frac{1}{x^3}\,\mathrm{d}x = \frac{3}{2}$. De plus, $\mathbb{E}[X^2] = 3\int_1^{+\infty} \frac{1}{x^2}\,\mathrm{d}x = 3$. Donc $\mathbb{V}(X) = 3 - \frac{9}{4} = \frac{3}{4}$.

4. Un calcul direct donne que $F_X(t) = \left(1 - \frac{1}{t^3}\right)\mathbb{1}_{[1;+\infty[}(t)$.

5. On suppose que $x > a$ sinon la probabilité recherchée est nulle. Un calcul direct donne que $\mathbb{P}\{X < x \mid X \geq a\} = \dfrac{F_X(x) - F_X(a)}{1 - F_X(a)}$.
Si $a, x \geq 1$, alors $\mathbb{P}\{X < x \mid X \geq a\} = 1 - \left(\frac{a}{x}\right)^3$. Si $a, x < 1$, $\mathbb{P}\{X < x \mid X \geq a\} = 0$. Enfin, si $a < 1$ et $x \geq 1$, alors $\mathbb{P}\{X < x \mid X \geq a\} = 1 - \frac{1}{x^3}$.

6. Il est facile de voir que $F_Y(t) = 1 - F_X\left(-\frac{t+b}{b}\right) = \begin{cases} -\dfrac{b^3}{(t+b)^3} \text{ si } t \leq -2b \\ 1 \text{ sinon} \end{cases}$. Or, $F_Y$ est dérivable par morceaux et continue sur $\mathbb{R}$, donc on peut la dériver pour obtenir la densité de $Y$ qui est $f_Y(x) = \dfrac{3b^3}{(x+b)^4}\mathbb{1}_{]-\infty;-2b]}(x)$.

**Exercise 8.3** *(Inf and sup of uniform distributions)*

**Exercice 39.** Soient $X$ et $Y$ deux variables aléatoires indépendantes de loi uniforme sur l'intervalle [0,1]. Posons $U = \inf(X, Y)$ et $V = \sup(X, Y)$

1. Calculer la fonction de répartition et la densité de $V$.
2. Calculer la fonction de répartition et la densité de $U$.
3. Calculer $\mathbb{E}[U]$.
4. Que peut-on dire de $U + V$ (à part du bien) ? En déduire $\mathbb{E}[V]$.

## Exercice 39

1. On a que $F_V(t) = \mathbb{P}\{\sup(X, Y) \leq t\} = F_X(t)F_Y(t)$ par indépendance de $X$ et $Y$. Or, $X$ et $Y$ ont même loi, donc $F_V(t) = F_X(t)^2$, avec $F_X(t) = t\mathbb{1}_{]0;1[}(t) + \mathbb{1}_{[1;+\infty[}(t)$. Donc $F_V(t) = t^2\mathbb{1}_{]0;1[}(t) + \mathbb{1}_{[1;+\infty[}(t)$. C'est une fonction continue sur $\mathbb{R}$ et dérivable par morceaux, donc la densité de $V$ est $f_V(x) = 2x\mathbb{1}_{]0;1[}(x)$.
2. On a que $F_U(t) = \mathbb{P}\{\inf(X, Y) \leq t\} = 1 - \mathbb{P}\{\inf(X, Y) > t\} = 1 - (1 - F_X(t))^2 = 2F_X(t) - F_X(t)^2$. Donc il est facile de calculer que $F_U(t) = (2t - t^2)\mathbb{1}_{]0;1[}(t) + \mathbb{1}_{[1;+\infty[}(t)$. C'est une fonction continue sur $\mathbb{R}$ et dérivable par morceaux. Donc on obtient la densité de $U$ en dérivant $F_U$ et on a $f_U(x) = (2 - 2x)\mathbb{1}_{]0;1[}(x)$.
3. On a que $\mathbb{E}[U] = \displaystyle\int_0^1 x(2 - 2x)\,\mathrm{d}x = \frac{1}{3}$.
4. Il est facile de voir que $U + V = \inf(X, Y) + \sup(X, Y) = X + Y$. Donc $\mathbb{E}[U + V] = \mathbb{E}[X + Y] = 2\mathbb{E}[X]$ car $X$ et $Y$ ont même loi. Or, $\mathbb{E}[X] = \frac{1}{2}$. Donc $\mathbb{E}[U + V] = 1 = \mathbb{E}[U] + \mathbb{E}[V]$. Donc $\mathbb{E}[V] = 1 - \frac{1}{3} = \frac{2}{3}$.

# Bibliography


[1] P. Billingsley, *Probability and Measure*. John Wiley & Sons, 2012.

[2] B. Jourdain, *Probabilités et statistique pour l'ingénieur*. 2018.

[3] E. H. Lieb and M. Loss, *Analysis*, 2nd ed. in Graduate Studies in Mathematics 14. American Mathematical Society, 2001.