Probabilité et Simulation

PolyTech INFO4 (Grenoble) – TD

Last updated: 2025-10-16

Contact: davide.legacci@univ-grenoble-alpes.fr

References

- Fundamentals: [1] B. Jourdain, *Probabilités et statistique pour l'ingénieur*. 2018.
- Further reading: [2] P. Billingsley, *Probability and Measure*. John Wiley & Sons, 2012.

Websites

- CM: https://github.com/jonatha-anselmi/INFO4-PS
- TD: https://github.com/davidelegacci/probasim24

Contents

1. Theory recap - Probability on finite set

Reminder: a set is

- *finite* (*fini*), if it has a finite number of elements
- *countable* (*dénombrable*), if either it is finite, or it can be made in one to one correspondence with the set of natural numbers

1.1. Finite probability space

- *Random process*: one random outcome out of finitely many
- *Sample space* $\Omega =$ *finite* set of possible *outcomes* ω
- *Probability* on $\Omega =$ set of weights $\mathbb{P}(\omega) \in \mathbb{R}$ on each $\omega \in \Omega$ such that
  - $\mathbb{P}(\omega) > 0 \forall \omega \in \Omega$
  - $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$

- *Event* $A \subseteq \Omega =$ any subset of the sample space

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega) \tag{1}$$

- *Complementary event* $A^c = \Omega / A$ (pronounced "*not* A")
- "A *and* B" $= A \cap B$
- "A *or* B" $= A \cup B$

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

- Probability of complementary event

$$\Omega = A^c \sqcup A \Rightarrow 1 = \mathbb{P}(A^c) + \mathbb{P}(A) \tag{3}$$

- *Indicator function* of event $A \subseteq \Omega$

$$\mathbb{1}_A : \Omega \to \mathbb{R}, \quad \mathbb{1}_A(\omega) = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{cases} \tag{4}$$

- Denoted by $|S|$ the *cardinality* of a set S
- Denoted by S^n the *cartesian product* of S with itself n times

$$S^n = S \times S \times ... \times S = \{(s_1, ..., s_n) : s_i \in S \text{ for all } i = 1, ..., n\} \tag{5}$$

- *Cardinality of cartesian product*

$$|S^n| = |S|^n \tag{6}$$

1.2. Uniform probability

- *Every outcome* $\omega \in \Omega$ *has the same weight*

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|} \tag{7}$$

- *Uniform probability of the event* $A \subseteq \Omega$

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|} \tag{8}$$

1.3. Counting

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

$$P_k = k! \tag{9}$$

Permutation of 5 elements

1	5	4	3	2

- 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 : $\{1, ..., k\} \to \{1, ..., n\}$

$$D_{n,k} = \underbrace{n(n-1)...}_{k \text{ times}} = n(n-1)...(n-k+1) = \frac{n!}{(n-k)!} \tag{10}$$

Disposition of 3 elements out of 5

	2	1	3	

- 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

$$C_{n,k} = \frac{D_{n,k}}{P_k} = \frac{n!}{k!(n-k)!} = \binom{n}{k} = \text{choose}(n, k) \tag{11}$$

Combination of 3 elements out of 5

		X	X	X

- *Binomial theorem*

$$(x + y)^n = \sum_{j=0}^{n} \binom{n}{j} x^{n-j} y^j \tag{12}$$

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.

$$\text{number of kisses} = \frac{f(f-1)}{2} + fg \tag{13}$$

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?

General strategy

- Identify sample space Ω (write in set-theoretic notation!) and its cardinality $|\Omega|$
- Identify event $A \subseteq \Omega$ (write in set-theoretic notation!) and its cardinality $|A|$
- Uniform probability? If so, use $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$

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

$$\Omega = \{\omega = (\omega_1, ..., \omega_n) : \omega_i \in \mathcal{F} \text{ for all } i \in \mathcal{N}\} = \mathcal{F}^n \tag{14}$$

with cardinality

$$|\Omega| = |\mathcal{F}^n| = |\mathcal{F}|^n = f^n \tag{15}$$

Endow the sample space with the uniform probability (since every outcome of the experiment is equiprobable).

The event we're looking at is

$$A = \left\{\omega \in \Omega : \omega_i \neq \omega_j \text{ for all } i \neq j \in \mathcal{N}\right\} \tag{16}$$

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

$$\begin{aligned} |A| &= \# \text{ of ways to choose and order } n \text{ elements out of } f \\ &= \underbrace{f(f-1)...}_{n} = f(f-1)...(f-n+1) = \frac{f!}{(f-n)!} \end{aligned} \tag{17}$$

$$\mathbb{P}(A) = \frac{f!}{f^n(f-n)!} \tag{18}$$

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

$$
\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}\left(\underbrace{\text{no two people have the same birthday}}_{\text{Ex. 2}}\right) \\
&= 1 - \frac{y!}{y^n(y-n)!}
\end{aligned} \tag{19}
$$

Formal solution Let $\mathcal{N} = \{1, \dots, n\}$ and $\mathcal{Y} = \{1, \dots, y\}$ with $n \leq y$. The sample space is

$$
\begin{aligned}
\Omega &= \text{distributions of possible birthdays of } n \text{ people} \\
&= \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \mathcal{Y} \text{ for all } i \in \mathcal{N}\} = \mathcal{Y}^n
\end{aligned} \tag{20}
$$

where ω_i is the birthday of the i-th person. The cardinality of the sample space is

$$
|\Omega| = |\mathcal{Y}^n| = |\mathcal{Y}|^n = y^n \tag{21}
$$

The event we're looking at is

$$
A = \left\{\omega \in \Omega : \exists i \neq j \in \mathcal{N} : \omega_i = \omega_j\right\} \tag{22}
$$

Note that this is the complementary event to the event defined in Equation 16 of Exercise 2. Thus we can compute its probability as

$$
\mathbb{P}(A) = 1 - \mathbb{P}(A^c) \tag{23}
$$

in agreement with Equation 19.

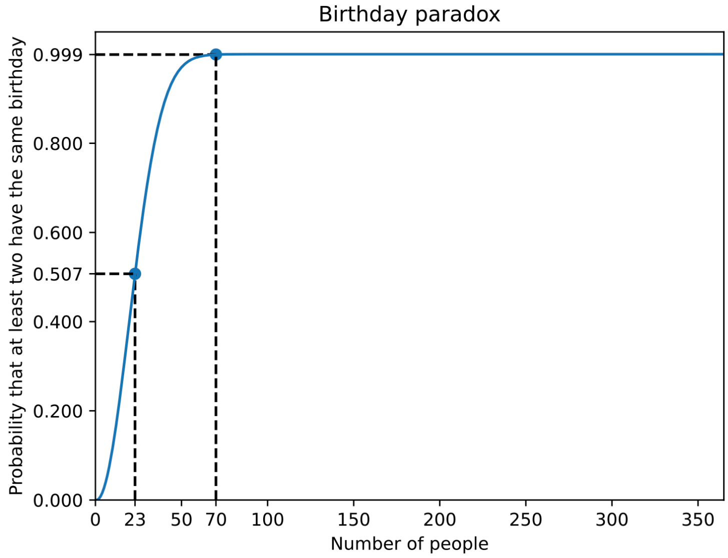

Figure 4: Birthday paradox probability. Code available.

> **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

$$
\begin{aligned}
\mathbb{P}(A) &= 1 - \mathbb{P}\left(\underbrace{\text{nobody has the prescribed birth date}}\right) \\
&= 1 - \left(\frac{y-1}{y}\right)^n
\end{aligned} \tag{24}
$$

Formal solution 1 As above $\mathcal{N} = \{1, ..., n\}$ and $\mathcal{Y} = \{1, ..., y\}$ with $n \leq y$. The sample space 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

$$A = \{\omega \in \Omega : \exists i \in \mathcal{N} : \omega_i = \omega_{n+1}\} \tag{25}$$

with complementary event

$$A^c = \{\omega \in \Omega : \omega_i \neq \omega_{n+1} \forall i \in \mathcal{N}\} \tag{26}$$

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

$$|A^c| = \underbrace{y}_{\text{prof}} \cdot \underbrace{(y-1)^n}_{\text{students}} \tag{27}$$

So, $\mathbb{P}(A) = 1 - \frac{y(y-1)^n}{y^{n+1}} = 1 - \left(\frac{y-1}{y}\right)^n$, in agreement with Equation 24.

Note that a factor $\frac{y}{y}$, corresponding to the prof's birthday, simplifies in the last step. Alternatively, you can fix the birthday of the prof and exclude it from the analysis from the beginning. In this case $|\Omega| = y^n$ and $|A^c| = (y-1)^n$, leading to the same result.

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

$$A = \sqcup_{j \in \mathcal{N}} A_j \tag{28}$$

with probability (cf Equation 2)

$$\mathbb{P}(A) = \sum_{j \in \mathcal{N}} \mathbb{P}(A_j) = \frac{\sum_{j \in \mathcal{N}} |A_j|}{|\Omega|} \tag{29}$$

The cardinality of A_j is

$$\begin{aligned} |A_j| &= \underbrace{1...1}_{j \text{ times}} \cdot \underbrace{(y-1)...(y-1)}_{n-j \text{ times}} \cdot \underbrace{y}_{\text{prof}} \cdot \underbrace{\binom{n}{j}}_{\text{number of ways to choose } j \text{ elements out of } n} \\ &= y(y-1)^{n-j}\binom{n}{j} \end{aligned} \tag{30}$$

By an application of the binomial theorem (Equation 12) and a short manipulation,

$$\sum_{j=1}^{n} |A_j| = y(y^n - (y-1)^n) \tag{31}$$

which leads back to Equation 24.

2. Theory recap - Conditional probability and independence

2.1. Conditional probability

- Let $\mathbb{P}$ be a probability on Ω, and consider the events $A, B \subseteq \Omega$.
- *Conditional probability*: probability of A *given* B

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \text{ if } \mathbb{P}(B) \neq 0 \tag{32}$$

- not reallty defined if $\mathbb{P}(B) = 0$, cf [2] pag. 427.
- often used as

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B) \tag{33}$$

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

$$\mathbb{P}(A \mid B) + \mathbb{P}(A^c \mid B) = 1 \tag{34}$$

- *Total Probability Theorem*

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c) \tag{35}$$

- *Bayes theorem*

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap A) \Rightarrow \mathbb{P}(A \mid B)\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A) \tag{36}$$

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

2.2. Independent events

Let Ω be equipped with a probability $\mathbb{P}$.

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \tag{37}$$

- equivalently, by definition of conditional probability, A and B are independent if

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \tag{38}$$

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

$$\mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i) \text{ for all } I \subseteq \{1, \ldots, n\} \tag{39}$$

- 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 5: 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,

$$\mathbb{P}(A_2 \cap A_1) = \underbrace{\mathbb{P}(A_2 \mid A_1)}_{\frac{5}{9}} \underbrace{\mathbb{P}(A_1)}_{\frac{6}{10}} = \frac{1}{3}. \tag{40}$$

Solution 2 The sample space 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

$$\mathbb{P}(A) = \frac{\binom{6}{2}}{\binom{10}{2}} = \frac{6 \cdot 5}{10 \cdot 9} = \frac{1}{3}. \tag{41}$$

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 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 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

$$\Omega = \{R_h \cap R_o, R_h \cap W_o, W_h \cap R_o, W_h \cap W_o\}. \tag{42}$$

For example, $W_h \cap R_o$ is the event "*the hidden face is white and the observed face is red*". We are given the information that the observed face is red, so we need to find the probability that the hidden face has a certain color *given* the fact that the observed face is red. In other words, we need to find the probabilities

$$\mathbb{P}(W_h \mid R_o) \text{ and } \mathbb{P}(R_h \mid R_o). \tag{43}$$

Clearly it suffices to find one of them, since by Equation 34 they sum to 1. By definition of conditional probability (Equation 32)

$$\mathbb{P}(R_h \mid R_o) = \frac{\mathbb{P}(R_h \cap R_o)}{\mathbb{P}(R_o)}. \tag{44}$$

The numerato of Equation 44 corresponds to the probability of drawing the red card (event denoted by r), which is $\frac{1}{3}$:

$$
\begin{aligned}
\mathbb{P}(R_h \cap R_o) &= \mathbb{P}(R_o \cap r) \\
&= \underbrace{P(R_o \mid r)}_{1} \underbrace{P(r)}_{\frac{1}{3}} = \frac{1}{3}.
\end{aligned}
\tag{45}
$$

The denominator of Equation 44 is the probability of drawing the red card, plus the probability of drawing the mixed card (event denoted by m) and observing the red face, that is $\frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{2}$. This follows formally form the Total Probability Theorem (Equation 35) and the definition of conditional probability (Equation 32):

$$
\begin{aligned}
\mathbb{P}(R_o) &= \mathbb{P}(R_o \cap R_h) + \mathbb{P}(R_o \cap W_h) \\
&= \frac{1}{3} + \qquad \mathbb{P}(R_o \cap m) \\
&= \frac{1}{3} + \qquad \underbrace{P(R_o \mid m)}_{\frac{1}{2}} \underbrace{P(m)}_{\frac{1}{3}} = \frac{1}{2}
\end{aligned}
\tag{46}
$$

In conclusion, given that the observed face is red, the hidden face is red with probability $2/3$:

$$
\mathbb{P}(R_h \mid R_o) = \frac{1/3}{1/2} = \frac{2}{3}. \tag{47}
$$

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 $x + 1 \bmod 6$:

$$
\Omega = \{12, 23, 34, 45, 56, 61\} \tag{48}
$$

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

$$
\mathrm{die}_1 = \{12, 61\} \tag{49}
$$

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

$$
\mathbb{P}(\mathrm{die}_2 \mid \mathrm{live}_1) = \frac{\mathbb{P}(\mathrm{die}_2 \cap \mathrm{live}_1)}{\mathbb{P}(\mathrm{live}_1)} \tag{50}
$$

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

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

3.1. Probability measure

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

Definition 3.1.1 *(sigma-field)* Let Ω be any set. A *σ-field $\mathcal{A}$* on Ω is a collection of subsets of Ω 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 Ω that is in $\mathcal{A}$ is called *event*.

Definition 3.1.2 *(Measure)* Given a set Ω and a σ-algebra $\mathcal{A}$ on Ω, a *measure* μ is a function

$$\mu : \mathcal{A} \to \mathbb{R}_{\geq 0} \tag{51}$$

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

$$\mu\left(\bigsqcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{52}$$

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

$$\mathbb{P}(\Omega) = 1. \tag{53}$$

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

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

3.2. Discrete random variables

Definition 3.2.1 *(Discrete random variable)* Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, a *discrete random variable* X is a function $X : \Omega \to F$ such that
1. F is a *countable* set; we call it the space of *values* of the random variable, and we say that any $x \in F$ is a *value*.
2. *the level sets of X are events*, that is

$$\{X = x\} := \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{A} \text{ for all } x \in F \tag{54}$$

- clearly, $\{X = x\} = \emptyset \in \mathcal{A}$ for all $x \in F \setminus \mathrm{Im}(X)$
- property 2. guarantees that $\mathbb{P}\{X = x\}$ 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 $\mu_X : F \to [0, 1]$ defined by

$$\mu_X(x) = \mathbb{P}\{X = x\} \text{ for all } x \in F. \tag{55}$$

In words: for any value, the law of a random variable returns the probability that the random variable takes the given value.

[1]In french, this set is called *tribu* on Ω. The term *σ-algebra* is also used – and is more common in the context of pure analysis, c.f. [3] – whereas the term *σ-field* is more common in the context of probabioloty theory, c.f. [2].

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

$$\mathbb{P}\{X = x, Y = y\} = \mathbb{P}\{X = x\}\mathbb{P}\{Y = y\} \text{ for all } x \in F, y \in G \tag{56}$$

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

$$\{\omega \in \Omega : X(\omega) = x\} \cap \{\omega \in \Omega : Y(\omega) = y\} \in \mathcal{A}. \tag{57}$$

- the definition generalises to collections of DRVs, see Section 2.2.3 in [1].

> **Tip** A discrete random variable is a *function on* Ω with *countably many values*. Think of it as an experiment with countably many random outcomes. Its *law, or distribution*, gives the probability to observe each of the possible (countable) values of the random variable.

> **Tip**
> - $(\Omega, \mathcal{A}, \mathbb{P})$ with $\mathbb{P} : \mathcal{A} \to [0, 1]$ and $\mathbb{P}(\Omega) = 1$
> - $X : \Omega \to F$ countable values space, with $\{X = x\} \in \mathcal{A}$ is an event for all $x \in F$
> - $\mu_X : F \to [0, 1]$ such that $\mu_X(x) = \mathbb{P}\{X = x\}$ is the distribution of 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 $p \in [0, 1]$, the probability to observe 1. We then have

$$\begin{aligned} \mu : \{0, 1\} &\to [0, 1] \\ 1 &\longmapsto \quad p \\ 0 &\longmapsto \quad 1 - p \\ x &\longmapsto \quad p^x (1 - p)^{1-x} \end{aligned} \tag{58}$$

- When a random variable X follows the Bernoulli distribution we write $X \sim \mathcal{B}(p)$.

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

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

$$\begin{aligned} \mu : F &\to \ [0, 1] \\ k &\longmapsto \binom{n}{k} p^k (1 - p)^{n-k} \end{aligned} \tag{59}$$

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

$$\begin{aligned} \mu : F &\to \ [0, 1] \\ n &\longmapsto e^{-\lambda} \frac{\lambda^n}{n!} \end{aligned} \tag{60}$$

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

$$\begin{aligned} \mu : F &\to [0,1] \\ k &\longmapsto p(1-p)^{k-1} \end{aligned} \tag{61}$$

3.4. Useful stuff

- Vandermonde's identity

$$\sum_{k_1=0}^{k} \binom{n_1}{k_1} \binom{n_2}{k-k_1} = \binom{n_1+n_2}{k} \tag{62}$$

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 [1] pag. 16.

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

$$\mu_i(k_i) = \mathbb{P}(X_i = k_i) = \binom{n_i}{k_i} p^{k_i} (1-p)^{n_i - k_i} \tag{63}$$

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

$$\begin{aligned} \mu(k) &= \mathbb{P}(X_1 + X_2 = k) \\ &= \mathbb{P}\left(\bigsqcup_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \{X_1 = k_1, X_2 = k_2\} \right) \\ &= \sum_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \mathbb{P}(X_1 = k_1)\mathbb{P}(X_2 = k_2) && \text{by c. add and indep.} \\ &= \sum_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \mu(k_1)\mu(k_2) && \text{by def of law} \\ &= \sum_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \binom{n_1}{k_1}\binom{n_2}{k_2} p^{k_1+k_2}(1-p)^{n_1+n_2-k_1-k_2} \\ &= p^k (1-p)^{n_1+n_2-k} \sum_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \binom{n_1}{k_1}\binom{n_2}{k_2} \end{aligned} \tag{64}$$

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 ≥ 0, k_1 can be at most equal to k. So the constraints

$$\begin{aligned} &k_1 \in \{0, ..., n_1\} \\ &k_2 \in \{0, ..., n_2\} \\ &k_1 + k_2 = k \end{aligned} \tag{65}$$

can be replaced by the constraints

$$\begin{aligned} &k_1 \in \{0, ..., k\} \\ &k_2 = k - k_1 \end{aligned} \tag{66}$$

namely

$$\sum_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \binom{n_1}{k_1}\binom{n_2}{k_2} = \sum_{k_1=0}^{k} \binom{n_1}{k_1}\binom{n_2}{k - k_1} = \binom{n_1 + n_2}{k} \tag{67}$$

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 67.
- If $k > n_1$, we have

$$\begin{aligned} \sum_{\substack{k_i \in F_i \\ k_1 + k_2 = k}} \binom{n_1}{k_1}\binom{n_2}{k_2} &= \sum_{k_1=0}^{n_1} \binom{n_1}{k_1}\binom{n_2}{k - k_1} \\ &= \sum_{k_1=0}^{n_1} \binom{n_1}{k_1}\binom{n_2}{k - k_1} + \sum_{k_1=n_1+1}^{k} \binom{n_1}{k_1}\binom{n_2}{k - k_1} \end{aligned} \tag{68}$$

since each summand in the the second sum is zero[2], and we get again Equation 67.

So in conclusion

$$\mu(k) = \binom{n_1 + n_2}{k} p^k (1 - p)^{n_1 + n_2 - k} \tag{69}$$

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

$$\begin{aligned} \mu(n) = \mathbb{P}(X_1 + X_2 = n) &= \sum_{\substack{n_i \in \mathbb{N} \\ n_1 + n_2 = n}} \mu_1(n_1)\mu_2(n_2) \\ &= \sum_{\substack{n_i \in \mathbb{N} \\ n_1 + n_2 = n}} e^{-\lambda_1} \frac{\lambda_1^{n_1}}{n_1!} e^{-\lambda_2} \frac{\lambda_2^{n_2}}{n_2!} \end{aligned} \tag{70}$$

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 $b > a$.

$$
\begin{aligned}
\mu(n) &= e^{-(\lambda_1+\lambda_2)} \sum_{n_1=0}^{n} \frac{\lambda_1^{n_1}}{n_1!} \frac{\lambda_2^{n-n_1}}{n-n_1!} \cdot \frac{n!}{n!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{n_1=0}^{n} \binom{n}{n_1} \lambda_1^{n_1} \lambda_2^{n-n_1} \\
&= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}
\end{aligned} \tag{71}
$$

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. Recall that the geometric series $\sum_{i=0}^{\infty} r^i$ with $r \in [0, 1)$ converges to $\frac{1}{1-r}$.

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

$$
\begin{aligned}
\mu(k) &= \mathbb{P}(Z = k) \\
&= \mathbb{P}(Z \geq k) - \mathbb{P}(Z \geq k+1) \\
&= \mathbb{P}(X_1 \geq k, X_2 \geq k) - \mathbb{P}(X_1 \geq k+1, X_2 \geq k+1) \\
&= \mathbb{P}(X_1 \geq k)\mathbb{P}(X_2 \geq k) - \mathbb{P}(X_1 \geq k+1)\mathbb{P}(X_2 \geq k+1)
\end{aligned} \tag{72}
$$

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

$$
\begin{aligned}
\mathbb{P}(X \geq k) &= \mathbb{P}\left(\bigsqcup_{i \geq k} (X = i)\right) \\
&= \sum_{i \geq k} \mathbb{P}(X = i) \\
&= \sum_{i \geq k} p(1-p)^{i-1} \\
&= p(1-p)^{k-1} + p(1-p)^k + p(1-p)^{k+1} + \dots \\
&= p(1-p)^{k-1}\left(1 + (1-p) + (1-p)^2 + \dots\right) \\
&= p(1-p)^{k-1} \sum_{j=0}^{\infty} (1-p)^j \\
&= (1-p)^{k-1}
\end{aligned} \tag{73}
$$

Plugging in Equation 72 we get

$$
\begin{aligned}
\mu(k) &= \mathbb{P}(X_1 \geq k)\mathbb{P}(X_2 \geq k) - \mathbb{P}(X_1 \geq k+1)\mathbb{P}(X_2 \geq k+1) \\
&= (1-p_1)^{k-1}(1-p_2)^{k-1} - (1-p_1)^k(1-p_2)^k \\
&= (1-p_1)^{k-1}(1-p_2)^{k-1}[1 - (1-p_1)(1-p_2)]
\end{aligned} \tag{74}
$$

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. $Z \sim \mathcal{G}(\beta)$.

Exercise 3.4 *(*Lost messages)* If there is extra time, do <u>Exercise 4.1</u>, else do in next session.

4. Theory recap - Expectation and variance

Recall from last week that

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

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

$$\mathbb{E}(X) := \sum_{x \in F} x \mathbb{P}(X = x). \tag{75}$$

Proposition 4.1.2 *(Linearity of expectation)*

$$\mathbb{E}(X + aY) = \mathbb{E}(X) + a\mathbb{E}(Y) \tag{76}$$

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

- $\mathbb{E}(\text{constant}) = \text{constant}$
- Sufficient condition, positivity, monotonicity: see [1] 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 / Théorème de transfert, ou de transport)* Let $X : \Omega \to F \subset \mathbb{R}$ be RV, and consider some function $f : F \to \mathbb{R}$. Then

$$\mathbb{E}(f(X)) = \sum_{x \in F} f(x) P(X = x) \tag{77}$$

whenever defined (see [1] Th. 2.3.6 for details).

Proposition 4.1.4 *(Expectation and independence)* Let X, Y be independent 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

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)) \tag{78}$$

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 $\sum_{x \in F} x^2 \mathbb{P}(X = x) < +\infty$, and in this case

$$\mathrm{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \tag{79}$$

- If X is square integrable then the variance is well defined, cf [1] 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)*

$$\operatorname{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \tag{80}$$

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

Lemma 4.2.3 *(Variance after scaling and shifting)*

$$\operatorname{Var}(aX + b) = a^2 \operatorname{Var}(X) \text{ for all } a, b \in \mathbb{R} \tag{81}$$

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

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

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i) \tag{82}$$

Some **examples**:
- $X \sim \mathcal{B}(p) \Rightarrow \operatorname{Var}(X) = p(1-p)$, see Exercise 4.3 and proof.
- $X \sim \mathcal{B}(n,p) \Rightarrow \operatorname{Var}(X) = np(1-p)$ (immediate by Proposition 4.2.4)
- $X \sim \mathcal{P}(\lambda) \Rightarrow \operatorname{Var}(X) = \lambda$, 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 $\{0, 1, 2, \dots\} \ni l$ contains the possible numbers l of lost messages in one time unit. The probability $\mathbb{P}(L = l)$ 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 $i \in \{1, 2, \dots\}$.

Let $L_a = \sum_{i=1}^{a} X_i$ be the DRV whose range $\operatorname{Im}(L_a) = \{0, 1, \dots, a\} \ni l$ 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:

$$L_a \sim \mathcal{B}(a, p). \tag{83}$$

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

The law of L is given by

$$\mathbb{P}(L = l) = \mathbb{P}\left[\bigsqcup_{a=l}^{\infty} \{L_a = l \cap A = a\}\right], \tag{84}$$

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,

$$\mathbb{P}(L = l) = \sum_{a=l}^{\infty} \mathbb{P}(L_a = l)\mathbb{P}(A = a) \tag{85}$$

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

$$\begin{aligned} \mathbb{P}(L = l) &= \sum_{a=l}^{\infty} \binom{a}{l} p^l (1-p)^{a-l} e^{-\lambda} \frac{\lambda^a}{a!} \cdot \frac{\lambda^l}{\lambda^l} \\ &= \lambda^l p^l \frac{1}{l!} e^{-\lambda} \sum_{a=l}^{\infty} \frac{1}{(a-l)!} (1-p)^{a-l} \lambda^{a-l} \\ &= \lambda^l p^l \frac{1}{l!} e^{-\lambda} \sum_{j=0}^{\infty} \frac{(\lambda - \lambda p)^j}{j!} \\ &= \lambda^l p^l \frac{1}{l!} e^{-\lambda} e^{\lambda - \lambda p} \\ &= \frac{(\lambda p)^l}{l!} e^{-\lambda p}. \end{aligned} \tag{86}$$

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

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \frac{1}{n+1} e^{-\lambda} \frac{\lambda^n}{n!} \tag{87}$$

Multiply and divide by λ 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:

$$\mathbb{E}(X_i) = 1 \cdot p + 0 \cdot (1-p) = p \tag{88}$$

$$\mathbb{E}(X_1 + ... + X_n) = \mathbb{E}(X_1) + ... + \mathbb{E}(X_n) = np \tag{89}$$

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 80:

$$\mathbb{E}(X_i^2) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p \tag{90}$$

$$\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 = p(1 - p) \tag{91}$$

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

$$\mathbb{E}(Y = Y_1 + ... + Y_n) = n[gp - l(1 - p)] \tag{92}$$

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

$$\frac{g}{l} = \frac{1 - p}{p}. \tag{93}$$

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)* If time allows, do Exercise 5.1, else do in next session.

5. Theory recap - 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 $\mathbb{E}[X] = \sum_{x \in F} x\mathbb{P}[X = x]$;
- the *law* of X is the function $F \to [0, 1]$ defined by $x \longmapsto \mathbb{P}[X = x]$.

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 the *product random variable* (X, Y) with values in $F \times G$, unless we have further information on how X and Y are related. Let $(X = x \cap Y = y) \equiv \{X = x\} \cap \{Y = y\}$ denote the intersection of two events, so that (this is just notation)

$$\mathbb{P}[(X, Y) = (x, y)] \equiv \mathbb{P}[\{X = x\} \cap \{Y = y\}] \equiv \mathbb{P}[X = x \cap Y = y]. \tag{94}$$

Then

- If X and Y are independent, then $\mathbb{P}[(X, Y) = (x, y)] = \mathbb{P}[X = x]\mathbb{P}[Y = y]$, namely the law of the product random variable (X, Y) is the product of the laws of the variables X and Y.
- Conversely, knowing the law of (X, Y) one can recover the laws of X and Y by the *marginal law*:

$$\mathbb{P}[X = x] = \sum_{y \in G} \mathbb{P}[X = x \cap Y = y]. \tag{95}$$

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

$$\mathbb{P}[X = x \cap Y = y] = c\mu(x)\nu(y) \text{ for all } x \in F, y \in G, \tag{96}$$

then X and Y are independent (cf. [1] Rem. 2.2.12.)

One can play with these concepts, looking at the behavior of the expected value of functions of random variables for a fixed value of one ofd them.

5.2. Conditional law

- The random variable $(X \mid Y = y)$ with values in F for some fixex $y \in G$ has the following law:

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

$$x \longmapsto \mathbb{P}[X = x \mid Y = y] = \frac{\mathbb{P}[X = x \cap Y = y]}{\mathbb{P}[Y = y]} \tag{97}$$

where we use the definition of conditional probability (Equation 32).

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).
- It's expected value is

$$\begin{aligned} \mathbb{E}[f(X, Y)] &= \sum_{x \in F} \sum_{y \in G} f(x, y)\mathbb{P}[X = x \cap Y = y] \\ &= \sum_{y \in G} \left(\sum_{x \in F} f(x, y)\mathbb{P}[X = x \mid Y = y] \right) \mathbb{P}[Y = y] \end{aligned} \tag{98}$$

The term in brackets is a function $G \to \mathbb{R}$. Call it ψ. So we have $\psi : G \to \mathbb{R}$,

$$\psi(y) = \sum_{x \in F} f(x, y)\mathbb{P}[X = x \mid Y = y] \tag{99}$$

By the Transfert Theorem, Equation 98 is the expected value of the random variable $\psi(Y)$. We call this random variable *conditional expectation of* $f(X, Y)$ *given* Y.

Definition 5.3.1 The *conditional expectation of* $f(X, Y)$ *given* Y is the **discrete random variable** $\psi(Y)$. It is denoted by $\mathbb{E}[f(X, Y) \mid Y] := \psi(Y)$.

Special cases

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

$$\mathbb{E}[X \mid Y] = \psi(Y) \text{ with } \psi(y) = \sum_{x \in F} x\mathbb{P}[X = x \mid Y = y]. \tag{100}$$

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

$$\psi(y) = \sum_{x \in F} f(x, y)\mathbb{P}[X = x] = \mathbb{E}[f(X, y)]. \tag{101}$$

Expected value As mentioned above, by Equation 98 and Equation 99,

$$\mathbb{E}[\psi(Y)] = \mathbb{E}[f(X,Y)] \quad (102)$$

whenever all the RVs involved are integrable. More generally, we have the following:

- Let $X : \Omega \to F$ and $Y : \Omega \to G$ be random variables
- consider functions $f : F \times G \to \mathbb{R}$ and $g : G \to \mathbb{R}$
- consider $\psi(Y)$ as above

Proposition 5.3.2

$$\mathbb{E}[\psi(Y)g(Y)] = \mathbb{E}[f(X,Y)g(Y)] \quad (103)$$

whenever all the RVs involved are integrable.

By setting $g \equiv 1$ one recovers Equation 102. See Proposition 2.5.6 in [1].

Tip
- Let $X : \Omega \to F$ and $Y : \Omega \to G$ be random variables.
- The law of X is $x \longmapsto \mathbb{P}[X = x]$
- The expectation of X is $\mathbb{E}[X] = \sum_{x \in F} x\mathbb{P}[X = x]$

Tip
- The random variable $(X \mid Y = y)$ is valued in F.
- Its law (called *conditional law of X given* $(Y = y)$) is by defintion $x \longmapsto \mathbb{P}[X = x \mid Y = y]$

Tip
- Define $\psi : G \to \mathbb{R}$ by Equation 100: $\psi(y) = \sum_{x \in F} x\mathbb{P}[X = x \mid Y = y]$
- The *conditional expectation of X given Y* is the random variable $\psi(Y)$ valued in $\mathbb{R}$.

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 $\mathbb{P}(X = 0) = q$ and $\mathbb{P}(X = 1) = p$ 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

$$B_n = \sum_{k=1}^{n} b_k = \sum_{k=1}^{n} b_1 2^{k-1} = b_1 \sum_{l=0}^{n-1} 2^l = b_1 \frac{1-2^n}{1-2} = b_1(2^n - 1). \tag{104}$$

by the formula for the sum of the first $n-1$ terms of a geometric series. 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

$$2b_n - B_n = 2^n b_1 - (2^n - 1)b_1 = b_1. \tag{105}$$

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 $F = \{1, 2, 3, \ldots\} \ni n$. 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 104. By Theorem 4.1.3 and Equation 61 the expected amount of money bet is

$$\mathbb{E}(B(X)) = \sum_{n=1}^{\infty} B_n \mathbb{P}(X = n) = \sum_{n=1}^{\infty} B_n (1-p)^{n-1} p \tag{106}$$

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

$$\begin{aligned}\mathbb{E}(B(X)) = b_1 \quad & \mathbb{P}(X_1 = 1) + \\ +(b_1 + b_2) \quad & \mathbb{P}(X_1 = 0, X_2 = 1) + \\ +(b_1 + b_2 + b_3) & \mathbb{P}(X_1 = 0, X_2 = 0, X_3 = 1) + \\ +\ldots + & \\ +(b_1 + \ldots + b_n) & \mathbb{P}(X_1 = 0, \ldots, X_{n-1} = 0, X_n = 1) + \ldots\end{aligned} \tag{107}$$

Then

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

which is equivalent to Equation 106.

---

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

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:

$$\text{Finite expected bet amount} \Leftrightarrow \text{prob. to win } p > \frac{1}{2}. \tag{109}$$

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

$$\begin{aligned} \mathbb{E}(B(X)) &= b_1 \frac{p}{q} \sum_{n=1}^{\infty} [(2q)^n - q^n] \\ &= b_1 \frac{p}{q} \left[\frac{2q}{1-2q} - \frac{q}{1-q} \right] \\ &= \frac{b_1}{2p-1}. \end{aligned} \tag{110}$$

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:

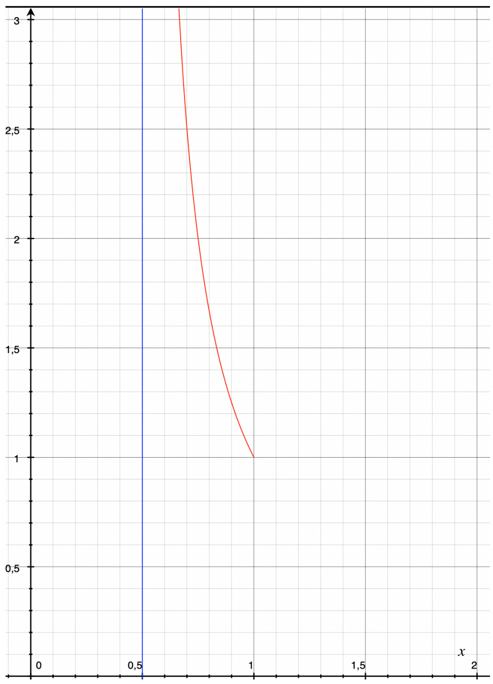

Figure 6: Horizontal: probability p of each winning Bernoulli event. Vertical: expected bet amount.

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 $X_1, \dots, X_n$ be independent and identically distributed (I.I.D.) random variables following $\mathcal{B}(p)$ with $p \in (0,1)$, and let $S = X_1 + \dots + X_n$. 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 100 to compute the conditional expectation.

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

$$\begin{aligned} x \longmapsto \mathbb{P}[X_1 = x \mid S = s] \\ &= \frac{\mathbb{P}[X_1 = x \cap S = s]}{\mathbb{P}[S = s]} \\ &= \frac{\mathbb{P}[X_1 = x \cap X_2 + \dots + X_n = s - x]}{\mathbb{P}[S = s]} \\ &= \mathbb{P}[X_1 = x] \frac{\mathbb{P}[X_2 + \dots + X_n = s - x]}{\mathbb{P}[S = s]} \end{aligned} \tag{111}$$

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

$$\mathbb{P}[X_1 = x \mid S = s] = \frac{p^x (1-p)^{1-x} \binom{n-1}{s-x} p^{s-x} (1-p)^{n-1-s+x}}{\binom{n}{s} p^s (1-p)^{n-s}} = \frac{\binom{n-1}{s-x}}{\binom{n}{s}} \tag{112}$$

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

$$\mathbb{P}[X_1 = 0 \mid S = s] = \frac{n-s}{n}, \tag{113}$$

while for $x = 1$ we get

$$\mathbb{P}[X_1 = 1 \mid S = s] = \frac{s}{n}. \tag{114}$$

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

$$\mathbb{P}[X_1 = x \mid S = s] = \left(\frac{s}{n}\right)^x \left(\frac{n-s}{n}\right)^{1-x} \tag{115}$$

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 100 the conditional expectation of X_1 given S is

$$\mathbb{E}[X_1 \mid S] = \psi(S) \text{ with } \psi(s) = \sum_{x \in \{0,1\}} x \mathbb{P}[X_1 = x \mid S = s] = \frac{s}{n} \tag{116}$$

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

$$\mathbb{E}[X_1 \mid S] = \frac{S}{n}. \tag{117}$$

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

$$\{0, 1, ..., s\} \ni x \longmapsto \mathbb{P}[X = x \mid S = s] \tag{118}$$

Following the steps of the previous exercise we readily arrive to

$$\mathbb{P}[X = x \mid S = s] = \frac{\mathbb{P}[X = x]\mathbb{P}[Y = s - x]}{\mathbb{P}[S = s]}. \tag{119}$$

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

$$\mathbb{P}[X = x \mid S = s] = \frac{e^{-\lambda}\lambda^x}{x!}\frac{e^{-\mu}\mu^{s-x}}{(s-x)!}\frac{s!}{e^{-(\lambda+\mu)}(\lambda+\mu)^s} \tag{120}$$

which can be manipulated – multiply and divide by $(\lambda + \mu)^x$ – into

$$\mathbb{P}[X = x \mid S = s] = \binom{s}{x} p^x (1-p)^{s-x} \text{ with } p = \frac{\lambda}{\lambda + \mu}. \tag{121}$$

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

$$\mathbb{E}[X \mid S] = \psi(S) \text{ with } \psi(s) = \sum_{x=0}^{s} x\binom{s}{x} p^x (1-p)^{s-x} = sp \tag{122}$$

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

$$\mathbb{E}[X \mid X + Y] = \frac{\lambda}{\lambda + \mu}(X + Y). \tag{123}$$

Exercise 5.4 *(An electronic component has a lifetime...)* Exercice corrigé 2.6.4. in [1].

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

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

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 - Continuous random variables

This theory recap covers the material of both this session (Section 6) and the next one (Section 7). During the TD present briefly only what is relevant for the exercises of this session.

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

Notation Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, 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:

$$\begin{aligned}\{a < X \leq b\} &:= \{\omega \in \Omega : X(\omega) \in (a, b]\} \\ &\equiv \{\omega \in \Omega : X(\omega) \leq b\} \cap \{\omega \in \Omega : X(\omega) > a\} \\ &\equiv \{\omega \in \Omega : X(\omega) \leq b\} \setminus \{\omega \in \Omega : X(\omega) \leq a\}.\end{aligned} \tag{124}$$

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

Definition 6.1.1 *(Continuous random variable)* Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, 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 $a < b \in \hat{\mathbb{R}}$,

$$\mathbb{P}\{a < X \leq b\} = \int_a^b \rho_X(x) dx. \tag{125}$$

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

$$F_X(x) = \mathbb{P}[X \leq x]. \tag{126}$$

Proposition 6.1.4 *(PDF = CDF')* The probability density function ρ_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):

$$\rho_X = F'_X \tag{127}$$

Proof By Equation 124, σ-additivity, and the fundamental theorem of integral calculus,

$$\begin{aligned}\mathbb{P}\{a < X \leq b\} &= \mathbb{P}[X \leq b] - \mathbb{P}[X \leq a] \\ &= F_X(b) - F_X(a) \\ &= \int_a^b F'_X(x) dx.\end{aligned} \tag{128}$$

□

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

$$\mathbb{E}[X] = \int_{\mathbb{R}} x \rho_X(x) dx. \tag{129}$$

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - \mathbb{E}[X]^2. \tag{130}$$

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

6.2. Vector-valued random variable

Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ 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 $D \subseteq \mathbb{R}^n$,

$$\mathbb{P}\{X \in D\} = \int_D \rho_X(x) dx = \int_{\mathbb{R}^n} \mathbb{1}_D(x) \rho_X(x) dx. \tag{131}$$

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 ρ_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

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^n} f(x) \rho_X(x) dx. \tag{132}$$

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,

$$\mathbb{E}[h(f(X))] = \int_{\mathbb{R}} h \circ f(x) \rho_X(x) dx. \tag{133}$$

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

$$\mathbb{E}[h(f(X))] = \int_{\mathbb{R}} h(y) \rho_{f(X)}(y) dy \tag{134}$$

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

$$\begin{aligned} \mathbb{E}[h(f(X))] &= \int_{\mathbb{R}} h \circ f(x) \rho_X(x) dx \qquad \text{change variable: } y = f(x) \\ &= \int_{\mathrm{Im}(f)} h(y) \frac{\rho_X(x)}{f'(x)} dy \qquad \text{with } x = f^{-1}(y), \end{aligned} \tag{135}$$

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

$$\rho_{f(X)}(y) = \frac{\rho_X(x)}{f'(x)} \mathbb{1}_{\{\mathrm{Im}(f)\}}(y) \qquad \text{with } x = f^{-1}(y). \tag{136}$$

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

$$\rho_{e^X}(y) = \frac{e^{-\frac{(\ln y)^2}{2}}}{\sqrt{2\pi}} \cdot \frac{1}{y} \cdot \mathbb{1}_{(0,\infty)}(y) \tag{137}$$

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 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 $F_X(x) = \mathbb{P}[X \leq x]$ is given by the expected value of the indicator function of $\{X \leq x\}$, that is

$$F_X(x) = \mathbb{E}\left[\mathbb{1}_{\{X \leq x\}}\right]. \tag{138}$$

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

$$\begin{aligned} x \longmapsto \mathbb{E}\left[\mathbb{1}_{\{X \leq x\}}\right] &= \int_{\mathbb{R}} \mathbb{1}_{\{\leq x\}}(y) \rho_X(y) dy && \text{by transfert theorem} \\ &= \int_{-\infty}^{x} 1 \cdot \rho_X(y) dy && \text{by def of indicator function} \\ &= \mathbb{P}[-\infty \leq X \leq x] && \text{by def of pdf} \\ &= F_X(x) && \text{by def of cdf,} \end{aligned} \tag{139}$$

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:

$$\rho_{(X,Y)}(x, y) = \rho_X(x) \rho_Y(y). \tag{140}$$

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 $f(X,Y)$.

$$\begin{aligned} z \longmapsto \mathbb{P}[f(X,Y) \leq z] &= \mathbb{E}\left[\mathbb{1}_{\{f(X,Y)\leq z\}}\right] && \text{by the indicator technique} \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{f(x,y)\leq z\}}(x,y)\rho_{X,Y}(x,y)dxdy && \text{by transfert theorem} \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\{f(x,y)\leq z\}}(x,y)\rho_X(x)\rho_Y(y)dxdy && \text{by independence} \end{aligned} \tag{141}$$

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 145, 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. [2] 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 ρ_{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

$$\begin{aligned} z \longmapsto \mathbb{P}[X+Y=z] \\ &= \sum_{\substack{x\in F \\ y\in G \\ x+y=z}} \mathbb{P}[X=x]\mathbb{P}[Y=y] \\ &= \sum_{x\in F} \mathbb{P}[X=x]\mathbb{P}[Y=z-x] \end{aligned} \tag{142}$$

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 ρ_X, ρ_Y respectively. The *convolution* of ρ_X and ρ_Y is the function $\rho_X * \rho_Y$ defined by

$$\rho_X * \rho_Y(z) := \int_{\mathbb{R}} \rho_X(x)\rho_Y(z-x)dx. \tag{143}$$

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:

$$\rho_{X+Y} = \rho_X * \rho_Y. \tag{144}$$

Proof We first find the CDF of $X + Y$ following the steps of Equation 141 specializing to $f(X,Y) = X + Y$, then differentiate to find the PDF.

$$
\begin{aligned}
z \longmapsto \mathbb{P}[X+Y \leq z] = \mathbb{E}\left[\mathbb{1}_{\{X+Y\leq z\}}\right] & & \text{by the indicator technique} \\
= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\leq z\}}(x,y)\rho_{X,Y}(x,y)dxdy & & \text{by transfert theorem} \\
= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\leq z\}}(x,y)\rho_X(x)\rho_Y(y)dxdy & & \text{by independence.}
\end{aligned}
\tag{145}
$$

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 11.

So the CDF of $X+Y$ is

$$
F_{X+Y}(z) = \mathbb{P}[X+Y\leq z] = \int_{\mathbb{R}} \rho_X(x)\left(\int_{-\infty}^{z-x} \rho_Y(y)dy\right)dx \tag{146}
$$

The PDF follows by differentiating with respect to z:

$$
\begin{aligned}
\rho_{X+Y}(z) = \frac{d}{dz}F_{X+Y}(z) &= \int_{\mathbb{R}} \rho_X(x)\frac{d}{dz}\left(\int_{-\infty}^{z-x} \rho_Y(y)dy\right)dx \\
&= \int_{\mathbb{R}} \rho_X(x)\rho_Y(z-x)dx
\end{aligned}
\tag{147}
$$

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

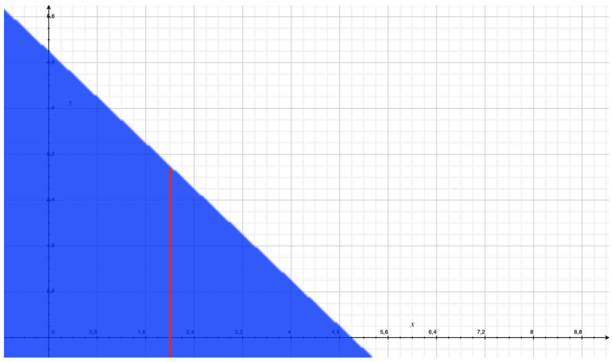

Figure 11: Domain where $\mathbb{1}_{\{x+y\leq z\}}$ is non-zero: $x \in (-\infty,+\infty)$ and $y \in (-\infty, z-x)$.

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

$$
\rho_X(x) = \frac{\mathbb{1}_{[a,b]}(x)}{b-a} \tag{148}
$$

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

$$
\rho_X(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ \lambda e^{-\lambda x} & \text{if } x > 0 \end{cases} = \lambda e^{-\lambda x}\mathbb{1}_{\{x>0\}} \tag{149}
$$

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

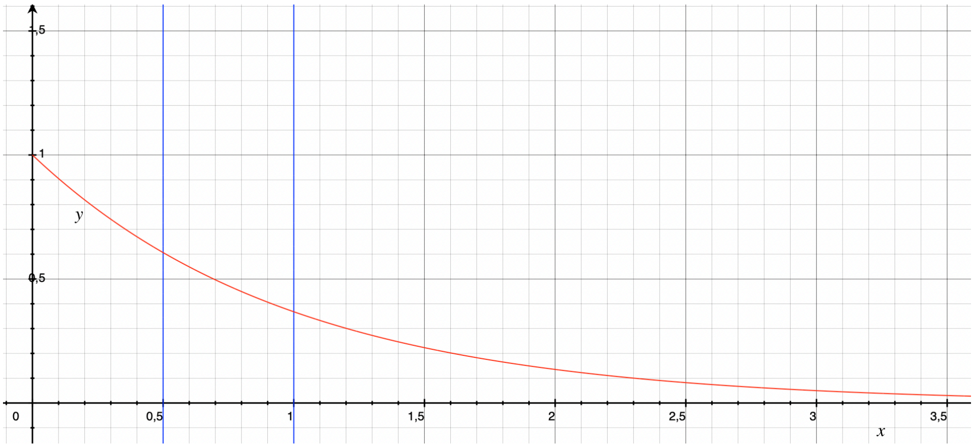

Figure 12: Exponential pdf. $\mathbb{P}[0.5 < X < 1] = \int_{0.5}^{1} \rho_X(x)dx$ is the area below the red curve delimited by the blue lines.

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

$$\rho_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{150}$$

6.6. Misc

For the following c.f. Prop. 3.3.6. in [1].

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

$$\rho_X(x) = \int_{\mathbb{R}} \rho_{X,Y}(x, y)dy \tag{151}$$

and similarly for ρ_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.

$$\begin{aligned} F_{\min(X,Y)}(z) &= \mathbb{P}[\min(X, Y) \leq z] \\ &= \mathbb{P}[X \leq z \text{ or } Y \leq z] \\ &= 1 - \mathbb{P}[X > z \text{ and } Y > z] \\ &= 1 - \mathbb{P}[X > z]\mathbb{P}[Y > z] \text{ by independence} \end{aligned} \tag{152}$$

We have $\mathbb{P}[X > z] = \int_z^\infty \mu e^{-\mu x} dx = e^{(-\mu z}$, 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,

$$\begin{aligned}
F_{X+Y}(z) &= \mathbb{P}[X + Y \leq z] = \mathbb{E}\left[\mathbb{1}_{\{X+Y\leq z\}}\right] \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\leq z\}}(x, y)\rho_X(x)\rho_Y(y)dxdy \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\leq z\}}(x, y)\mathbb{1}_{\{x\geq 0\}}(x)\mathbb{1}_{\{y\geq 0\}}(y)\lambda e^{-\lambda x}\mu e^{-\mu y}dxdy \\
&= \int_0^z \lambda e^{-\lambda x}\left(\int_0^{z-x} \mu e^{-\mu y}dy\right)dx \\
&= \text{...some standard calculations...} \\
&= 1 - e^{-\lambda z} - \frac{\lambda}{\lambda - \mu}\left[e^{-\mu z} - e^{-\lambda z}\right]
\end{aligned} \tag{153}$$

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

$$\rho_{X+Y}(z) = \frac{\lambda\mu}{\lambda - \mu}\left[e^{-\mu z} - e^{-\lambda z}\right]. \tag{154}$$

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 147:

$$\begin{aligned}
\rho_{X+Y}(z) &= \int_{\mathbb{R}} \rho_X(x)\rho_Y(z - x)dx \\
&= \int_0^z \lambda e^{-\lambda x}\mu e^{-\mu(z-x)}dx
\end{aligned} \tag{155}$$

which after some standard steps leads back to Equation 154.

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

$$\begin{aligned}
F_{X^2}(z) &= \mathbb{P}[X^2 \leq z] = \mathbb{E}\left[\mathbb{1}_{\{X^2\leq z\}}\right] \\
&= \int_{\mathbb{R}} \mathbb{1}_{\{x^2\leq z\}}(x)\rho_X(x)dx \\
&= \int_0^{\sqrt{z}} \rho_X(x)dx = 1 - e^{-\lambda\sqrt{z}}
\end{aligned} \tag{156}$$

From which the pdf is

$$\frac{d}{dz}F_{X^2}(z) = \rho_{X^2}(z) = \frac{\lambda}{2}\frac{e^{-\lambda\sqrt{z}}}{\sqrt{z}}. \tag{157}$$

Exercise 6.4 *(Split a stick into a triangle)* Exercice corrigé 3.5.8. in [1].

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