People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University of 8 mai 1945 Guelma

Faculty of Mathematics, Computer Science
and Sciences of Matter
Department of Mathematics

Course

Presented to students of the second year

Computer Engineer

Title

Probability and Statistics 2

By:

Dr. EZZEBSA Abdelali

Academic year 2024/2025

Contents

Part I
Properties of Mathematical Expectation

1 General Introduction

This course is specifically designed for second-year students in computer engineering. It is also valuable for students preparing for exams that require a solid understanding of mathematical tools, particularly the concepts and terminology related to probability theory.

Objectives

The primary objectives of the course are:

- To incorporate the role of randomness in decision-making processes.
- To understand and develop strategies for prediction in uncertain environments.

Fields of Application

- Medicine
- Quality control
- Environmental studies
- Insurance
- Games
- Computer science

Historically, probability theory emerged in the 17^{th} century through the study of games of chance, focusing on situations involving a finite number of outcomes. More advanced developments, which address infinite or continuous sample spaces, rely on more sophisticated tools from probability theory. However, all fundamental concepts can still be introduced and understood within the context of finite probability spaces.

Course Structure

The course is divided into two main parts:

1. **First Part**: Introduction to key concepts such as expectation, conditional expectation, covariance, and correlation. It also includes a section on different modes of convergence to emphasize the importance of approximating discrete and continuous probability distributions and their simulations.

2. **Second Part**: Focus on statistical inference and hypothesis testing. It also introduces topics like information theory, including notions of surprise, uncertainty, entropy, and information encoding.

This course aims to equip computer engineering students with the foundational tools necessary for modeling and analyzing data effectively.

2 Random variables

Definition 1 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable (often abbreviated r.v subsequently) is an application*

$$\begin{cases} X : \Omega \to \mathbb{E} \subset \mathbb{R} \\ \omega \longmapsto X(\omega) \end{cases}$$

who checks the condition, for all $x \in \mathbb{R}$

$$X^{-1}(]-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$$

Example 2 *If we toss a coin 2 times in a row and we are interested in the number of tails obtained.We denote by X: "**The number of tails obtained during the two tosses**": Let us show that X represents a random variable on $(\Omega, \mathcal{F})$ such that*

$$\Omega = \{(a_1, a_2); a_i = \textit{heads or tails } (F = T \vee P = H) \ \ i = 1, 2\}$$

and $\mathcal{F} = \mathcal{P}(\Omega)$, Indeed

$$\begin{cases} \textit{if } x < 0\textit{:} \ X^{-1}(]-\infty, x]) = \phi, \\ \textit{if } 0 \leq x < 1 : X^{-1}(]-\infty, x]) = \{(T;T)\}, \\ \textit{if } 1 \leq x < 2 : X^{-1}(]-\infty, x]) = \{(H;T);(T;H)\}, \\ \textit{if } x \geq 2 : X^{-1}(]-\infty, x]) = \Omega. \end{cases}$$

So it is obvious that for every thing $x \in \mathbb{R} : X^{-1}(]-\infty, x]) \in \mathcal{P}(\Omega)$, so X represents a random variable on $(\Omega, \mathcal{F}, P)$.

Remark 3 *In general we denote the set of values taken by the random variable X; by D_X and we call it the support of the random variable X.*

2.1 Discrete Random Variable

Definition 4 *A discrete random variable is a function $X(s)$ from a finite or countably infinite sample space $\mathcal{F}$ to the real numbers*

$$X : \mathcal{F} \to \mathbb{R}.$$

Example 5 *Toss a coin* **2** *times in sequence. The sample space is*

$$\mathcal{F} = \{HH, HT, TH, TT\},$$

and examples of random variables are
► $X(s) = 1$ *The number of Heads in the sequence ;* ***e.g***, $X(HT) = 1$,
► $Y(s) =$*The index of the first H;* ***e.g***, $Y(TH) = 2$,
0 *if the sequence has no H ,* ***i.e.***, $Y(TT) = 0$.

NOTE : In this example $X(s)$ and $Y(s)$ are actually integers.
Value-ranges of a random variable correspond to events in $\mathcal{F}$.

Example 6 *For the sample space*

$$\mathcal{F} = \{HH, HT, TH, TT\},$$

with $X(s)$ *=the number of Heads, the value* $X(s) = 1$*, corresponds to the event* $\{HT, TH\}$ *,and the values* $1 < X(s) \leq 2$*, correspond to*

$$\{HH\}$$

Definition 7 *You can say that a random variable* X *has a discrete probability distribution when: exists finite or enumerable set of real numbers* $M = \{x_1, ..., x_n, ...\}$ *that*

$$\begin{cases} P(X = x_i) \geq 0; \ i = 1, 2, ... \\ \sum_i P(X = x_i) = 1 \end{cases}$$

Function $P(X = x_i) \Leftrightarrow P(x_i)$ *is called probability function of random variable* X*. A distribution function of such a distribution is a step function with steps in* $x_!, ..., x_n, ...$ *For a distribution function of a discrete random variable the following is true:*

$$F(x) = \sum_{x_i \leq x} P(X = x_i).$$

Example 8 *Throwing a dice,* X *a number of dots obtained*

x_i	P_i	$F(x_i)$
1	$\frac{1}{6}$	0
2	$\frac{1}{6}$	$\frac{1}{6}$
3	$\frac{1}{6}$	$\frac{2}{6}$
4	$\frac{1}{6}$	$\frac{3}{6}$
5	$\frac{1}{6}$	$\frac{4}{6}$
6	$\frac{1}{6}$	$\frac{5}{6}$

Definition 9 *Let* X *be a real random variable with support* D_X*: If* D_X *is a countable subset of* $\mathbb{R}$*; then* X *is called a* ***discrete random variable****.*

Properties

Proposition 10 *If* X *and* Y *are two real discrete random variables on* $(\Omega, \mathcal{F}, P)$*; then*

$$\begin{cases} i) \text{ for allr eal } a; b : aX + bY \text{ is a discrete random variable,} \\ ii) \ XY \text{ is a discrete random variable,} \\ iii) \ \sup(X;Y) \text{ and } \inf(X;Y) \text{ is a discrete random variable.} \end{cases}$$

Proposition 11 *Let X be a random variable defined on Ω, and with valuesin $D_X = \{x_1, x_2, ..., x_n\}$. If for all $i = 1, ..., n$:*

$$A_i = (X = x_i) = \{\omega \in \Omega : X(\omega) = x_i\}$$

So, the family $\{A_i;\ i = 1, ..., n\}$ forms a complete system of Ω and $X = \sum_{i=1}^{n} x_i 1_{A_i}$ where 1_{A_i} denotes the indicator function of the set A_i, that is to say

$$1_{A_i}(\omega) = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ Otherwise } (\text{if } \omega \notin A) \end{cases}$$

Example 12 *Let $(\Omega, \mathcal{F}, P)$ the probability space of the balanced die*

$$\begin{aligned} X_1(\omega) &= 1_{\{2,4,6\}} : P(\{\omega \in \Omega; X_1(\omega) = 1\}) = P(\{2,4,6\}) = \frac{3}{6}. \\ X_2(\omega) &= 1_{\{1,3,5\}}(\omega) : P(\{\omega \in \Omega; X_2(\omega) = 1\}) = P(\{1,2,3\}) = \frac{3}{6} = \frac{1}{2}. \end{aligned}$$

2.2 Distribution function of a random variable

$$F_X(t) = P(\{X \leq t\});\ t \in \mathbb{R}.$$

Remark 13 *If the range of the random variable function is discrete, then the random variable is called **a discrete random** variable. Otherwise, if the range includes a complete interval on the real line, the random variable is **continuous**.*

Proposition 14 *The data of F_X is equivalent to that of P_X Law of a R.V.*

Properties of the probability distribution function:

1.$0 \leq F(x) \leq 1$ for $+\infty < x < \infty-$

2. The distribution function is a monotonic increasing function of x, i.e. $\forall x_1, x_2 \in \mathbb{R} : x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$

3. The distribution function $F(x)$ is left-continuous.

4. $\lim_{x \to +\infty} F(x) = 1,\ \lim_{x \to -\infty} F(x) = 0.$

5. $\forall a, b \in \mathbb{R};\ a < b : P(a \leq X < b) = F(b) - F(a)$

6. $P(X = x_0) = \lim_{x \to x_0^+} F(x) - F(x_0)$

3 Mathematical expectation and Properties

3.1 Mathematical Expectation

The mathematical expected value of a discrete random variable X is

$$\mathbb{E}[X] = \sum_k x_k \times P(X = x_k) = \sum_k x_k \times P_X(x_k).$$

Thus $\mathbb{E}[X]$ represents the weighted average value of X ($\mathbb{E}[X]$ is also called **the mean** of X).

Example 15 *The expected value of rolling a die is*

$$\mathbb{E}[X] = 1 \times \frac{1}{6} + 2 \times \frac{2}{6} + ... + 6 \times \frac{1}{6} = \frac{1}{6}\sum_{k=1}^{6} k = \frac{7}{2}.$$

Proposition 16 *Prove the following*
▶ *The mathematical expectation is linear* $\mathbb{E}[aX] = a\mathbb{E}[X]$ *where* a *is a real number*
▶ *Expectation of a sum of random variables:* X *and* Y *two random variables* $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$*, where* a, b *two real numbers*
▶ *The expected value of a function of a random variable is*

$$\mathbb{E}[g(X)] = \sum_k g(x_k) P(x_k).$$

Example 17 *The pay-off of rolling a die is* $k^2 DZD$*, where* k *is the side facing up.*

What should the entry fee be for the betting to break even?

Solution 18 *Here* $g(X) = X^2$*, and*

$$\mathbb{E}[g(X)] = \sum_{k=1}^{6} k^2 \frac{1}{6} = \frac{1}{6} \frac{6(6+1)(2.6+1)}{6} = \frac{91}{6} \simeq 15,17DZD.$$

4 Covariance, Sum Variance, Correlation

4.1 Variance and Standard Deviation

Let X have mean

$$\mu = \mathbb{E}[X].$$

Then the variance of X is

$$Var(X) = \mathbb{E}[(X-\mu)^2] = \sum_k (x_k - \mu)^2 p(x_k),$$

which is the average weighted square distance from the mean. We have

$$\begin{aligned} Var(X) &= \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2. \end{aligned}$$

The standard deviation of X is

$$\sigma_X = \sqrt{Var(X)} = \sqrt{\mathbb{E}[X^2] - \mu^2}.$$

which is the average weighted distance from the mean.

Example 19 *The variance of rolling a* ***die*** *is*

$$\begin{aligned} Var\,(X) &= \sum_{k}^{6}\left(k^2.\frac{1}{6}\right)-\mu^2 \\ &= \frac{1}{6}\frac{6(6+1)(2{\cdot}6+1)}{6}-\left(\frac{7}{2}\right)^2=\frac{35}{12}. \end{aligned}$$

The standard deviation is

$$\sigma=\sqrt{\frac{35}{12}}\simeq 1.70.$$

4.2 Covariance

Let X and Y be random variables with mean

$$\mathbb{E}[X]=\mu_X,\ \mathbb{E}[Y]=\mu_Y.$$

Then the covariance of X and Y is defined as

$$Cov(X,Y)\equiv\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]=\sum_{k,l}(x_k-\mu_X)(y_l-\mu_Y)p(x_k,y_l).$$

We have

$$\begin{aligned} Cov(X,Y) &\equiv \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \mathbb{E}[XY-\mu_XY-\mu_YX+\mu_X\mu_Y] \\ &= \mathbb{E}[XY]-\mu_X\mu_Y-\mu_Y\mu_X+\mu_X\mu_Y \\ &= \mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

Remark 20 *$Cov(X,Y)$ measures "**concordance** " **or** "**coherence** " of X and Y:*

▶If $X>\mu_X$ when $Y>\mu_Y$ and $X<\mu_X$ when $Y<\mu_Y$ then

$$Cov(X,Y)>0.$$

▶If $X>\mu_X$ when $Y<\mu_Y$ and $X<\mu_X$ when $Y>\mu_Y$ then

$$Cov(X,Y)<0.$$

Proposition 21 *Here are some properties of variance*
▶ $Var(aX+b)=a^2Var(X)$; $a,b\in\mathbb{R}$
▶ $Cov(X,Y)=Cov(Y,X)$,
▶ $Cov(cX,Y)=cCov(X,Y)$; $c\in\mathbb{R}$
▶ $Cov(X,cY)=cCov(X,Y)$,
▶ $Cov(X+Y,Z)=Cov(X,Z)+Cov(Y,Z)$,
▶ $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$.

Proposition 22 *If X and Y are independent then $Cov(X, Y) = 0$.*

Proof. We have already shown (with $\mu_X \equiv \mathbb{E}[X]$ and $\mu_Y \equiv \mathbb{E}[Y]$) that

$$Cov(X, Y) \equiv \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y],$$

and that if X and Y are independent then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

from which the result follows. ■

Example 23

X/Y	$y = 6$	$y = 8$	$y = 10$	$P_X(x)$
$x = 1$	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{2}{5}$
$x = 2$	0	$\frac{1}{5}$	0	$\frac{1}{5}$
$x = 3$	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{2}{5}$
$P_Y(y)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	1

Show that

$$\begin{cases} \blacktriangleright\ \mathbb{E}[X] = 2, \\ \blacktriangleright\ \mathbb{E}[Y] = 8, \\ \blacktriangleright\ \mathbb{E}[XY] = 16, \\ \blacktriangleright\ Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0. \end{cases}$$

X and Y are not independent. Thus if

$$Cov(X, Y) = 0,$$

then it does not necessarily follow that X and Y are independent!

Proposition 24 *If X and Y are independent then*

$$Var(X + Y) = Var(X) + Var(Y).$$

Indeed. *We have already shown (in an exercise !) that*

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y),$$

and that if X and Y are independent then

$$Cov(X, Y) = 0,$$

from which the result follows.

Example 25 *Compute*

$$\mathbb{E}[X],\ \mathbb{E}[Y],\ \mathbb{E}[X^2],\ \mathbb{E}[Y^2]$$

$$\mathbb{E}[XY], Var(X), Var(Y)$$

$$Cov(X,Y)$$

for Joint probability mass function $P_{X,Y}(x,y)$

X/Y	$y=0$	$y=1$	$y=2$	$y=3$	$P_X(x)$
$x=0$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$x=1$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$x=2$	0	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$
$x=3$	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
$P_Y(y)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	1

4.3 Correlation

Definition 26 *The linear correlation coefficient of two random variables X and Y is the real number*

$$\rho = Corr = Corr\,(X,Y) = \frac{Cov\,(X,Y)}{\sqrt{V\,(X)}\sqrt{V\,(Y)}}.$$

Remark 27 ρ *it's a number such that*

$$-1 \le \rho \le 1,\ \ with\ \ |\rho| = 1 \Leftrightarrow \exists a \in \mathbb{R}^*, \exists b \in \mathbb{R} : Y = aX + b.$$

4.4 Examples

Exercise 1

A **6**-sided die is rolled. Let p_i be the probability of the face marked i: This die is rigged so that the probabilities of the faces are

$$p_1 = 0,15;\ p_2 = 0,2;\ p_3 = 0,3;\ p_4 = 0,1;\ p_5 = 0,1.$$

1. What is the probability of the face marked **6** coming up.

2. What is the probability of getting an even number.

Exercise 2

We have two urns U_1 and U_2, Urn U_1 contains three white balls and one black ball. Urn U_2 contains one white ball and two black balls.

We roll an unrigged die. If the die gives a number less than or equal to 2; we draw a ball from urn U_1. Otherwise we draw a ball from urn U_2. (We assume that the balls are indistinguishable by touch).

1. Calculate the probability of drawing a white ball.

2. We have drawn a white ball. Calculate the probability that it comes from urn U_1.

Exercise 3

The total production of a factory is carried out by three machines A; B and C according to the percentages 75%; 15% and 10% respectively. The proportions of defective production are 3%; 5% and 6% respectively. A unit of the production of this factory is chosen at random.

1. What is the probability that this unit will be defective.

2. Knowing that the chosen unit is good, what is the probability that it would be produced by machine C?

Exercise 4

Let (Ω, F, P) be a probability space, with Ω finite. Consider the application $X : (\Omega, F) \to \mathbb{R}$.

a) If $n = 2, F = \{\Phi, \Omega\}$, $X(\omega_1) = x_1$ and $X(\omega_2) = x_2$ where $x_1 \neq x_2$.

■Is X a random variable?

b) If $n = 5$, $F = P(\Omega)$ the set of all parts of Ω, $X(\omega_1) = 0$ and $X(\omega_2) = X(\omega_3) = 1$ where $X(\omega_4) = X(\omega_5) = 2$.

1. Verify that X is a random variable on (Ω, F).

2. Determine the mass function and the distribution function of this variable by assuming that the $\{\omega_i; i = 1, ..., n\}$ are equiprobable.

3. Deduce the probabilities $P(-1 < X \leq 2)$ and $P(X \geq 2)$.

Exercise 5

A player throws a balance dice and wins **1DA** if the result is even, he loses **1DA** if the result is 1 or 3 and **loses** or **wins** nothing if the result is 5. We denote X as the random variable equal to the player's gain.

1. Determine the distribution of X.

2. Calculate $\mathbb{E}(\mathbb{X})$ and $Var(X)$.

3. Determine the distribution of the random variable $Y = X^2$ and calculate $\mathbb{E}(\mathbb{Y})$.

answer

Exercise 1

1. The event space Ω is the set $\{1,2,3,4,5,6\}$ we then have

$$\begin{aligned} p(\Omega) &= p(\{1\})+p(\{2\})+p(\{3\})+p(\{4\})+p(\{5\})+p(\{6\})=1 \\ &\Leftrightarrow p_1+p_2+p_3+p_4+p_5+p_6=1 \Leftrightarrow p_6=1-(p_1+p_2+p_3+p_4+p_5) \\ &= 1-0,85=0,15, \end{aligned}$$

that is, the probability of the face marked 6 coming out is $0,15$,

2. The probability of getting an even number is

$$P(\{2,4,6\})=p(\{2\})+p(\{4\})+p(\{6\})=0,2+0,1+0,15=0,45.$$

Exercise 2

1. Consider the events

B: "the drawn ball is white",

A_1: "the drawn ball comes from urn U_1",

A_2: "the drawn ball comes from urn U_2",

So

$$\begin{aligned} P(A_1) &= P(\{1,2\})=\frac{2}{6}=\frac{1}{3}, \\ P(A_2) &= P(\{3,4,5,6\})=\frac{4}{6}=\frac{2}{3}, \\ P(B/A_1) &= \frac{C_3^1}{C_4^1}=\frac{3}{4}, \text{ and } P(B/A_2)=\frac{C_1^1}{C_3^1}=\frac{1}{3}. \end{aligned}$$

Using the Total Probability formula, we find

$$P(B)=P(A_1)P(B/A_1)+P(A_2)P(B/A_2)\simeq 0,47.$$

2. This is to calculate $P(A_1/B)$; using Bayes Theorem, we find

$$P(A_1/B)=\frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1)+P(A_2)P(B/A_2)}=\frac{9}{17}\simeq 0,53.$$

Exercise 3

Indeed: We consider the following events:

M1: "the chosen unit is produced by machine A"

M2: "the chosen unit is produced by machine B"

M3: "the chosen unit is produced by machine C"

D: "the chosen unit is defective"

B: "the chosen unit is good".

It is clear that $\Omega=\{U/U$ unit produced by the machine A or else B or else $C\}$, the events $M1$; $M2$ and $M3$ forms a complete system of Ω

et $P(M_1) = 0,75$; $P(M_2) = 0,15$; $P(M_3) = 0,1$; $P(D/M_1) = 0,03$; $P(D/M_2) = 0,055$; $P(D/M_3) = 0,06$. 1. Using the Total Probability Theorem, we have

$$\begin{aligned} P(D) &= P(M_1)\,P(D/M_1) + P(M_2)\,P(D/M_2) + P(M_3)\,P(D/M_3) \\ &= 0,750 \times 0.03 + 0,15 \times 0,55 + 0,10 \times 0.06 \simeq 037. \end{aligned}$$

2. Event B is the opposite event of event D; therefore

$$P(B) = 1 - P(D) = 1 - 0,037 \simeq 0,963.$$

Using Bayes Theorem, we find

$$P(M_3/B) = \frac{P(M_3)\,P(D/M_3)}{\sum_i P(M_i)\,P(B/M_i)} = \frac{0.1 \times 0.06}{0,963} \simeq 0.006.$$

Exercise 4

a) X is a random variable on $(\Omega, F) \Leftrightarrow \forall x \in \mathbb{R} : X^{-1}(]-\infty, x]) \in F$. Suppose $x_1 < x_2$:if $x \in [x_1, x_2[$,

$$X^{-1}(]-\infty, x]) = \{\omega \in \Omega; X(\omega) \leq x\} = \{\omega_1\} \notin \mathcal{F},$$

So X is not a random variable on (Ω, F).

b) 1. It is clear that for all $x \in \mathbb{R}, X^{-1}(]-\infty, x]) \in P(\Omega)$, so X is a random variable on $(\Omega, P(\Omega))$.

2. The support of the random variable X is $D_X = \{0, 1, 2\}$, and its mass function P_X is defined by:

$$\begin{cases} P_X(0) = P(X=0) = P(\{\omega_1\}) = \dfrac{1}{5} \\ P_X(1) = P(X=1) = P(\{\omega_2, \omega_3\}) = \dfrac{2}{5} \\ P_X(2) = P(X=2) = P(\{\omega_4, \omega_5\}) = \dfrac{2}{5} \end{cases}$$

the distribution function F_X of X is defined by

$$\begin{cases} \text{if } x < 0;\ F_X(x) = 0 \\ \text{if } 0 \leq x < 1;\ F_X(x) = P_X(0) = \dfrac{1}{5} \\ \text{if } 1 \leq x < 2\ ;\ F_X(x) = P_X(0) + P_X(1) = \dfrac{3}{5} \\ \text{if } x \geq 2\ ;\ F_X(x) = P_X(0) + P_X(1) + P_X(2) = 1. \end{cases}$$

hense

$$F_X(x) = \begin{cases} 0 \text{ if } x < 0 \\ \dfrac{1}{5} \text{ if } 0 \leq x < 1 \\ \dfrac{3}{5} \text{ if } 1 \leq x < 2 \\ 1 \text{ if } x \geq 2 \end{cases}$$

Let's calculate

$$\begin{aligned} P(-1 < X \leq 2) &= F_X(2) - F_X(-1) = 1 - 0 = 1 \text{ et } P(X \geq 2) \\ &= 1 - P(X < 2) = 1 - F_X(1) \\ &= 1 - \frac{3}{5} = \frac{2}{5} = 0,4 \end{aligned}$$

Exercise 5

1. The support of the *R.V* is $D_X = \{-1, 0, 1\}$ and its mass function P_X is defined by

$$\begin{cases} P_X(-1) = P(\{1,3\}) = \dfrac{2}{6} = \dfrac{1}{3} \\ P_X(0) = P_X(\{5\}) = \dfrac{1}{6} \\ P_X(1) = P(\{2,4,6\}) = \dfrac{3}{6} = \dfrac{1}{2}. \end{cases}$$

2.
$$\mathbb{E}(X) = (-1) \times \frac{1}{3} + 0 \times \frac{1}{6} + 1 \times \frac{1}{2} = \frac{1}{6} \simeq 0,17$$
and
$$\mathbb{E}\left(X^2\right) = (-1)^2 \times \frac{1}{3} + 0^2 \times \frac{1}{6} + 1^2 \times \frac{1}{2} = \frac{5}{6}$$
hence
$$Var(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}^2(X) = \frac{5}{6} - \frac{1}{36} = \frac{29}{36} \simeq 0,8.$$
It is clear that the support of the $R.V$ Y is $D_Y = \{0,1\}$ and its mass function P_Y is defined by
$$\begin{aligned} P_Y(0) &= P_X(0) = \frac{1}{6} \\ P_Y(1) &= P_X(-1) + P_X(1) = \frac{1}{3} + \frac{1}{2} = \frac{5}{6} \end{aligned}$$
So
$$\mathbb{E}(Y) = 0 \times \frac{1}{6} + 1 \times \frac{5}{6} = \frac{5}{6} \simeq 0,83.$$

5 Usual laws

5.1 Bernoulli's law of parameter p, B(p)

This is the law of a random variable X which can only take 2 values, noted 1 and 0, and $p \in [0;1]$ is the probability of the value 1:

$$P(X=1) = p \text{ and } P(X=0) = 1-p$$

This is therefore the law of the indicator function 1_A of an event A such that $P(A) = p$. We have

$$\mathbb{E}[X] = p \text{ and } Var(X) = p(1-p)$$

5.2 Binomial law with parameters n and p, B(n,p)

Let n random variables $X_1, X_2, ..., X_n$ be independent and have the same distribution $B(p)$. The binomial distribution $B(n,p)$ is the distribution of the random variable $S_n = X_1 + X_2 + +X_n$. It is therefore the distribution of the number of events among $A_1, ..., A_n$ that are realized, if $A_1, ..., A_n$ are independent and have the same probability p. (Above, $X_n = 1_{A_n}$) S_n has values in $\{0, 1, ..., n\}$ and we have:

$$\text{for } k = 0, 1, ..., n; \; P(S_n = k) = C_n^k p^k (1-p)^{n-k}$$

What's more

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = np$$

and since the random variables are independent,

$$Var(X) = \sum_{i=1}^{n} Var(X_i) = np(1-p)$$

5.3 Poisson's law

Let $\lambda > 0$. A random variable X follows the Poisson distribution with parameter λ if

$$X(\Omega) = \mathbb{N} \text{ and for everything } k \in \mathbb{N}; P(X=k) = \exp(-k) \frac{\lambda^k}{k!}$$

we have

$$\mathbb{E}[X] = \lambda \text{ and } Var(X) = \lambda$$

This is the **"law of small probabilities"** because the limit law of the binomial law $B(n;p)$, with $np \sim \lambda$.

Remark 28 *Fields of application: The Poisson law models rare phenomena, it can also be used to approximate the binomial law as we will see later.*

Example 29 *Assume that out of* 1000 *people traveling by train at a given time, there is on average* 1 *doctor. Let* X *be the rv representing the number of doctors on the train.*
1. What is the probability distribution of the r v X*?*
2. What is the probability of finding:
a) No doctors.
b) Between 1 *and* 2 *doctors (in the broad sense).*
c) At least two doctors.

Answer
1. The probability distribution of the $r.v.X$:
The r.v. X follows a Poisson distribution with parameter $\lambda = 1$, we write $X \to \mathcal{P}(1)$, and we have

$$P(X = k) = exp(-\lambda)\frac{\lambda^k}{k!} = exp(-1)\frac{1}{k!}$$

2. The probability of finding
a) No doctor is

$$P(X = 0) = e^{-1}\frac{1^0}{0!} = e^{-1} = 0.368$$

b) Between 1 and 2 doctors (in the broad sense) is

$$\begin{aligned} P(1 &\leq X \leq 2) = P(X = 1) + P(X = 2) \\ &= e^{-1}\left(1 + \frac{1}{2!}\right) = 0,552. \end{aligned}$$

c) At least two doctors are

$$\begin{aligned} P(X &\geq 2) = 1 - P(X < 2) \\ &= 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \left(0,368 + e^{-1}\frac{1}{2!}\right) \\ &= 0,448. \end{aligned}$$

Proposition 30 *If* $X \to P(\lambda_1)$ *and* $Y \to P(\lambda_2)$*, the random variables* X *and* Y *being independent, then* $X + Y \to P(\lambda_1 + \lambda_2)$*.*

5.4 Geometric law

We say that a $r.v.$ X follows a geometric law of parameter p if

$$X(\Omega) = \mathbb{N}^* \text{ and for evrything } k \in X(\Omega), P(X = k) = p(1-p)^{k-1}.$$

we note

$$X \to \mathcal{G}(p).$$

She admits for moments:

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

Remark 31 *Application situation: The geometric law models the rank of the first success by repeating a Bernoulli trial identically and independently to infinity (theoretically).*

5.5 Hypergeometric Law

We say that a random variable X defined on a probability space (Ω, F, P) follows a Hypergeometric distribution of parameters N, N_1, n and we denote $X \to H(N, N_1, n)$

if $D_X = \{0, 1, ..., min(N_1; n)\}$ and its mass function p_X is given by

$$P_X(k) = \begin{cases} \dfrac{C_{N_1}^k \times C_{N-N_1}^{n-k}}{C_N^n} & \text{if } k \in D_X \\ 0 & \text{otherwise} \end{cases}$$

Example 32 *Consider an urn containing N balls of which N_1 are white balls and $N - N_1$ are black balls. We draw n balls at once and we note by X:* ***"The number of white balls among the n drawn".***

Using classical probability theory, it is easy to notice that

$$P(X = k) = \frac{C_{N_1}^k \times C_{N-N_1}^{n-k}}{C_N^n} \text{ for evrything } k \in \{0, 1, ..., \min(N_1, n)\}$$

so it is a random variable which follows a hypergeometric law with parameters N, N_1, n, that is to say $X \to H(N, N_1, n)$.

Remark 33 *If $X \to H(N, N_1, n)$, then we have*

$$\mathbb{E}(X) = np \text{ and } var(X) = np(1-p)\frac{N-n}{N-1},$$

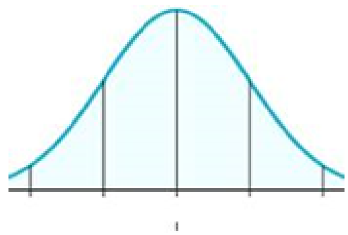
where $p = \dfrac{N_1}{N}$.

5.6 Normal Distribution

How to explain the normal distribution?

The normal distribution, or normal distribution, defines a representation of data according to which most of the values are grouped around the mean and the others deviate symmetrically on both sides.

The normal distribution is the most widespread and useful statistical distribution. It accounts for many random phenomena. m is the mean and σ is the standard deviation, when we write X follows $N(m, \sigma)$. The following graph shows the shape of the density of this law



Such as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \frac{-1}{2} \left(\frac{x-m}{\sigma}\right)^2 ; x \in \mathbb{R}$$

A *r.v.* follows a normal distribution $N(40, 5)$, this means that the mean value (expectation) of is worth $E(X) = 40$ and 5 designates its variance, therefore $\sigma(X) = 5$ and its distribution function is given by

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

5.6.1 Properties of the normal law

Let X be a random variable following a reduced $\sigma = 1$, centered normal $E(X) = 0$ and F its distribution function, we have:

$$\begin{cases} 1. P(X \geq t) = 1 - P(X < t) = 1 - F_X(x) \\ 2. \text{ If } t \text{ is positif } F_X(-x) = 1 - F_X(x) \\ 3. \text{For everything } a, b \in \mathbb{R}, \text{ with } a \leq b: \\ P(a \leq X \leq b) = F_X(a) - F_X(b) \\ 4. \text{ For everything } t \geq 0: \; P(-t \leq X \leq t) = 2F_X(t) - 1. \end{cases}$$

Example 34 *If we have*

$$F(t) = 0,9750.$$

We read on the table of the Normal law $N(0, 1)$ the value of t which is equal to $1,96$.

5.7 Exponential laws

Motivation

When one wants to establish a real mathematical model, it is often necessary to make many simplifying assumptions to make the model flexible from a computational point of view. A simplifying assumption often made in practice is that some random variables are distributed according to an exponential law. This is justified by the simplicity of calculation linked to this law but also by the fact that it often constitutes a good approximation of the real phenomenon.

The exponential law is the law of the life of a material that does not wear out over time. A material having a constant failure rate over time.

Remark 35 *The exponential law is the only law that has such a property.*

Definition 36 *A real random variable X follows an exponential distribution with parameter $\lambda > 0$ if its density is expressed by*

$$f(x) = \begin{cases} \lambda \exp(-\lambda) & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Its distribution function is

$$F(x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 1 - \exp(-\lambda x) & \text{if } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

Mathematical expectation and variance of X

$$\mathbb{E}(X) = \int_{0}^{+\infty} x f(x)\, dx = \frac{1}{\lambda}$$

and

$$Var(X) = \mathbb{E}\left(X^2\right) - \mathbb{E}^2(X)$$

such as

$$\mathbb{E}\left(X^2\right) = \int_{0}^{+\infty} x^2 f(x)\, dx = \frac{2}{\lambda^2}.$$

From the two previous equations, we deduce that

$$Var(X) = \frac{1}{\lambda^2}.$$

5.8 Generating function of the moments of a discrete Random variable

The moment generating function of a discrete *r.v.* X is given by:

$$G_X(t) = \mathbb{E}(e^{tX}) = \sum_{k \in X(\Omega)} e^{tk} P(X = k).$$

Example 37 *The generating function of the moments of a r.v.$X \to B(p)$:*

$$\begin{aligned} G_X(t) &= \mathbb{E}(e^{tX}) = \sum_{k \in X(\Omega)} e^{tk} P(X = k). \\ &= e^{t \times 0}(1-p) + e^{t \times 1} p \\ &= (1-p) + pe^t. \end{aligned}$$

The generating function of the moments of a r.a.$Y \to G(p)$:

$$\begin{aligned}
G_Y(t) &= \mathbb{E}(e^{tY}) = \sum_{k\in X(\Omega)} e^{tk}P(Y=k). \\
&= \sum_{k=1}^{\infty} e^{tk}p(1-p)^{k-1} \\
&= \frac{p}{1-p}\sum_{k=1}^{\infty}((1-p)e^t)^k \\
&= \frac{p}{(1-p)}\frac{1}{(1-(1-p)e^t)}.
\end{aligned}$$

Example 38 *A certain equipment has a probability $p = 0.02$ of failure each time it is put into service. The following experiment is carried out: the equipment is started, stopped, restarted, stopped, until it breaks down. Let X be the r.v representing the number of trials required to obtain the failure.*
1. What is the probability distribution of the r.v X?
2. What is the probability that this equipment breaks down (for the first time) on the tenth trial?

Answer

1. The probability law of the *r.v.* X:

The *r.v.* X follows a geometric law with parameter $p = 0.02$, we write $X \to G(0.02)$, and we have:

$$P(X=k) = p(1-p)^{k-1} = (0,02)(0,98)^{k-1}.$$

2. The probability that this equipment will fail (for the first time) on the tenth attempt:

$$P(X=10) = (0,02)(0,98)^{10-1} = 0,016.$$

5.9 Examples

Exercise 1

The mass function of a random variable X is

$$f(x) = \begin{cases} 2P \text{ if } x = 1 \\ P \text{ if } x = 2 \\ 4P \text{ if } x = 3 \\ 0 \text{ otherwise} \end{cases}$$

1) Determine the value of P.

2) Calculate $P(0 \leq X \leq 3)$ et $P(X > 1,5)$.

Exercise 2

We randomly choose a ball from an urn containing 8 numbered balls:

$$\mathbf{-2}, \mathbf{-1}, \mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3},$$

Let X be the random variable representing the number of the chosen ball.

1) Determine the probability law, the distribution function, the mean of this variable.

2) Repeat question 1, for the random variables $|X|$ and X^2.

Exercise 3

A mouse is placed in a cage. It faces **4** gates, only one of which allows it to exit the cage. For each unsuccessful attempt, the mouse receives an electric shock and is placed back in its original location. It is assumed that the mouse memorizes the unsuccessful attempts and chooses equiprobably between the gates it has not yet tried. Let X be the rv representing the number of attempts to exit the cage.

1. Determine the probability distribution of the $R.V$ X. Recognize the distribution.

2. Calculate $\mathbb{E}(\mathbb{X})$ and $Var(X)$.

Exercise 4

An urn contains **7** balls: **One** red ball, **2** yellow balls and **4** green balls. A player randomly draws a ball, if the ball is red, he wins **2 points**, if it is yellow, he loses **2 points**, if it is green, he draws a second ball from the urn without replacement, if this second ball is red, he wins **1 points**, otherwise he loses **1 points**. Let X be the $R.V.$ associating the player's gain with each draw.

1) Determine the probability law of $R.V.X$.

2) Calculate the expectation and variance of $R.V.X$.

3) The game conditions remain identical. Indicate the amount of gain that must be awarded to a player when the ball drawn in the second draw is red, so that the expectation of $R.V.X$ is **zero**.

Exercise 5

A company produces pens in large quantities. The probability that a pen has a defect is equal to **0.1**. Eight pens are taken from this production, successively and with replacement. We note X the $R.V$ that counts the number of pens with a defect among the eight pens taken.

1. What is the probability distribution of the $R.V\ X$?
2. What is the probability that there is **no pen** with a defect?
3. What is the probability that there is **at least one pen** with a defect?
4. What is the probability that there are **fewer than two pens** with a defect?

Answer

Exercise 1

1. f is a mass function of the $R.V$ X, then its support is $D_X = \{1, 2, 3\}$.

$$\begin{cases} f \text{ is a mass function of the } r.v.X \Leftrightarrow \sum_{x=1}^{3} f(x) = 1. \\ \sum_{x=1}^{3} f(x) = 1 \Leftrightarrow 2p + p + 4p = 1 \Leftrightarrow p = \frac{1}{7} \end{cases}$$

It is easy to deduce that the distribution function of the $R.V$ X is

$$F_X(x) = \begin{cases} 0 \text{ if } x < 1 \\ \frac{2}{7} \text{ if } 1 \leq x < 2 \\ \frac{3}{7} \text{ if } 2 \leq x < 3 \\ 1 \text{ if } x \geq 3 \end{cases}$$

so using the properties of the distribution function we obtain

$$\begin{cases} P(0 \leq X < 3) = F_X(3) - F_X(-0) + p_X(0) = 1 - 0 + 0 - \frac{4}{7} = \frac{3}{7} \\ \text{and} \\ P(X > 1,5) = 1 - F_X(1,6) = 1 - \frac{2}{7} = \frac{5}{7}. \end{cases}$$

Exercise 2

1. The support of X is $D(X) = \{-2, -1, 0, 1, 2, 3,\}$ and its mass function P_X is defined by:

$$\begin{aligned} P(X &= -2) = P(X = -1) = P(X = 0) = P(X = 1) = P(X = 2) = P(X = 3) \\ &= \frac{1}{6}, \end{aligned}$$

that is, X follows uniform law on the set $\{-2, -1, 0, 1, 2, 3\}$ $(X \to U_{\{-2,-1,0,1,2,3\}})$. It is easy to deduce that the distribution function of the $R.V$ X east

$$F_X(x) = \begin{cases} 0 \text{ if } x < -2 \\ \frac{2}{6} \text{ if } -2 \leq x < -1 \\ \frac{3}{6} \text{ if} -1 \leq x < 0 \\ . \\ . \\ 1 \text{ if } x \geq 3 \end{cases}$$

and

$$\mathbb{E}(X) = \frac{1}{6}(-2 - 1 + 0 + 1 + 2 + 3) = \frac{1}{2}.$$

2. We pose $Y = h_1(X) = |X|$ and $Z = h_2(X) = X^2$ and we have:

$$D_Y = h_1(X) = \{0, 1, 2, 3\} \text{ and } D_Z = h_2(Y) = \{0, 1, 4, 9\}$$

and mass functions P_Y and P_Z are given by:

$$\begin{aligned}
P(Y &= 0) = P(X = 0) = \frac{1}{6} \\
P(Y &= 1) = P(X = 1) + P(X = -1) = \frac{2}{6}, \\
P(Y &= 2) = P(X = 2) + P(X = -2) = \frac{2}{6}, \\
P(Y &= 3) = P(X = 3) + P(X = -3) = \frac{1}{6},
\end{aligned}$$

et

$$\begin{aligned}
P(Z &= 0) = P(X = 0) = \frac{1}{6} \\
P(Z &= 1) = P(X = 1) + P(X = -1) = \frac{2}{6}, \\
P(Z &= 4) = P(X = 2) + P(X = -2) = \frac{2}{6}, \\
P(Y &= 9) = P(X = 3) = \frac{1}{6},
\end{aligned}$$

Distribution functions F_Y and F_Z of Y and Z are defined by

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ \frac{1}{6} & \text{si } 0 \leq y < 1 \\ \frac{2}{6} & \text{si } 1 \leq y < 2 \\ \frac{3}{6} & \text{si } 2 \leq y < 3 \\ 1 & \text{si } 3 \leq y \end{cases}$$

et

$$F_Z(z) = \begin{cases} 0 & \text{if } y < 0 \\ \frac{1}{6} & \text{if } 0 \leq z < 1 \\ \frac{2}{6} & \text{if } 1 \leq z < 4 \\ \frac{4}{6} & \text{if } 4 \leq z < 9 \\ 1 & \text{if } z \geq 9. \end{cases}$$

the expectation of the $R.V$ are given by:

$$\mathbb{E}(Y) = 0 \times \frac{1}{6} + (1 + 2) \times \frac{2}{6} + 3 \times \frac{1}{6} = \frac{3}{2}.$$

and
$$\mathbb{E}(Z) = 0 \times \frac{1}{6} + (1+4) \times \frac{2}{6} + 9 \times \frac{1}{6} = \frac{19}{6} = 3,16.$$

Exercise 3
Let us determine the probability law of $R.V\ X$: We have:

$$\begin{cases} P(X=1) = \frac{1}{4} = P_1 \\ P(X=2) = (1-P_1)\frac{1}{3} = \frac{3}{4} \times \frac{1}{3} = \frac{1}{4} \\ P(X=3) = (1-P_1)(1-P_2) \times \frac{1}{2} = \frac{3}{4}\frac{2}{3}\frac{1}{2} = \frac{1}{4} \\ P(X=4) = (1-P_1)(1-P_2)(1-P_3) = \frac{3}{4}\frac{2}{3}\frac{1}{2}1 = \frac{1}{4}. \end{cases}$$

The law of probability of the $R.V\ X$:

x_i	1	2	3	4	$\sum$
$P(X=x_i)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	1

Since $P(X=1) = P(X=2) = P(X=3) = P(X=4) = \frac{1}{4}$, we deduce that the $R.V\ X \to U_{\{1,2,3,4\}}$

Let's calculate:
$$\begin{cases} \mathbb{E}(X) = \frac{n+1}{2} = \frac{5}{2} \\ Var(X) = \frac{n^2+1}{12} = \frac{5}{4} \end{cases}$$

Exercise 4
1. Let us determine the probability law of the $R.V\ X$:

$$D_X = \{-2,-1,1,2\}.$$

and

$$\begin{aligned} P(X=-2) &= \frac{2}{7},\ P(X=-1) = P(V_1).P(J_2 \text{ or } V_2 |V_1) \\ &= P(V_1)[P(J_2|V_1) + P(V_2|V_1)] \\ &= \frac{4}{7} \times \frac{2}{6} + \frac{4}{7} \times \frac{3}{6} = \frac{10}{21} \text{ or } J \text{ or } R \text{ is Yellow or Red} \end{aligned}$$

and

$$\begin{aligned} P(X=1) &= P(V).P(J \text{ or } V|V) \\ &= \frac{4}{7} \times \frac{5}{6} = \frac{2}{21} \text{ or } J \text{ or } V \text{ is Yellow or Green,} \\ P(X=2) &= \frac{1}{7} \end{aligned}$$

Summary of the law

$X(\Omega)=x_i$	-2	-1	1	2	$\sum$
$P(X=x_i)$	$\frac{2}{7}$	$\frac{10}{21}$	$\frac{2}{21}$	$\frac{1}{7}$	1

2. Let us calculate the mathematical expectation of the $R.V.X$ as well as the variance $Var(X)$:

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{4} x_i P(X=x_i) = \frac{-14}{21} = -0,66 \text{ and} \\
Var(X) &= \mathbb{E}(X^2) - \mathbb{E}^2(X) \text{ such as:} \\
\mathbb{E}(X^2) &= \sum_{i=1}^{4} x_i^2 P(X=x_i) = \frac{36}{21} = 1,71 \\
\text{From where } Var(X) &= 1,71 - (0,66)^2 = 1,27
\end{aligned}$$

Let us denote "α" the algebraic gain corresponding to the event $V_1 \cap R_2$:
We have:

$$P(V_1 \cap R_2) = P(V_1).P(R_2 | V_1) = \frac{4}{7} \times \frac{1}{6} = \frac{2}{21}.$$

We therefore obtain

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{i=1}^{4} x_i P(X=x_i) = (-2) \times \frac{2}{7} + (-1) \times \frac{10}{21} + \alpha \times \frac{2}{21} + 2 \times \frac{1}{7} \\
&= \frac{2\alpha - 16}{21}
\end{aligned}$$

just solve the equation:

$$\mathbb{E}(X) = 0 \Leftrightarrow 2\alpha - 16 = 0 \Leftrightarrow \alpha = 8 \text{ points.}$$

Exercise 5
1. The probability law of the $R.V.X$:

$$X = \sum_{i=1}^{8} X_i \text{ such as } X_i \to B(0,1)$$

So, we deduce that the $R,V\ X$ follows a Binomial law with parameters $n=8$ and $P=0,1$, that is $X \to B(8,0,1)$;

$$P(X=k) = C_8^k (0,1)^k (0,8)^{8-k} ; k \in \{0,1,...,8\}$$

2. The probability that there is no pen with a defect:

$$P(X=0) = C_8^0 (0,1)^0 (0,8)^8 = 0,43$$

3. The probability that there is at least one pen with a defect:

$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0,43 = 0,57$$

4. The probability that there are less than two pens with a defect:

$$P(X < 0) = P(X = 0) + P(X = 1) \approx 0,813.$$

6 Conditional Expectation and prediction

Remark 39 *Let $Z = (X, Y)$ a pair of real random variables such that $X(\Omega) = \{x_1, ..., x_n\}$ and $Y(\Omega) = \{y_1, ..., y_p\}$. So, the family of events*

$$([X = x_i] \cap [Y = y_i])_{\substack{1 \le i \le n \\ 1 \le j \le p}}$$

forms a complete event system of Ω.

Remark 40 *The event $[X = x_i]$ and $[Y = y_j]$ can also be noted $[X = x_i, Y = y_j]$ or even $[(X, Y) = (x_i, y_j)]$.*

6.1 Laws associated with a pair of random variables

The pair of real random variables (X, Y) is confused with a probability law $P_{(X,Y)}$ on $\Omega = \mathbb{R}^2$. In this case, F is constructed with the "blocks" $(a, b) \times (c, d).(X, Y)$ represents a random experiment whose outcome is a stream of reals. The law of the pair (X, Y) is used to characterize the laisons and mutual influences of the two characters in experimentec. Note

$$P_{(X,Y)}(\Delta) = P((X, Y) \in \Delta)$$

The probability that the results of the experiment belong to Δ. In the case where $\Delta = \Delta_1 \times \Delta_2$ we can write down the previous quantity

$$P_{(X,Y)}(\Delta) = P(X \in \Delta_1, Y \in \Delta_2)$$

6.2 Marginal laws

The law of a pair is associated with two marginal laws, which are the laws of each of the elements of the pair taken separately, defined by the set of possible values and the associated probabilities obtained by summation:

$$\begin{aligned} P_X(X = x_i) &= \sum_{j \in J} P(X = x_i, Y = y_j) = \sum_{j \in J} P_{ij} = P_{i.} \\ P_Y(Y = y_j) &= \sum_{i \in I} P(X = x_i, Y = y_j) = \sum_{i \in I} P_{ij} = P_{.j} \end{aligned}$$

If the couple's law is presented in a table, these laws are obtained in the margins, by row or column summation

Y/X	x_i	
.	.	.
.	.	.
y_j	P_{ij}	$P_{.j}$
.	.	.
.	.	.
	$P_{i.}$	1

6.3 Conditional laws

We can also associate two conditional laws with the law of a pair, i.e. the law of one variable, the other having a fixed value (law in a given row or column). For example, for a fixed $Y = y_j$, the conditional law of X is defined by the set of possible values and the associated probabilities

$$P\left(X = x_i \middle| Y = y_j\right) = \frac{P\left(X = x_i, Y = y_j\right)}{P\left(Y = y_j\right)} = \frac{P_{ij}}{P_{.j}} = P_i^j$$

We check that it is a probability law on

$$\Omega_X = \{x_i; i \in I\} : \sum_{i \in I} P_i^j = \frac{1}{P_{.j}} \sum_{i \in I} P_{ij} = 1$$

Example 41 *The law of a pair (X, Y) is given by the following table*

Y/X	-2	0	2	
-1	$0,1$	$0,2$	$0,1$	$0,4$
2	$0,2$	$0,2$	$0,2$	$0,6$
	$0,3$	$0,4$	$0,3$	1

The conditional distribution of X for $Y = -1$ is shown in the table below

$X\| Y = -1$	-2	0	2	
	$\frac{0,1}{0,4}$	$\frac{0,2}{0,4}$	$\frac{0,1}{0,4}$	1

Recall that the two random variables X and Y are independent if for all $i \in I$ and $j \in J$:

$$P\left(X = x_i, Y = y_j\right) = P\left(X = x_i\right) P\left(Y = y_j\right) \left(P_{ij} = P_{i.} P_{.j}\right)$$

In this case, of course, conditional laws are confused with marginal laws; for example

$$P\left(X = x_i \middle| Y = y_j\right) = P_i^j = \frac{P_{i.} P_{.j}}{P_{.j}} = P_{i.}$$

This is one of the few cases where the marginal laws can be used to reconstruct the law of the couple.

Definition 42 *Conditional laws are associated with moments, such as the expectation of the law defined by the pairs $\left\{\left(y_j, P_j^i; j \in J\right)\right\}$, or:*

$$\mathbb{E}\left(Y \middle| X = x_i\right) = \sum_{j \in J} y_j P\left(Y = y_j \middle| X = x_i\right) = \sum_{j \in J} y_j P_j^i.$$

Remark 43 *The plot of this conditional expectation as a function of x_i is called the **(non-linear) regression** curve from Y to X.*

Example 44 *In the previous example, the conditional distribution of Y for $X = 2$ is given by the following table*

$Y \setminus X = 2$	-1	2	
	$\frac{0,1}{0,3}$	$\frac{0,2}{0,3}$	1

From this table, we can calculate the conditional expectation

$$\mathbb{E}\left(Y \mid X = 2\right) = (-1)\frac{1}{3} + 2\left(\frac{2}{3}\right) = 1.$$

Remark 45 *Note that $\mathbb{E}\left(Y \mid X\right)$ is a function of X, so is a discrete random variable whose probability law is defined by the set of possible values, in this case $\{\mathbb{E}\left(Y \mid X = x_i\right); i \in I\}$ and the associated probabilities $P_{i.} = P\left(X = x_i\right)$. We can therefore calculate the mean value of this random variable, i.e*

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}\left(Y \mid X\right)\right] &= \sum_{i \in I} P_{i.} \mathbb{E}\left(Y \mid X = x_i\right) \\
&= \sum_{i \in I} P_{i.} \sum_{j \in J} y_j P\left(Y = y_j \mid X = x_i\right) \\
&= \sum_{i \in I} P_{i.} \sum_{j \in J} y_j P_j^i \\
&= \sum_{i \in I} \sum_{j \in J} y_j P_{i.} \frac{P_{ij}}{P_{i.}} \\
&= \sum_{j \in J} y_j \sum_{i \in I} P_{ij} \\
&= \sum_{j \in J} y_j P_{.j} = \mathbb{E}\left(Y\right).
\end{aligned}$$

6.4 Moments associated with a couple

If h:$\mathbb{R}^2 \to \mathbb{R}$ is a continuous application, it defines a real random variable whose moments, such as its expectation, can be calculated:

$$\mathbb{E}\left(h\left(X, Y\right)\right) = \sum_{i \in I} \sum_{j \in J} P_{ij} h\left(x_i, y_j\right)$$

In the special case where $h\left(X, Y\right) = \left[X - \mathbb{E}\left(X\right)\right]\left[Y - \mathbb{E}\left(Y\right)\right]$, we define the covariance of X and Y

$$\begin{aligned}
Cov\left(X, Y\right) &= \mathbb{E}\left\{\left[X - \mathbb{E}\left(X\right)\right]\left[Y - \mathbb{E}\left(Y\right)\right]\right\} \\
&= \mathbb{E}\left(XY\right) - \mathbb{E}\left(X\right)\mathbb{E}\left(Y\right)
\end{aligned}$$

If X and Y are independent, then $\mathbb{E}\left(XY\right) = \mathbb{E}\left(X\right)\mathbb{E}\left(Y\right)$ and consequently $Cov\left(X, Y\right) = 0$.

Example 46 *Consider the pair (X, Y) whose distribution is defined by the following table*

$Y \backslash X$	-1	0	1
-1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

The distributions of X and Y are symmetrical with respect to 0*, so* $\mathbb{E}(X) = 0 = \mathbb{E}(Y)$ *and*

$$Cov\,(X, Y) = \mathbb{E}\,(XY) = 1 \times \frac{1}{8} + (-1)\,\frac{1}{8} + (-1)\,\frac{1}{8} + 1 \times \frac{1}{8} = 0,$$

and yet these two variables are not independent since, for example

$$P\,(X = -1, Y = -1) = \frac{1}{8} \neq P\,(X = -1)\,P\,(Y = -1) = \frac{5}{16} \times \frac{3}{8}$$

$$Var\,(X) = \sigma_X^2 \text{ and } Var\,(Y) = \sigma_Y^2$$

We'll establish that this coefficient lies between -1 and $+1$, due to the Schwarz inequality

$$|\mathbb{E}\,(XY)| \leq \sqrt{\mathbb{E}\,(X^2)}\sqrt{\mathbb{E}\,(Y^2)}$$

which we obtain by considering the polynomial in λ, always positive

$$\mathbb{E}\,(X - \lambda Y)^2 = \lambda^2 - \mathbb{E}\left(Y^2\right) - 2\lambda\mathbb{E}\,(XY) + \mathbb{E}\left(X^2\right) \geq 0$$

which implies $\mathbb{E}^2\,(XY) - \mathbb{E}\left(X^2\right)\mathbb{E}\left(Y^2\right) \leq 0$, or by applying this inequality to centered variables

$$|Cov\,(X, Y)| \leq \sigma_X \sigma_Y$$

and therefore $|\rho| \leq 1$. The case $|\rho| = 1$ corresponds to the existence of an affine relationship between the two variables

$$|\rho| = 1 \iff \exists a \in \mathbb{R}^*,\ \exists b \in \mathbb{R};\ Y = aX + b.$$

This affine relationship, when it exists, can be written as follows

$$Y = \mathbb{E}\,(Y) + \frac{Cov\,(X, Y)}{Var\,(X)}\,[X - \mathbb{E}\,(X)]\,.$$

7 Convergences

In this section we will study two types of convergence of such sequences, in probability and in law.

We consider sequences of random variables $(X_n)_{n\in\mathbb{N}}$ where the variables X_n defined on the same probability space $(\Omega, \mathcal{F}, P)$, that is to say they concern the result of a sequence of random experiments all of the same type.

7.1 Inequalities

7.2 Markov inequality

If X is a positive random variable $(R.V)$ whose expectation exists, Markov's inequality states that for all $\lambda > 0$

$$P\left(X \geq \lambda \mathbb{E}\left(X\right)\right) \leq \frac{1}{\lambda} \text{ or } P\left(X \geq \lambda\right) \leq \frac{\mathbb{E}\left(X\right)}{\lambda}$$

7.3 Bienaymé-Tchebychev inequality

We obtain the Bienaymé-Tchebychev inequality by applying the Markov inequality (for all k such that $\mathbb{E}(X^k)$ exists: $P\left(\left|X^k\right| \geq \varepsilon\right) \leq \dfrac{\mathbb{E}\left|X\right|^k}{\varepsilon^k}; \varepsilon > 0)$ to the v.a $X - \mathbb{E}(X)$ for $k = 2$, therefore for a $R.V$ whose variance exists, that is for all $\varepsilon > 0$ fixed

$$P\left(\left(\left|X - \mathbb{E}\left(X\right)\right|\right) \geq \varepsilon\right) \leq \frac{Var\left(X\right)}{\varepsilon^2}.$$

Example 47 *Let* $Y \to N\left(m, \sigma\right)$.

$$\begin{aligned} P\left(\left|Y - m\right| \geq \varepsilon\right) &= 1 - P\left(\left|Y - m\right| < \varepsilon\right) \\ &= 1 - P\left(m - \varepsilon < Y < m + \varepsilon\right) \\ &= 1 - \left[\Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(\frac{-\varepsilon}{\sigma}\right)\right] = 2\left(1 - \Phi\left(\frac{\varepsilon}{\sigma}\right)\right) \end{aligned}$$

Let us take for example the case where:

$$\varepsilon = 4, \sigma = 2.5. \textit{ So } \frac{\varepsilon}{\sigma} = 1.6 \textit{ and } \Phi\left(\frac{\varepsilon}{\sigma}\right) = 0.9452.$$

So in this case

$$P\left(\left|Y - m\right| \geq \varepsilon\right) = 0.1096.$$

let us compare with the increase obtained by the Bienaymé-Tchebychev inequality:

$$\frac{\sigma^2}{\varepsilon^2} = 0.3906.$$

it is indeed an upper bound of the calculated probability.

Example 48 *Let* $Y \to B\left(20, \dfrac{1}{2}\right)$. *So* $\sigma = \sqrt{5}$ *and* $\mathbb{E}(Y) = 10$. *Let's take for example* $\varepsilon = 3$:

$$\begin{aligned} P(|Y-10| \geq 3) &= 1 - P(7 < Y < 13 < \varepsilon) \\ &= 1 - \sum_{k=8}^{12} P(Y=y) = 1 - \left(\frac{1}{2}\right)^{20} \sum_{k=8}^{12} C_{20}^{k} \\ &= 1 - 0.7368 = 0.2632. \end{aligned}$$

Or $\dfrac{\sigma^2}{\varepsilon^2} = \dfrac{5}{9} = 0.555.$

The Bienaymé-Tchebychev inequality is therefore verified.

Remark 49 *The upper bound obtained with this inequality is often too large, but it is universal. In some problems it is a question of improving this inequality, that is to say of finding a smaller upper bound of the probability in question, in the case of random variables of a particular type.*

8 Weak law of large numbers

8.1 Convergence in probability

The definition of convergence in probability involves a numerical sequence of probabilities whose convergence will often be established using the **Bienaymé-Tchebychev** ($B.T$) inequality, which links a probability and a variance.

If (X_n) is a sequence of $R.V$ that converges to a $R.V$ X, this means that X_n **"gets closer"** to X when n increases. We measure the distance between X_n and X by $|X_n - X|$ which will be all the smaller as n is large; but, concerning $R.V$, we must consider the event $|X_n - X| < \varepsilon$ which will be realized with a probability all the higher as n is large. We will therefore associate with the random sequence (X_n) the numerical sequence of the probabilities of these events, which must converge to one.

Definition 50 *We say that the sequence of random variables* (X_n) *converges in probability to a variable* X *if, for all* $\varepsilon > 0$

$$P\left[(|X_n - X|) < \varepsilon\right] \to 1 \text{ When } n \to \infty$$

or, equivalently

$$P\left[(|X_n - X|) > \varepsilon\right] \to 0 \text{ When } n \to \infty$$

we write

$$X_n \xrightarrow{p} X.$$

In the general case, it is difficult to evaluate $P\left[(|X_n - X|) > \varepsilon\right]$, *and therefore its limit. On the other hand, when* X *is a certain random variable of value* C, *we have*

$$P\left[(|X_n - X|) > \varepsilon\right] = P\left[(|X_n - C|) > \varepsilon\right] = 1 - P\left(C - \varepsilon < X_n < C + \varepsilon\right).$$

The interesting case is that of a sequence of variables X all with the same expectation m. We then study the convergence in probability of this sequence towards the certain variable with values m.

Theorem 51 *(weak law of large numbers) Let $(X_n)_{n\in N^*}$ be a sequence of two-by-two independent random variables, with the same expectation m and the same standard deviation.σ*
We set $S_n = \sum_{i=1}^{n} X_i$ and $\overline{X}_n = \frac{S_n}{n}$ (mean of the n first X_i). Then the sequence $\left(\overline{X}_n\right)_{n\in N^*}$ converges in probability towards the certain variable of value m.

Indeed

To apply the inequality **B.T**. to the variable $\overline{X_n}$, we calculate its expectation and its variance.

$$\mathbb{E}\left(S_n\right) = \sum_{i=1}^{n} \mathbb{E}\left(X_i\right) = nm,$$

from where

$$\mathbb{E}\left(\overline{X_n}\right) = \mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\mathbb{E}\left(S_n\right) = m.$$

$$Var\left(S_n\right) = \sum_{i=1}^{n} Var\left(X_i\right) = n\sigma^2,$$

because the variables X are independent two by two, where

$$Var\left(\overline{X_n}\right) = Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2}Var\left(S_n\right) = \frac{\sigma^2}{n}.$$

According to **B.T**.'s inequality,

$$\forall \varepsilon > 0 : P\left(\left(\left|\overline{X}_n - m\right|\right) \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2},$$

therefore according to the "**gendarmes' theorem**", a probability being a positive number,

$$\forall \varepsilon > 0 : \lim_{n\to+\infty} P\left(\left(\left|\overline{X}_n - m\right|\right) \geq \varepsilon\right) = 0,$$

hence the result.

Example 52 ***Case of a sequence of Bernoulli variables.***
*If the variables X all follow the same Bernoulli distribution of parameter p, while maintaining the independence hypothesis, the variable S_n follows a **Binomial distribution** of parameters n and p: it represents the number of successes out of n attempts of an event of probability p. We can then say that X represents **the frequency of success out of n attempts**, and the sequence X converges in probability towards the certain variable of value p. We can specify this fact with **Bernoulli's Theorem**:*

Theorem 53 *In a succession of n independent Bernoulli trials, such that the probability of success is p, the frequency $\overline{X_n}$ verifies*

$$\forall \varepsilon > 0 \ P\left(\left|\overline{X}_n - p\right| > \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}$$

and the sequence $\left(\overline{X}_n\right)$ converges in probability towards the certain variable of value p.

Remark 54 *This theorem justifies the fact that one can have an approximate value of the probability of success, by taking the frequency of success over a large number of attempts. For example, to see if a coin is fair, one tosses it a large number of times, and the frequency of "**heads**" must be close to $\frac{1}{2}$.*

9 Convergence in law

Convergence in law is a convergence of the laws of random variables, without taking into account the behavior of the sequence for a fixed eventuality.

Definition 55 *We say that the sequence of random variables $(X_n)_{n\geq 1}$ converges in law towards the random variable X when n tends towards infinity, and we write $X_n \underset{n\to\infty}{\overset{law}{\to}} X$, if for any function $f : \mathbb{R} \to \mathbb{R}$ bounded continuous*

$$\mathbb{E}\left(f\left(X_n\right)\right) \underset{n\to\infty}{\to} \mathbb{E}\left(f\left(X\right)\right)$$

This definition is not necessarily the easiest to handle, and there are many equivalent definitions of this convergence.

Proposition 56 *The convergence in law of X_n towards X is equivalent to each of the following properties:*
1. Simple convergence of characteristic functions

$$\varphi_{X_n}\left(t\right) \underset{n\to\infty}{\to} \varphi_X\left(t\right) \ \ \forall t \in \mathbb{R}$$

2. Simple convergence of distribution functions at points of continuity

$$F_{X_n}\left(x\right) \to F_X\left(x\right)$$

for all x such that F_X is continuous in x.

Remark 57 *The following lemma follows immediately from the definition of convergence in law.*

Lemma 58 *If the following $(X_n)_{n\geq 1}$ converges in law to X and if: $f : \mathbb{R} \to \mathbb{R}$ is a continuous function, then $f\left(X_n\right)$ converges in law to $f\left(X\right)$.*

Theorem 59 *Let* $(X_n)_{n\geq 1}$ *a sequence of independent real random variables with the same distribution, with mean* m *and variance* σ^2 *finished, either*

$$S_n = X_1 + ... + X_n.$$

So

$$\frac{S_n - nm}{\sigma\sqrt{n}} \underset{n\to\infty}{\overset{law}{\to}} S \underset{n\to\infty}{\overset{law}{\to}} N(0,1)$$

that's to say

$$P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \leq t\right) \to \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) dx = \phi(t) \ \ \forall t \in \mathbb{R}$$

Remark 60 *The distribution function* ϕ *of the normal law* $N(0,1)$ *above all* $\mathbb{R}$*.*

Example 61 *Let* $X_n \to \mathcal{U}\left(\left\{\frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n}\right\}\right)$*: So* $P\left(X_n = \frac{k}{n}\right) = \frac{1}{n}$.
for all real $x \in [0,1]$ *,*

$$F_n(x) = \sum_{\frac{k}{n} \leq x} \frac{1}{n} = \sum_{k \leq nx} \frac{1}{n} = \frac{1}{n}[nx],$$

or $[y]$ *denotes the integer part of* y*, for any real* y*.Or we know that*

$$[nx] \leq nx < [nx] + 1,$$

from where

$$nx - 1 < [nx] \leq nx,$$

and therefore

$$x - \frac{1}{n} < \frac{1}{n}[nx] \leq x.$$

According to the gendarmes' theorem we therefore have

$$\lim_{n\to+\infty} \frac{1}{n}[nx] = x$$

that's to say

$$\forall x \in [0,1] \qquad \lim_{n\to+\infty} F_n(x) = x.$$

Let $F(x) = x$ *on* $[0,1]$*, and* $F(x) = 0$ *if* $x < 0$*,* $F(x) = 1$ *if* $x > 1$*. So* F *is the distribution function of a variable* X *following a uniform distribution on* $[0.1]$*: The sequence* $(X_n)_{n\geq 1}$ *converges in distribution (**converges in law**)towards a variable* $X \to U([0,1])$ *.*

10 Approximations

10.1 Approximation of the Binomial distribution by the Poisson distribution

The binomial distribution depends on two parameters n and p, while the Poisson distribution depends on only one parameter. For a binomial distribution to be as close as possible to a Poisson distribution, we must at least hope that these two distributions have the same expectation. The expectation of the binomial distribution being $n \times p$ and that of the Poisson distribution being ■ , it is necessary that $\lambda = np$. This necessary condition is not sufficient to achieve such an approximation, theoretically the approximation is perfect when:

$$\begin{cases} n \to +\infty \\ p \to 0 \\ np = \text{constant} \end{cases}$$

In practice, the condition

$$\begin{cases} n \geq 30 \\ np \leq 5. \end{cases}$$

or

$$\begin{cases} n \geq 50 \\ p \leq 0, 1. \end{cases}$$

is sufficient to consider the approximation.

Example 62 *We consider a binomial distribution with parameters* $n = 35$ *and* $p = 0.1$. *We are in the conditions of approximation of this distribution by a Poisson distribution with parameter*

$\lambda = 0.1 \times 35 = 3.5.$

10.2 Approximation of a hypergeometric law by a binomial law

Let $X_N \to H(N, n, p)$, for any integer N such that N_p is an integer. (We note $J = \left\{ \dfrac{N}{N_p} \in N \right\}$). the sequence $(X_n)_{n \in J}$ converges in law towards a random variable $X \to B(n, p)$.

Remark 63 *We make this approximation from* $N > 10n$.

11 Central Limit Theorem

Theorem 64 *Let* (X_n) *a sequence of independent random variables with the same distribution, with expectation* m *and standard deviation* σ. *We pose*

$$S_n = \sum_{i=1}^{n} X_i,\ \overline{X}_n = \frac{S_n}{n},\ Z_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\overline{X}_n - m}{\frac{\sigma}{\sqrt{n}}}.$$

Then the sequence (Z_n) converges in law towards a variable $T \to N(0,1)$.

We know that $\mathbb{E}(S_n) = nm$, $\mathbb{E}(\overline{X}) = m$, $Var(\overline{X}_n) = \dfrac{\sigma^2}{n}$. We notice that

$$Z_n = \frac{S_n - nm}{\sigma\sqrt{n}} = \frac{\overline{X}_n - m}{\frac{\sigma}{\sqrt{n}}}.$$

Z_n is therefore the reduced centered variable associated with both S_n and $\overline{X}_n$.

Remark 65 *The sum of n independent variables and of the same law can be approximated, for n large enough, by a variable of normal law, whose parameters are the expectation and the variance of this sum.*
The average of n independent variables and of the same law can be approximated, for n large enough, by a variable of normal law, whose parameters are the expectation and the variance of this average.

11.1 Approximation of a Binomial distribution by a Normal distribution

If the variables X_i are independent Bernoulli variables of parameter p, the sums S_n follow a binomial distribution of parameters n and p, of expectation np, of variance npq, with $q = 1 - p$. The application of the central limit theorem to the sequence S_n gives:

Theorem 66 *Let $X \to B(n,p)$. For n large enough the law of X can be approximated by the normal law $N(np, npq)$.*

In practice: this approximation is done as soon as $n > 20$, with an average parameter p.

11.2 Law of Frequency of Success

Similarly, with the same notations, the variable $F_n = \dfrac{S_n}{n}$ which gives the frequency of success of an event E of probability p on n identical and independent draws is the mean $\overline{X}$ of the n Bernoulli variables $X_i \to B(p)$.

The central limit theorem then gives:

Theorem 67 *Let F be the frequency of the success of an event E of probability p on n identical and independent draws. For n large enough the law of F can be approximated by the normal law $N\left(p, \dfrac{pq}{n}\right)$.*

11.3 Approximation of a Poisson distribution by a Normal distribution

If the variables X_n of the initial sequence all follow a Poisson distribution of parameter p, and if they are independent, their sums S_n follow a Poisson distribution of parameter np. The expectation and variance of S_n are both equal to np. According to the central limit theorem, the general sequence converges in distribution to a reduced centered normal distribution variable, which allows the following approximation:

Theorem 68 *Let $X \rightarrow \mathcal{P}(\lambda)$ be. The law of X can be approximated by the normal law $N(\lambda, \lambda)$.*

Remark 69 *In practice: we can make this approximation from $\lambda > 15$.*

11.4 Examples

Discrete Usual Laws

Bernoulli (p) with $p \in [0;1]$: $P(X=0) = 1 - p$ and $P(X=1) = p$.

Binomiale (n,p) with $n > 0$ and $p \in [0;1]$: $P(X=k) = C_n^k p^k (1-p)^{n-k}$ for $k = 0, ..., n$.

Géométrique(p) with $p \in [0;1]$: $P(X=k) = p(1-p)^{k-1}$ for $k \in \mathbb{N}^*$?.

Poisson(λ) with $\lambda > 0$: $P(X=k) = \exp(-\lambda)\dfrac{\lambda^k}{k!}$ for $k \in \mathbb{N}$.

Exercise 1

A certain equipment has a probability $p = 0.02$ of failure each time it is put into service. The following experiment is carried out: the equipment is started, stopped, started again, stopped, until it breaks down. Let X be the $r.v.$ representing the number of trials required to obtain the failure.

1. What is the probability distribution of the $r.v.X$?

2. What is the probability that this equipment breaks down (for the first time) on the tenth trial?

Exercise 2

In a marble quarry, an inspection is carried out on slabs intended for construction. The surface of the slabs is checked to detect any chips or stains. It was found that on average there are $1,2$ defects per slab and that the number of defects per slab follows a **Poisson** distribution.

1. What is the parameter of this variable? What are the possible values of the variable?

2. What is the probability of observing more than 2 defects per slab?

3. The company presents its customers with two categories of slabs: those with less than two defects (**quality ***) and those with at least two defects (**quality ****). What is the probability of observing at least two defects on a slab? What is then the proportion of slabs of quality **?

4. Out of 500 slabs inspected, what is the expected number with no defects?

Exercise 3. Parts 1 and 2 are independent

Part 1

Let X,Y be two independent $r.a.$ of Poisson distributions with respective parameters λ and μ.

Determine the conditional distribution of X when the sum $S = X + Y$ has a fixed value $S = s$. Deduce the expression of the regression function of X on S then the value of $\mathbb{E}[\mathbb{E}(X/S)]$

Part 2

The following model can be used to represent the number of injured people in traffic accidents over a weekend. The number of accidents follows a Poisson distribution with parameter λ. The number of injured people per accident, follows a Poisson distribution with parameter m. The total number of injured people is therefore

$$S\,|N = n = X_1 + X_2 + \cdots + X_N$$

such that $S\,|N = n \to \mathcal{P}(nm)$

S is the sum of a random number of Poisson variables, independent and with the same distribution.

1. Give an expression for $P(S = s)$.
2. Calculate $P(S = 0)$.
3. Calculate $\mathbb{E}(S)$ and $V(S)$.

Exercise 4

A laptop manufacturer wants to check that the warranty period it must associate with the hard drive corresponds to a not too large number of returns of this component under warranty. Laboratory tests have shown that the law followed by the life span, in years, of this component is the exponential law with an average $\mathbf{4}$.

1. Specify the distribution function of this law as well as its expected value. $\mathbb{E}(X)$ and are standard deviation .
2. What is the probability that a hard drive will operate without failure for more than four years?
3. What is the probability that a hard drive will operate without failure for at least six years, given that it has already operated for five years?
4. What is the probability that the lifetime belongs to the interval: $[\mathbb{E}(X) - \sigma, \mathbb{E}(X) + \sigma]$?
5. How long do 50% of hard drives operate without failure?
6. Give the optimum warranty period to replace less than 15% of hard drives under warranty.

Exercise 5

The distance (in meters) traveled by a projectile follows a normal distribution. During training, we find that:

- The probability that a projectile exceeds $\mathbf{60}$ meters is $\mathbf{0.0869}$.
- The probability that a projectile travels a distance less than $\mathbf{45}$ meters is $\mathbf{0.6406}$.
- Calculate the average distance traveled by a projectile, as well as its standard deviation.

Answer

Exercise 1

1. The probability law of the $R.V..$ X:

The random variable X follows a geometric distribution with parameter $\mathbf{p = 0,02}$, we write $X \to G(0,02)$, and we have

$$P(X = k) = p(1-p)^{k-1} = (0,02)(0,98)^{k-1}; k \in \{1, 2, ...\}.$$

2.The probability that this equipment will fail (for the first time) on the tenth attempt

$$P(X = 10) = (0,02)(0,98)^{10-1} = 0,016.$$

Exercise 2

1. Let X be the variable: **"number of defects per slab"**. The distribution of X is a **Poisson** distribution. Its parameter is equal to the mean observed on the sample: $\lambda = 1,2$. The possible valuesof X are positive integers

2.

$$\begin{aligned} P(X &> 2) = 1 - P(X \leq 2). \\ \text{Or } P(X &= 0) = e^{-1,2}, P(X = 1) = 1,2 \times e^{-1,2}, P(X = 2) = e^{-1,2} \times \frac{1,2^2}{2!} \\ P(X &= 0) = 0,301 \ , P(X = 1) = 0,361 \ , P(X = 2) = 0,217. \\ P(X &> 2) = 1 - P(X \leq 2) = 1 - 0,879 = 0,122 \end{aligned}$$

3. The probability of observing at least two defects on a slab is then

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - e^{-1,2} - 1,2 \times e^{-1,2} = 0,338.$$

The proportion of quality ** slabs is therefore $33,8\%$.

4. Of the **500** slabs checked, the expected number showing no defects is

$$500 \times P(X = 0) \approx 150.$$

Exercise 3

Part 1

X and Y are two independent **Poisson** such that

$$X \to \mathcal{P}(\lambda) \text{ and } Y \to \mathcal{P}(\mu)$$

Let us determine the conditional law of $X/S = X + Y$?

We know that

$$S \to \mathcal{P}(\lambda + \mu)$$

and therefore for

$$\begin{aligned} 0 &\leq x \leq s, \ P(X = x/S = s) = \frac{P(X = x, S = s)}{P(S = s)} \\ &= \frac{P(X = x) P(S = s/X = x)}{P(S = s)} \\ &= \frac{P(X = x) P(Y = s - x)}{P(S = s)} \end{aligned}$$

by calculation we find

$$\begin{aligned}
\frac{P(X=x)\,P(Y=s-x)}{P(S=s)} &= \frac{s!}{x!\,(s-x)!} \times \frac{\lambda^x \mu^{s-x}}{(\lambda+\mu)^s} \\
&= C_s^x \left(\frac{\lambda}{\lambda+\mu}\right)^x \left(1-\frac{\lambda}{\lambda+\mu}\right)^{s-x} \\
&= C_s^x P^x (1-P)^{s-x}
\end{aligned}$$

We deduce that $X/S = s \to B\left(s, \dfrac{\lambda}{\lambda+\mu}\right)$ so of mean

$$\mathbb{E}(X/S=s) = s\frac{\lambda}{\lambda+\mu} \text{ constant.}$$

So

$$\mathbb{E}(X/S) \text{ is a random variable}$$

such as

$$\begin{aligned}
\mathbb{E}(X/S) &= S\frac{\lambda}{\lambda+\mu} \text{ is a random variable and} \\
\mathbb{E}(\mathbb{E}(X/S)) &= \mathbb{E}\left(S\frac{\lambda}{\lambda+\mu}\right) = \frac{\lambda}{\lambda+\mu}\mathbb{E}(S) = \frac{\lambda}{\lambda+\mu} \times \lambda+\mu = \lambda.
\end{aligned}$$

Part 2

Remark 70 *What to remember from this exercise The application of the theorems of total expectation and total variance finds its full meaning here.*

1. If we know the number of accidents over the weekend, we can then know the number of injured people over the **weekend** using the sum of Poisson variables: we will therefore use the conditional probability law

$$\begin{cases}
P(S=s/N=n) = \dfrac{e^{-\mu n}(\mu n)^s}{s!} \Rightarrow P(S=s) = \sum_{n=0}^{\infty} \dfrac{e^{-\mu n}(\mu n)^s}{s!} \dfrac{e^{-\lambda}\lambda^n}{n!} \\
P(S=s) = \dfrac{e^{-\lambda}\mu^s}{s!} \sum_{n=0}^{\infty} \dfrac{\lambda^n n^s e^{-\mu n}}{n!}.
\end{cases}$$

2.
$$P(S=0) = e^{-\lambda}\sum_{n=0}^{\infty} \frac{\lambda^n e^{-\mu n}}{n!} = \exp[-\lambda(1-e^{-\mu})].$$

3.
$$\begin{cases}
\mathbb{E}(S) = \mathbb{E}[\mathbb{E}(S/N)] \\
\mathbb{E}(S/N=n) = n\mu \Rightarrow \mathbb{E}(S/N) = N\mu \\
\mathbb{E}(S) = \mathbb{E}(N\mu) = \mu\mathbb{E}(N) = \mu\lambda \\
V(S) = \mathbb{E}[V(S/N)] + V[\mathbb{E}(S/N)] \\
\mathbb{E}(S/N) = N\mu \text{ and } V(S/N) = N\mu \\
\mathbb{E}[V(S/N)] = \mathbb{E}(N\mu) = \mu\mathbb{E}(N) = \mu\lambda \\
V[\mathbb{E}(S/N)] = V(N\mu) = \mu^2 V(N) = \mu^2\lambda \Rightarrow V(S) = \mu\lambda + \mu^2\lambda \\
= \mu\lambda(1+\mu).
\end{cases}$$

Exercise 4

1. The law followed by the lifetime, in years, of this component is the exponential law with mean **4**. Its density is

$$f(x) = 0,25e^{-0,25x}\text{for } x \geq 0$$

The average lifespan is equal to $1/0,25 = 4$ years, and its standard deviation is $\sigma = 4$. The distribution function is

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \int_0^x f(t)dt = 1 - e^{-0,25x} & \text{if } x \geq 0 \end{cases}$$

2.

$$P(X > 4) = 1 - F(4) = exp(-1) = 0,368.$$

3. By definition of a conditional probability

$$P(X \geq 6/X > 5) = \frac{P(X \geq 6)}{P(X > 5)} = e^{-0,25(6-5)} = e^{-0,25} = 0,78.$$

This is a phenomenon without memory

4.

$$P[\mathbb{E}(X) - \sigma < X < \mathbb{E}(X) + \sigma] = P(0 < X < 8) = F(8) = 0,865.$$

5. We are looking for the duration d during which 50% hard drives work without failure.

$$P(X > d) = 1 - F(d) = 0,5. \text{ From where } \exp(-0,25d) = 0,5.$$

We obtain $d = 2,77$years.

6. We seek the duration t such that: $P(X < t) \leq 0,15$. From where

$$P(X \geq t) = \exp(-0,25t) = 0,85 \Rightarrow t = -\frac{\ln 0,85}{0,25} \simeq 0,61.$$

Exercise 5

$X \to N(m, \sigma)$.

We seek to calculate the mean m and the standard deviation σ. We pose:

$$Z = \frac{X - m}{\sigma} \to N(0,1).$$

we have

$$\begin{cases} P(X>60)=0,0869 \\ P(X<45)=0,6406 \end{cases} \Rightarrow \begin{cases} P\left(Z>\dfrac{60-m}{\sigma}\right)=0,0869 \\ P\left(Z<\dfrac{45-m}{\sigma}\right)=0,6406 \end{cases}$$

$$\Rightarrow \begin{cases} 1-P\left(Z\leq\dfrac{60-m}{\sigma}\right)=0,0869 \\ \quad =F\left(\dfrac{60-m}{\sigma}\right)=0,9131 \end{cases}$$

$$\Rightarrow \begin{cases} F\left(\dfrac{60-m}{\sigma}\right)=0,9131 \\ F\left(\dfrac{45-m}{\sigma}\right)=0,6406 \end{cases}$$

Using the reduced centered normal distribution table, we have

$$\begin{cases} \dfrac{60-m}{\sigma}=1,36 \\ \dfrac{45-m}{\sigma}=0,36 \end{cases} \Rightarrow \begin{cases} \sigma=15. \\ m=39,6. \end{cases}$$

Part II

Inferential Statistics

If probability theory is the deductive method of statistics, then by implication, theory in statistical science must be represented by some well-defined population with a known probability distribution and data by the sample drawn from that population. Statistical inference then becomes the inductive methods for using sample data to make inferences about the probability distribution of the population from which the sample was drawn.

Population (Ω) **Known Probability Distribution** **of Population**	$\longrightarrow\longrightarrow\rightarrow$	**Sample** **Sampling Distributions** **of Sample Statistics**

Population
Probability Theory $\searrow\nearrow$ **Statistical Inference**
Sample

Statistical inference then is the inverse of the probability theory. It is the process of making statements about an unknown population on the basis of a known sample from that population.

12 Sampling Theory

Introduction

The study of the characteristics of all elements of a population is often impossible to achieve due to cost and time constraints. This impossibility leads to studying a subset from the parent population: The sample.Sampling consists of deducing from the supposedly known knowledge of the characteristics of a population, the characteristics of the samples taken from this population.

Estimation is the inverse problem. It involves estimating, from the characteristics calculated on one or more samples, the value of the characteristics of the parent population.

	Parent population Ω	**Samples** E_i
Number or size	N	n
Average	m	$\overline{X}_i$
Frequency or proportion	P	f_i
Variance	σ^2	σ_i^2
Standard deviation	σ	σ_i

Probabilistic methods of sample formation consist of randomly selecting elements from the population and are the only ones that respect statistical laws. The sampling of elements from the sample can be carried out.

With replacement:

The sampled element is immediately returned to the parent population before sampling the next one. Since an element may be sampled several times, the draws are independent and the sample is said to be **non-exhaustive**.

Without replacement:

The sample is exhaustive, but the draws are not independent since the composition of the parent population is modified at each draw.In the following, in order to apply the rules of probability calculation, the samples will be assumed to be constituted with replacement, or to be samples without replacement whose size is negligible compared to that of the population which is large or infinite (the draw is then similar to a draw with replacement). Let us indicate the procedure to be implemented to constitute a sample using a table of random numbers.

This table is made up of the numbers $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ and each of the integer values has the same probability of appearing. Any number in the table has no relationship with the number above, below, to the right or to the left of it.the numbers are randomly scattered in the table we are using, the numbers are grouped into columns of **5** digits, each line has **50** numbers (**10 groups of 5**). To choose numbers from the table, it is simply a matter of:

a) Choosing an entry point in the table.

b) Choosing a reading route. We can read the numbers in a line (from left to right or from bottom to top).

We could also skip every other number,

12.1 L'enquête statistique (Techniques de sondage)

Randomly drawing a sample of size $n = 10$ using the random number table. The delegate of a student association of a university wants to randomly draw a sample of **10 individuals** who are part of the association. Let's suppose that the **AEU** has **300 individuals** listed on a file. We therefore have the following:

Population: The individuals who are members of the association.

Sampling frame: The list of names of the individuals on the file.

Statistical unit: The individuals.

Population size: $N = 300$.

Required sample size: $N = 10$.

Drawing method: Without replacement (**exhaustive drawing**). We start by numbering each individual in the sampling frame from **001** to **300**. Since the sampling frame has **300** individuals, we will choose 3-digit numbers from the table. To read the table, we propose the following rule: Start from the **3rd** line, considering only the last **3 digits** of the **4th column** (and the following ones if applicable) with reading from top to bottom, only retain the reading results that are between **001** and **300**. Since we are drawing without replacement, we reject any number already taken out that appears again in the selection procedure. We then obtain the following **10 numbers**: The individuals with the following numbers in the sampling frame will constitute the sample of size **n=10**.

251	045	075	157	199
267	278	026	238	051

12.2 Sampling distribution

Let in a parent population Ω of size N, a random variable X for which the mathematical expectation m, the proportion P and the standard deviation are known.

From this population come k samples $E_1, E_2, ..., E_k$ of size n which will have different means and standard deviations. The notion of sampling distribution can be summarized and schematized:

Parent population: Ω	**Sample 1**
Size: N	Size: n
Average: m (connue)	Average:$\overline{X}_1$
proportion: P (connue)	proportion:f_1
Standard deviation σ (connue)	Standard deviation:σ_1

Sample 2	**Sample k**
Size: n	Size: n
Average:$\overline{X}_2$	Average:$\overline{X}_k$
proportion: f_2	proportion: f_k
Standard deviation:σ_2	Standard deviation:σ_k

Remark 71 *Deduce the characteristics of a sample of knowledge characteristics of the parent population.*

12.3 Sampling distribution of means

The averages $\overline{X}_i$ of each sample vary from sample to sample and represent the distribution of the means of the random variable $\overline{X}_i$ which associates with any sample of size n the mean of that sample.

The random variable $\overline{X}_n$ therefore takes the values: $\overline{X}_1, \overline{X}_2, ..., \overline{X}_k$.

12.4 Characteristic values of $\overline{X}_n$

· The mathematical expectation of the random variable $\overline{X}_n$ is equal to that of the parent population:

$$\mathbb{E}\left(\overline{X}_n\right) = m$$

· The variance of the random variable $\overline{X}_n$ is equal to that of the parent population reported to the sample size:

$$V\left(\overline{X}_n\right) = \frac{\sigma^2}{n}$$

· The standard deviation of the random variable $\overline{X}_n$ is deduced from the variance:

$$\sigma\left(\overline{X}_n\right) = \frac{\sigma}{\sqrt{n}}$$

Remark 72 *If the samples are taken from a finite parent population and are constituted without replacement. The mathematical expectation of* $\overline{X}_n$ *is always equal to* m*, but the standard deviation is corrected by the completeness factor (correction factor)*

$\sigma\left(\overline{X}_n\right) = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}} \approx \dfrac{\sigma}{\sqrt{n}}\sqrt{1-\dfrac{n}{N}}$ *such as* $\dfrac{n}{N}$ *represents the polling rate.*

Mathematical justification: Mathematical expectation and variance of $\overline{X}$***, the sample mean***

A sample of size n is randomly selected (drawing with replacement) whose elements have a measurable characteristic X following a probability distribution with mean $\mathbb{E}(X) = m$ and of variance $Var(X) = \sigma^2$.

By randomly taking a sample of size n from this population, we create a sequence of n independent random variables $X_1, X_2, ..., X_n$ each of which has the same distribution as X.

a) The sample mean $\overline{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ is a random variable whose mathematical expectation is $\mathbb{E}(X) = m$.

Indeed

$$\begin{aligned}
\mathbb{E}\left(\overline{X}\right) &= \mathbb{E}\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n}\mathbb{E}\left(X_1 + X_2 + ... + X_n\right) \\
&= \frac{1}{n}\left[\mathbb{E}(X_1) + \mathbb{E}(X_2) + ... + \mathbb{E}(X_n)\right] = \frac{n}{n}m.
\end{aligned}$$

b) The variance of $\overline{X}$ is equal to the variance σ^2 of the population divided by n the sample size

$$Var\left(\overline{X}\right) = \sigma^2\left(\overline{X}\right) = \mathbb{E}\left(\overline{X} - m\right)^2 = \frac{\sigma^2}{n}$$

Indeed

$$Var\left(\overline{X}\right) = Var\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n^2}\left[Var\left(X_1 + X_2 + ... + X_n\right)\right]$$

Since $X_1, X_2, ..., X_n$ are independent, we can write

$$\begin{aligned}
Var\left(\overline{X}\right) &= \frac{1}{n^2}\left[Var(X_1) + Var(X_2) + ... + Var(X_n)\right] \\
&= \frac{1}{n^2}\left(\sigma^2 + \sigma^2 + ... + \sigma^2\right) \\
&= \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

12.5 Frequency sampling distribution

The probability of occurrence of an event A is equal to P. We consider random samples of size n drawn from a population of size N. For each sample we determine the proportion f of occurrence of event A. The population and the samples follow binomial distributions $B(N, P)$ and $B(n, f)$ respectively. The average m_f and the standard deviation σ_f of the sampling distribution of frequencies are:

$$m_f = P \text{ and } \sigma_f = \sqrt{\frac{P(1-P)}{n}}$$

if the draw is non-exhaustive or if the population is infinite.

$$\sigma_f = \sqrt{\frac{P(1-P)}{n}}\sqrt{\frac{N-n}{N-1}}$$

if the draw is exhaustive.

Remark 73 ** If the sample size n is large enough (in practice $n \geq 30$) the sampling distribution of the mean approaches the normal distribution regardless of the population distribution, that is:*

$$\overline{X} \to N\left(m, \frac{\sigma}{\sqrt{n}}\right).$$

** Si la population est normalement distribuée, la distribution d'échantillonnage de la moyenne est une loi normale quelle que soit la valeur n de la taille des échantillons.*
** Si la population parente possède une distribution pratiquement symétrique, il semble qu'un échantillon d'au moins **15** observations soit convenable pour que la distribution de la moyenne soit approximativement normale.*

13 Parameter Estimation

Introduction

Random variables $X_1, X_2, ..., X_n$ are defined by their distribution function:

$$F(x_1, x_2, ..., x_n; \theta) = P(X < x_1, X_2 < x_2, ..., X_n < x_n; \theta)$$

dependent on a parameter θ belonging to a set Θ included in $\mathbb{R}$. The problem of estimation is that of measuring θ from the observation of random variables $(X_1, X_2, ..., X_n)$.

An estimator of θ is a random variable, denoted $\widehat{\theta} = \widehat{\theta}(X_1, X_2, ..., X_n)$, sufficiently regular function of random variables $X_1, X_2, ..., X_n$ and only of these random variables (that's to say that θ must not intervene in $\widehat{\theta}$ explicitly).

Remark 74 *We ask to $\widehat{\theta}$ the following two qualities:*

$$\begin{cases} 1) \text{ absence of bias: } \mathbb{E}\left(\widehat{\theta}\right) = \theta \ \left(biais = b(\theta) = \mathbb{E}\left(\widehat{\theta}\right) - \theta\right) \\ 2) \text{ Convergence: } \forall\varepsilon > 0; \lim\limits_{n\to+\infty} P\left[\left|\widehat{\theta} - \theta\right| > \varepsilon\right] = 0 \end{cases}$$

The first property is related to the absence of systematic error while the second is convergence in probability.

Definition 75 *An estimator $\widehat{\theta}$ asymptotically unbiased*

$\lim\limits_{n\to\infty} \mathbb{E}\left(\widehat{\theta}\right) = \theta$ *and whose variance verifies* $\lim\limits_{n\to\infty} Var\left(\widehat{\theta}\right) = 0$ *is convergent.*

Theorem 76 *Any unbiased estimator whose variance tends to zero is convergent.*

Proof. This result is directly deduced from the Bienaymé-Tchebychev inequality:

$$P_\theta\left[\left|\widehat{\theta}_n - \theta\right| > \varepsilon\right] \leq \frac{Var_\theta\left(\widehat{\theta}\right)}{\varepsilon^2} \to 0 \text{ When } n \to \infty$$

■

Example 77 *Estimated mean*
When $X_1, X_2, ..., X_n$ are independent random variables (i.r.v), of the same average m and of the same variance σ^2, the estimator of the mean, $\widehat{m} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is such that,

$$\begin{cases} \mathbb{E}\left(\overline{X}\right) = m \\ Var\left(\overline{X}\right) = \frac{\sigma^2}{n} \to 0 \ \textit{When } n \to \infty \end{cases}$$

So $\widehat{m} = \overline{X}$ is an unbiased and consistent estimator of m.

Example 78 $S^2 = \sigma^2_{\acute{e}chantillon} = \sigma^2_{\acute{e}cha}$ *is an unbiased estimator of σ^2. Because*

$$S^2 = \frac{\sum_i \left(X_i - \overline{X}\right)^2}{n-1}; \ \mathbb{E}\left(S^2\right) = \sigma^2$$

We can write

$$(n-1)\, S^2 = \sum_i \left(X_i - \overline{X}\right)^2 = \sum_i \left[(X_i - m) - \left(\overline{X} - m\right)\right]^2$$

which gives, in development the right member

$$(n-1)\, S^2 = \sum_i (X_i - m)^2 - n\left(\overline{X} - m\right)^2$$

13.1 Optimal (efficient) estimator

13.1.1 Quality of an estimator

The quality of an estimator will be measured using a distance to the parameter which can be for example $\left|\widehat{\theta}_n - \theta\right|$ or $\left(\widehat{\theta_n} - \theta\right)^2$. To obtain a numerical indicator we can then determine the average value of this distance.

The indicator generally chosen, because it lends itself easily to calculations, is the mean square error defined for all θ by

$$\mathbb{E}_Q\left(\widehat{\theta}_n\right) = \mathbb{E}_\theta\left(\widehat{\theta_n} - \theta\right)^2 = Var\left(\widehat{\theta}_n\right) + b_n^2(\theta)$$

In the particular case of an unbiased estimator, this squared error coincides with the variance of the estimator.

13.1.2 Efficient estimator

An unbiased estimator is more efficient (or simply efficient) if its variance is the lowest among the variances of the other unbiased estimators. Thus if $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are two unbiased estimators of the parameter θ, the estimator $\widehat{\theta}_1$ is more effective if

$$Var\left(\widehat{\theta}_1\right) < Var\left(\widehat{\theta}_2\right) \text{ and } \mathbb{E}\left(\widehat{\theta}_1\right) = \mathbb{E}\left(\widehat{\theta}_2\right) = \theta$$

The notion of an efficient estimator can be illustrated in the following way.

The distribution of $\widehat{\theta}_1$ is more concentrated around θ than that of $\widehat{\theta}_2$. If we sample a normal population $\overline{X}$ and M_e are unbiased estimators of m because

$$\mathbb{E}\left(\overline{X}\right) = \mathbb{E}\left(M_e\right) = m.$$

On the other hand, the variance of $\overline{X}$ is smaller than that of the median since

$$Var\left(\overline{X}\right) = \frac{\sigma^2}{n} \text{ and } Var\left(M_e\right) = 1,57\frac{\sigma^2}{n}.$$

For the same sample size, $\overline{X}$ is more effective than M_e to estimate m; $Var\left(\overline{X}\right) < Var\left(M_e\right)$.

Comparison of two estimators of m.

Example 79 *Let X_1, X_2, X_3 a random sample taken from an infinite population with $\mathbb{E}\left(X_i\right) = m$ and $Var\left(X_i\right) = \sigma^2$. Show that*

$$\widehat{m} = \frac{X_1 + 2X_2 + 3X_3}{6}$$

is an unbiased estimator of m but is less efficient than

$$\overline{X} = \frac{\sum_i X_i}{3}$$

That is to say that

$$\mathbb{E}\left(\widehat{m}\right) = m \text{ and } Var\left(\widehat{m}\right) > Var\left(\widehat{X}\right)$$

Let's check that $\widehat{m}$ is an unbiased estimator of m

$$\begin{aligned} \mathbb{E}\left(\widehat{m}\right) &= \mathbb{E}\left(\frac{X_1 + 2X_2 + 3X_3}{6}\right) = \frac{1}{6}\left[\mathbb{E}\left(X_1\right) + 2\mathbb{E}\left(X_2\right) + 3\mathbb{E}\left(X_3\right)\right] \\ &= \frac{6m}{6} = m \end{aligned}$$

Now let's determine $Var\left(\widehat{m}\right)$

$$Var\left(\widehat{m}\right) = \frac{14}{36}\sigma^2$$

Or

$$Var\left(\overline{X}\right) = \frac{\sigma^2}{n} = \frac{\sigma^2}{3}, \text{ from where } 0,333\sigma^2 < \frac{14}{36}\sigma^2.$$

Therefore $\overline{X}$ is more effective than $\widehat{m}$ to estimate m.

13.2 Fréchet-Darmois-Cramer-Rao (F.D.C.R) inequality

We will see in some conditions there is a lower bound for all variances of unbiased estimators, which will constitute a limit that does not allow to constantly improve the estimators. On the other hand, if this bound is reached by an estimator, it will become the best and will be qualified as optimal in the class of unbiased estimators.

Definition 80 *We call the likelihood of the sample* $(X_1, X_2, ..., X_n)$ *the probability law of this* $n-uple$, *notée* $L(X_1, X_2, ..., X_n; \theta)$ *and, defined by*

$$L(x_1, x_2, ..., x_n; \theta) = \Pi_{i=1}^{n} P(X_i = x_i; \theta)$$

if X *is a discrete random variable and by*

$$L(x_1, x_2, ..., x_n; \theta) = \Pi_{i=1}^{n} f(x_i; \theta)$$

if is a continuous random variable of density $f(x, \theta)$.

Le théorème suivant préciser la borne inférieure pour la variance des estimateurs sans biais.

Theorem 81 *Under the Cramer-Rao assumptions, especially if* $E = X(\Omega)$ *is independent of the parameter to be estimated* θ, *for any unbiased estimator* $\widehat{\theta}$ *of* θ *we have:*

$$Var_\theta\left(\widehat{\theta}\right) \geq \frac{1}{I_n(\theta)} = B_F(\theta)$$

such as $I_n(\theta)$ *is the amount of Fisher information which is defined by*

$$I_n(\theta) = \mathbb{E}\left(-\frac{\partial^2 \ln l}{\partial \theta^2}\right)$$

and $B_F(\theta)$ *is the lower bound of* ***F.D.C.R***.

Remark 82 *If* $E = X(\Omega)$ *depends on the parameter to be estimated* θ, *we obtain* $I_n(\theta) = \mathbb{E}\left(\frac{\partial \ln l}{\partial \theta}\right)^2$.

Example 83 *Let* X *a random variable with exponential distribution of parameter* $\frac{1}{\theta}$, *or Gamma law noted* $\Gamma\left(1, \frac{1}{\theta}\right)$, *with* $\theta > 0$ *density for* $x > 0$:

$$f(x, \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right)$$

The likelihood here admits the following expression

$$L(x_1, x_2, ..., x_n; \theta) = \Pi_{i=1}^{n} f(x_i; \theta) = \frac{1}{\theta^n} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right)$$

To calculate the Fisher amount of information we write

$$\ln L\left(x_1, x_2, ..., x_n; \theta\right) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^{n} x_i.$$

$$\frac{\partial \ln l}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} x_i$$

as $X\left(\Omega\right) = \mathbb{R}_+$ *is independent of* θ *we have*

$$\frac{\partial^2 \ln l}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3} S_n$$

such as

$$S_n = \sum_{i=1}^{n} x_i$$

From where

$$I_n\left(\theta\right) = \mathbb{E}\left(-\frac{\partial^2 \ln l}{\partial \theta^2}\right) = \frac{n}{\theta^2} + \frac{2n\theta}{\theta^3} = \frac{n}{\theta^2}.$$

Example 84 *If we now take the example of the exponential law on* $[\theta, +\infty[$, *of density*

$$f\left(x, \theta\right) = \begin{cases} \exp\left(-\left(x - \theta\right)\right); & \textit{if } x \geq \theta \\ 0; & \textit{othewise.} \end{cases}$$

Likelihood is written

$$L\left(x_1, x_2, ..., x_n; \theta\right) = \exp\left(-\sum_{i=1}^{n} \left(x_i - \theta\right)\right)$$

if all the x_i *are bigger than* θ, *that is to say if*

$$\min\left\{x_i / 1 \leq i \leq n\right\} \geq \theta$$

we then have

$$\ln L\left(x_1, x_2, ..., x_n; \theta\right) = -\sum_{i=1}^{n} x_i + n\theta.$$

from where

$$\frac{\partial \ln l}{\partial \theta} = n \textit{ and } I_n\left(\theta\right) = \mathbb{E}\left(\frac{\partial \ln l}{\partial \theta}\right)^2 = n^2.$$

13.3 Efficient estimator

Definition 85 *An unbiased estimator* $\widehat{\theta}$ *is said to be efficient if its variance is equal to the lower bound of* $(F.D.C.R)$:

$$Var\left(\widehat{\theta}\right) = \frac{1}{I_n\left(\theta\right)}$$

Example 86 *If we take the example of the expenontial law with parameter* $\frac{1}{\theta}$, *as* $\mathbb{E}_\theta(X) = \theta$, *we know that* $\widehat{\theta}_n = \overline{X}$ *is an unbiased estimator and is consistent. Moreover*

$$Var_\theta\left(\widehat{\theta}_n\right) = Var_\theta\left(\overline{X_n}\right) = \frac{Var_\theta(X)}{n} = \frac{\theta^2}{n} = \frac{1}{I_n(\theta)}$$

so this estimator is also efficient.

13.4 Methods for constructing an estimator

13.5 Maximum likelihood method

Likelihood $L(x_1, x_2, ..., x_n; \theta)$ represents the probability of observing the $n-uple$ $(x_2, ..., x_n)$ for a fixed value of θ. In the opposite situation, where $(x_2, ..., x_n)$ without knowing the value of θ, we will assign to θ the value which appears most likely, taking into account the observation available, that is to say the one which will give it the highest probability. We therefore set ourselves the following rule: $(x_2, ..., x_n)$ fixed we consider the likelihood L as a function of θ and it is attributed to θ the value that maximizes this function. Hence the following definition:

Definition 87 *We call the maximum likelihood estimator* (*e.m.v*) *any function* $\widehat{\theta}$ *of* $(x_2, ..., x_n)$ *who checks:*

$$L\left(x_1, x_2, ..., x_n; \widehat{\theta}\right) = \max_{\theta\in\Theta} L(x_1, x_2, ..., x_n; \theta)$$

This definition does not provide any information on the existence or uniqueness of such an estimator. The search for e.m.v can be done directly by searching for the maximum of L, *or in the special case where the function* L *is twice differentiable with respect to* θ, *as a solution to the equation* $\frac{\partial \ln l}{\partial\theta} = 0$ *who also checks* $\frac{\partial^2 \ln l}{\partial\theta^2} < 0$.

Example 88 *Let us find the* ***emv*** *for the exponential family of laws with parameter* $\frac{1}{\theta}$. *The log-likelihood is infinitely differentiable for* $\theta > 0$ *and we had obtained in example 1*

$$\frac{\partial \ln l}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \theta = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}_n$$

With

$$\frac{\partial^2 \ln l}{\partial \theta^2} = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_{i=1}^{n} x_i = \frac{n}{\theta^3}\left(\theta - 2\overline{X}_n\right)$$

either for

$$\theta = \overline{X}_n$$

$$\left(\frac{\partial^2 \ln l}{\partial \theta^2}\right)_{\theta=\overline{X}_n} = -\frac{n}{\overline{X}_n^2} < 0,$$

so the emv is

$$\theta = \overline{X}_n.$$

13.6 Estimation of a mean by confidence interval

We propose to estimate, by confidence interval, the mean m of a measurable characteristic of a population. This involves calculating, from the mean X of the sample, an interval in which it is likely that the true value of m is found. We obtain this interval by calculating two limits to which is associated a certain assurance of containing the true value of m. This interval is defined according to the following equation

$$P\left(\overline{X} - k \leq m \leq \overline{X} + k\right) = 1 - \alpha$$

and the limits will take, after having taken the sample and calculated the estimate $\overline{X}$, the following form

$$\overline{X} - k \leq m \leq \overline{X} + k.$$

where k will be determined using the standard deviation of the sampling distribution of $\overline{X}$ and the confidence level $1 - \alpha$ chosen a priori. We know that if we take a random sample of size n from a normal population of known variance,

$$\overline{X} \rightarrow N\left(m, \frac{\sigma^2}{n}\right)$$

If the distribution of the measurable trait (the population) is unknown or the population variance is unknown, a sample of size $n \geq 30$ allows us, according to the central limit theorem, to consider that $\overline{X}$ follows approximately a normal distribution. Consequently, the quantity

$$Z = \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \text{ or } \left(\frac{\overline{X} - m}{\frac{S}{\sqrt{n}}} \text{ as the case may be}\right)$$

follows a reduced centered normal distribution.

Let us start from this fact to deduce a random interval having, a priori, a probability $1 - \alpha$ to contain the true value of m, which amounts to determining k such that

$$P\left(\overline{X} - k \leq m \leq \overline{X} + k\right) = 1 - \alpha$$

From where

$$P\left(-Z_{\frac{\alpha}{2}} \leq \frac{\overline{X} - m}{\frac{\sigma}{\sqrt{n}}} \leq Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

So

$$P\left(\overline{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq m \leq \overline{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

which is of the form

$$P\left(\overline{X}-k \leq m \leq \overline{X}+k\right)=1-\alpha$$

from where

$$k=Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Example 89 *An independent laboratory has verified on behalf of the consumer protection office, the bursting strength (in kg/cm^2) of a gasoline tank from a certain manufacturer. Similar tests carried out a year ago allow us to consider that the bursting strength is normally distributed with a variance of 9. Tests on a sample of 10 tanks lead to an average bursting strength of $219kglcm^2$. Estimate by confidence interval the average bursting strength of this type of tank with a confidence level of 95%.*

Example 90 *The elements necessary for calculating the confidence interval are indicated as follows:*

$$\begin{aligned} \overline{X} &= 219 \\ \sigma &= \sqrt{9}=3kg\,\mathrm{lcm}^2 \end{aligned}$$

Sample size $n=10$*, the level of confidence:* $1-\alpha=0,95$ *So* $\alpha=0,05$.

14 Hypothesis testing

Introduction

Let a hypothesis H_0 concerning a population on the basis of the results of samples taken from this population we are led to accept or reject the hypothesis H_0. The decision rules are called statistical tests.

H_0 denotes the hypothesis called null hypothesis and by H_1 we denote the hypothesis called alternative hypothesis we have:

$$\left\{ \begin{array}{ll} H_0 \text{ true} & H_1 \text{ false or} \\ H_0 \text{ false} & H_1\text{true} \end{array} \right.$$

There are four solutions of which only the first two are correct:

$a)H_0$ is true and we chose H_0

$b)H_0$ is false and has been rejected H_0

$c)H_0$is true and has been rejected H_0

$d)H_1$ is true and we chose H_0

There are two types of errors:

a) If H_0 is true and we rejected it, we say that we have an error of 1^{st} species. The probability of the error of $1^{t\grave{e}re}$ species is noted α.

b) If H_1 is true and we accepted H_0, we say that we have an error of 2^{end} species. The probability of the error of 2^{end} species is noted β. α is the significance threshold of the test and $1-\alpha$ his confidence threshold.

15 Test categories

1. A test is said to be a simple hypothesis test if we want to choose between two values of a parameter $\theta\,(\theta_0$ and $\theta_1)$ we have:

$$\left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{array} \right.$$

2. A test is said to be bilateral if

$$a) \left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_1 \end{array} \right.$$

a) right-tailed test or superiority test

$$b) \left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_1 \end{array} \right.$$

b) left-tailed test or inferiority test

$$c) \left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_1 \end{array} \right.$$

or θ_0 and θ_1 are two values of the same parameter in two different populations.

3. A test is said to be of adjustment if:

$$\begin{cases} H_0 : F(x) = F_0(x) \\ H_1 : F(x) \neq F_0(x) \end{cases}$$

Or $F(x)$ is the distribution function of the sampled variable and $F_0(x)$ is the distribution function of a known random variable.

5. A test is said to be independent if:

$$\begin{cases} H_0 : X \text{ and } Y \text{ are two independent random variables} \\ H_1 : X \text{ and } Y \text{ are not independent variables.} \end{cases}$$

15.1 Critical and acceptance region of a hypothesis

Constructing a test involves determining the critical region ω_0 of $\mathbb{R}^n$. The value of α, error of 1^{st} species being fixed (in general $\alpha = 0.05, 0.01$ or 0.1) the set of values of the decision variable which allow us to exclude H_0 and to choose H_1 is called critical region, the complementary $\overline{\omega}_0$ of this critical region is called the acceptance region. We have

$$\begin{aligned} P(\omega_0 \,|H_0\,) &= \alpha; P(\overline{\omega}_0 \,|H_0\,) = 1 - \alpha \\ P(\omega_0 \,|H_1\,) &= 1 - \beta; P(\overline{\omega}_0 \,|H_1\,) = \beta \end{aligned}$$

we extract a random sample from the population and accept H_0 if the value of the decision variable belongs to the acceptance region. Otherwise we reject it and accept H_1. For a fixed value α, we maximize the quantity $1 - \beta$ called the power of the test.

15.2 Test between two simple hypotheses (Neymane and Pearson method)

We test

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

we set the risk of 1^{st} species α

$$L(x; \theta) = L(x_1, ..., x_n; \theta)$$

is the likelihood function with $x = (x_1, ..., x_n)$, ω_0, critical region, is defined by

$$P(\omega_0 \,|H_0\,) = \alpha = \int_{\omega_0} L(x; \theta_0)\, dx.$$

we maximize the quantity

$$1 - \beta = \int_{\omega_0} L(x; \theta_1)\, dx = P(\omega_0 \,|H_1\,) = 1 - \beta = \int_{\omega_0} \frac{L(x; \theta_1)}{L(x; \theta_0)} L(x; \theta_0)\, dx$$

to maximize $1-\beta$, we are looking for the set of points of $\mathbb{R}^n$ such as

$$A = \frac{L(x;\theta_1)}{L(x;\theta_0)} \geq K_\alpha$$

the constant K_α is determined by

$$\int_{A \geq K_\alpha} L(x;\theta_0)\, dx = \alpha.$$

15.3 Homogeneity test

From a sample of size n_1 taken from a population P_1 and a sample of size n_2 taken from a population P_2, the test allows us to decide between:

$$\begin{cases} H_0 : \theta_0 = \theta_1 \\ H_1 : \theta_0 \neq \theta_1 \end{cases}$$

θ_0 and θ_1 are the two values of the same parameter of the two populations P_1 and P_2.

15.4 Test of homogeneity of two means

In case sample sizes are high $(n_1, n_2 \geq 30)$, the variables $\overline{X}_1$ et $\overline{X}_2$ (corresponding to populations P_1 and P_2) following the respective normal laws:

$$N\left(m_1, \frac{\sigma_1}{\sqrt{n_1}}\right), N\left(m_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$$

or m_i and σ_i are the mean and standard deviation of the population $P_i\,(i = 1, 2)$.

The random variable $\left(\overline{X}_1 - \overline{X}_2\right)$ also follows a normal law

$$N\left(m_{1-}m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

we choose between the two hypotheses

$$\begin{cases} H_0 : m_1 = m_2 \\ H_1 : m_1 \neq m_2 \end{cases}$$

or

$$\begin{cases} H_0 : m_1 - m_2 = 0 \\ H_1 : m_1 - m_2 \neq 0 \end{cases}$$

if H_0 is true, $m_1 - m_2 = 0$ and

$$\overline{X}_1 - \overline{X}_2 \to N\left(0, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

We then have

$$P(|Z| \leq U_\alpha) = 1 - \alpha$$

or

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (m_1 - m_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - 0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

follows a normal law $N(0,1)$. We accept H_0 if the value $z = \dfrac{\left(\overline{X}_1 - \overline{X}_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ of Z such as

$$-U_\alpha \leq z \leq U_\alpha$$

or the value of U_α is obtained by reading the normal law table $N(0,1)$. We reject H_0 if $|Z| > U_\alpha$. We will say that the difference is significant between $\overline{X}_1$ and $\overline{X}_2$.

Remark 91 *i) If σ_1^2 and σ_2^2 are unknown, we replace them with the estimators*

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(X_i - \overline{X}_1\right)^2$$

and

$$S_2^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_2} \left(X_i - \overline{X}_2\right)^2 \text{ respectively.}$$

as the samples are of large sizes we consider that

$$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

follows a normal law $N(0,1)$ (if H_0 is true that is to say $m_1 - m_2 = 0$)
ii) If the samples are of respective sizes $n_1 < 30$ and $n_2 < 30$,the test is no longer valid because the central limit theorem no longer applies. but for two populations $P_1$and P_2 following normal distributions $N(m_1, \sigma_1)$ and $N(m_2, \sigma_2)$ respectively having standard deviations σ_1 and $\sigma_2$equal and unknown, that is to say $\sigma_1 = \sigma_2 = \sigma$ we have:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

*follows a **Student** distribution at $n_1 + n_2 - 2$ degrees off reedom,*

S^2being the pointestimate of σ^2.

we then accept H_0 if $|t| < U_\alpha$ or U_α is a value obtained by reading the Student-Fisher t-distribution table (number of degrees of freedom $n_1 + n_2 - 2$; significance threshold α)

15.5 Test of homogeneity of two proportions

Each individual of the two populations p_1 and p_2 may or may not possess a certain characteristic. We say that this characteristic is present in proportion P_1 and P_2 in the populations p_1 and p_2 respectively: we test at the significance threshold α:

$$\begin{cases} H_0 : P_1 = P_2 \\ H_1 : P_1 \neq P_2 \end{cases}$$

or

$$\begin{cases} H_0 : P_1 - P_2 = 0 \\ H_1 : P_1 - P_2 \neq 0 \end{cases}$$

Of the population P_i, a sample of size is extracted n_i. It corresponds to a proportion f_i $(i = 1, 2)$.

If the samples are of large sizes ($n_1 \geq 30$ and $n_2 \geq 30$), the central limit theorem allows us to assert that f_i follows a normal law

$$N\left(P_i, \sqrt{\frac{P_i Q_i}{n_i}}\right)$$

with $P_i + Q_i = 1$. The random variable $f_{1-} f_2$ then follows a normal law $N\left(P_1 - P_2, \sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}\right)$.

If H_0 is true, $P_1 - P_2 = 0$ and $f_1 - f_2 \to N\left(0, \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$ because $P_1 = P_2 = P$. But since P is unknown, we make an approximation. We estimate P by

$$f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$$

and so

$$f_1 - f_2 \to N\left(0, \sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

and

$$Z = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \to N(0, 1).$$

Let U_α the constant such that $P(|Z| \geq U_\alpha) = \alpha$. We then have

$$P(|Z| < U_\alpha) = 1 - \alpha,$$

we accept H_0 if the value

$$z = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

of Z is such that

$$-U_\alpha \leq z \leq U_\alpha$$

we reject H_0 if $|Z| \geq U_\alpha$.

Remark 92 U_α *is obtained by reading the normal law table* $N(0,1)$.

16 Some selected themes of probability

16.1 Poisson process

One of the most important counting processes is the Poisson process. Intuitively, it involves counting the number of occurrences of events that occur randomly and independently of each other over time.

Definition 93 *The counting process* $(N_t)_{t\geq 0}$ *is a Poisson process with rate* $\lambda > 0$*, if*

$i)$ The process is independently increases;

$ii)$ The number of occurrences in any time interval of length t follows the Poisson distribution with parameter λ_t.

$$\forall s, t \geq 0,\ P\left(N_{t+s} - N_s = n\right) = e^{-\lambda t}\frac{(\lambda t)^n}{n!},\ n = 0, 1, ...$$

It follows immediately from such a definition that a Poisson process is a process with stationary increments and moreover.

$$\mathbb{E}\left(N_t\right) = \lambda t.$$

16.2 Markov Chaîns

16.3 Discrete-time processes

Definition 94 *A discrete-time random process is a family* $\{X_n; n \in \mathbb{N}\}$ *of random variables indexed by positive integers. A discrete-time process is characterized by the data of the law of each of the vectors* $(X_0, X_1, ..., X_n)$ *for everything* $n \in IN$*. We then speak of process law to describe the set of finite-dimensional laws. Many results have been obtained concerning processes consisting of independent variables and the same law. The simplest notion of dependence that can be introduced into a sequence of random variables consists of assuming that the value observed for the variable* X_n *depends only on the observed value for the variable* X_{n-1}*. In this case, we only consider processes with values in a finite or countable set* E *(discrete-valued processes).*

Definition 95 *A random process* $\{X_n; n \in IN\}$ *with values in* E *is a homogeneous Markov chain denoted* (X_n) *if, for all* $n \geq 0$*,*

$$\forall i_0, i_1, ..., i, j \in E,\ P(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, ..., X_0 = i_0) = p_{ij}.$$

For a Markov chain, the conditional distribution of the variable X_n given the past of the process at time n depends on the past only through the last observation of the process. This property is often called the Markov property. The value p_{ij} represents the probability that the process makes a transition to state j when it is in state i. Since these transition probabilities are positive numbers and the process must necessarily make a transition to a state in E, we have

$$\forall i, j \in E, p_{ij} \geq 0 \text{ and} \sum_{j \in E} p_{ij} = 1, \forall i \in E.$$

We note P the probability matrix p_{ij} state transition i to the state j

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & \cdots & \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{pmatrix}$$

The matrix P is called the **transition matrix** of the chain.

Example 96 ***Predicting the weather***. *A simplistic model of weather evolution is as follows. Suppose that the probability of rain tomorrow is a function of the weather conditions of the previous days through today's weather only. If it rains today, it will rain tomorrow with probability α. If it is sunny today, it will rain tomorrow with probability β. The evolution of weather is described using a two-state process:* 0 *for rain and* 1 *for sunny.*

Solution. This two-state process is a transition Markov chain

$$P = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix}.$$

17 Surprise, insertitude, entropy

Introduction (Information theory?)

►In everyday language, "information" is used in various contexts: news, intelligence, etc.

►In the field of telecommunications, the notion of information is linked to the efficiency of communication systems

►the information contained in a message is a measurable quantity (Shannon Engineer at Bell Labs).

Information theory touches on several areas

►coding,

►data compression,

►cryptography.

17.1 Amount of information and entropy of a source

17.2 Informative content of a message

For efficient encoding of a message into a sequence of 0 and 1, it is necessary to determine its informative content.

Example 97 ►*The source always transmits the same message, consisting of the letter A: The informative content of this message is zero, because the receiver learns nothing by receiving it.*

►*The source emits either yes or no: the receiver receives binary information.*

►*The source broadcasts tomorrow's weather: the information content is very rich, and the information transmitted is m-ary if there are m weather possibilities.*

►*A message is "rich" in information if its knowledge leads to a more "predictable" system (Case of a message that can take many different values.).*

17.3 Informational content of a message and coding

In telecommunications, we always want to save the number of bits transmitted for:

►Save time (eg: download a web page quickly from the Internet).

►Send as many messages as possible on the same medium (ex: several users on the same optical fiber).

►Direct influence on the cost of transmissions. . .

So, we would like to encode the relevant information of the message and only that!

►But how do we measure the information content of a message? Information Theory provides a measure of the amount of information.

18 Uncertainty, Information and Entropy

Definition 98 ***Discrete source****: a system that regularly emits symbols from a finite alphabet.*
Alphabet: *a finite set of symbols from the source*

$$A = \{S_0, S_1, ..., S_{k-1}\}$$

Random source*: symbols are randomly emitted according to the probabilities:*

$$P(S = s_k) = p_k;\ k = 0, 1, ..., k-1$$

with $\sum_{k=0}^{k-1} p_k = 1$.
Memoryless source*: random source whose emitted symbols are statistically independent.*

Definition 99 *the amount of information gained from observing the event* $S = s_k$ *of probability* p_k*, is defined by*

$$I(s_k) = -\log(p_k)$$

Remark 100 *if* $p_k = \frac{1}{2}$*, so* $I(s_k) = 1bit$.

Properties
- ► $I(s_k) = 0$ if $p_k = 1$;
- ► $I(s_k) \geq 0$ for $0 \leq p_k \leq 1$;
- ► $I(s_k) > I(s_i)$ for $p_k < p$;
- ► $I(s_k s_i) = I(s_k) + I(s_i)$; if s_k and s_i are statistically independent.

19 Entropy of a source

Consider a source that can send N different messages. Let p_i be the probability of sending the message m_i.

$$\textbf{Source of information} \begin{cases} \blacktriangleright \text{ Symbol } s_1\text{,probability } p_1 \\ \blacktriangleright \text{ Symbol } s_2\text{,probability } p_2 \\ \blacktriangleright \text{ Symbol } s_3\text{,probability } p_3 \\ \quad\quad\quad\quad . \\ \quad\quad\quad\quad . \\ \quad\quad\quad\quad . \\ \blacktriangleright \text{ Symbol } s_N\text{,probability } p_N \end{cases}$$

The entropy of the source S is called the mathematical expectation of $I(s_k)$ taken as a random variable.

$$H(s_k) = \sum_{k=1}^{N} -p_k \log_2(p_k) = \sum_{k=1}^{N} p_k \log_2\left(\frac{1}{p_k}\right).$$

Remark 101 *Entropy provides a measure of the average amount of information per symbol from the source, expressed in* ***bits/symbol****.*

Example 102 ►*Consider a source emitting successive symbols equal to* 0 *or* 1*. The probability of* 1 *is* $0,3$*. That of* 0 *is* $0,7$*. Calculate its entropy*

$$H(S) = -0,7\log_2(0,7) - 0,3\log_2(0,3) = 0,88sh.$$

►*The source in question reports the result of a rigged die roll:*

$$P(1) = P(6) = 0,2;\ P(2) = P(3) = P(4) = P(5) = 0,15$$

Calculate its entropy

$$H(S) = 2\times[-0,2\log(0,2) + 4[-0,15\log(0,15)]] = 2,571sh.$$

►*Calculate the entropy of the source if the die is not rigged*

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$$

$$H(S) = 6\times\left[-\frac{1}{6}\log\left(\frac{1}{6}\right)\right] = 2585sh.$$

Entropy is greater when messages are equally likely.

Conclusion 103 ►*Information theory provides a mathematical model for quantifying the information emitted by the source of a communication.*
►*Finding 1: The more different values an experiment's results can take, the greater the amount of information measured.*
►*Intuitively, this result makes sense. A source that can transmit many different messages provides more information than a source that transmits a single value.*
►*Finding 2: When a source can produce many different values, the uncertainty about the outcome of the experiment is high. Now, the amount of information transmitted is all the greater as the number of possible outcomes is different.*

Remark 104 *The amount of information received is all the more important as the uncertainty is great!*

19.1 Coding theory and entropy

Communication System: **Source Coding** and **Channel Coding**

►Now that we know how to measure the information contained in a message, we can encode it.

►Source encoding represents the message in the most economical form possible in terms of number of bits.

►Channel coding adds information that allows the receiver to reconstruct the message despite any errors that may appear due to noise on the channel.

19.1.1 Source coding

▶The goal of source coding is to find a binary translation of the messages emitted by the source that saves bits and takes into account their information content.

Example 105 *If the source emits the letter A with probability* $0,8$*, and and letters B, C, D and E with probability* $0,05$*, we intuitively feel that it is better to encode A with fewer bits, because this letter occurs often, while B, C, D and E can be encoded in a larger number of bits.*

20 Simulation

In the preceding, in particular with the technique of change of variables, we were interested in this problem which amounts to determining the law of the image by a known function of a random vector of given law. But when the system is complex such as a meteorological model, simulation becomes the only alternative to obtain information on the output variables. Finally, we will see in the next section that the strong law of large numbers ensures that the average of n independent and identically distributed integrable random variables converges when n tends to infinity towards the common expectation of these variables. To calculate this expectation numerically, we can simulate on a computer a realization of these n variables with n large and calculate the corresponding average. This is the principle of the Monte Carlo method which is very widely used in physics, reliability but also in financial mathematics.

To perform probabilistic simulations on a computer, a pseudo-random number generator is used. Such a generator returns a sequence $(x_n)_n$ of real numbers between 0 and 1. These real numbers are calculated by a deterministic algorithm but imitate a realization of a sequence of independent and identically distributed random variables according to the uniform law on $[0,1]$. The correct behavior of the sequence is verified using statistical tests.

A commonly used method to construct the sequence $(x_n)_n$ is the congruence: $x_n = \dfrac{y_n}{N}$ where the y_n are integers between 0 and $N-1$ calculated using the recurrence relation

$$y_{n+1} = (ay_n + b) \bmod (N)$$

The choice of the integers a, b, N is made so that the period of the generator (always smaller than N) is as large as possible and that the properties of the sequence $(x_n)_n$ are close to those of a realization of a sequence of $I.I.D.$ variables following the uniform law on $[0,1]$.

20.1 Simulation of discrete random variables

20.1.1 Bernoulli's law of parameter P$\in [0,1]$

If $U \to \mathcal{U}([0,1])$ then

$$X = 1_{\{U \leq P\}} \to B(P).$$

In fact X takes the values 0 or 1 and

$$P(X=1) = P(U \leq P) = \int_0^1 1_{\{u \leq P\}} du = P.$$

20.1.2 Binomial distribution with parameters n∈N* and P∈[0,1]

If $U_1, U_2, ..., U_n$ are n independent uniform variables on $[0,1]$ then

$$X = 1_{\{U_1 \leq P\}} + ... + 1_{\{U_n \leq P\}} \sum_{i=1}^{n} 1_{\{U_i \leq P\}} \rightarrow B(n, P)$$

From the above, the variables $1_{\{U_i \leq P\}}$,$1 \leq i \leq n$ are independent Bernoulli variables with parameter P. The random variable X, sum of these n variables therefore follows the binomial distribution with parameters n and P

20.1.3 Geometric law with parameter P∈]0, 1]

This is the law of the time of first success in a sequence of independent random experiments with probability of success P. Thus, if the $(U_i)_{i \geq 1}$ are independent uniform variables on $[0,1]$

$$N = \inf \{i \geq 1; U_i \leq P\} \rightarrow Geo(P).$$

20.1.4 Simulation following any discrete law

It is however still possible to obtain a variable which takes the values $(x_i)_{i \in N^*}$ with respective probabilities $(P_i)_{i \in N^*}$ (with the $P_i \geq 0$ such tha $\sum_{i \in N^*} P_i = 1$) using a single uniform variable Uon $[0,1]$ by setting

$$X = x_1 1_{\{U \leq P_1\}} + x_2 1_{\{P_1 \leq U \leq P_1 + P_2\}} + ... + x_i 1_{\{P_1 + ... + P_{i-1} \leq U \leq P_1 + ... + P_i + ...\}}.$$

Remark 106 *To implement this very general method, it is necessary to program a loop on i with the stopping test $P_1 + ... + P_i \geq U$. This can be costly in terms of computation time when the series with general term P_i slowly converges towards 1.*

20.2 Simulation of random variables with density

20.2.1 Uniform law on [a,b] with a<b∈R

If U is a uniform variable on $[0,1]$ then

$$X = a + (b-a) U \rightarrow \mathcal{U}([0,1])$$

20.2.2 Distribution function inversion method

Let P be a strictly positive probability density on $\mathbb{R}$ and

$$F(x) = \int_{-\infty}^{y} P(y)\, dy$$

the associated distribution function. Comme F est continue et strictement croissante sur $\mathbb{R}$ and check $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$, it admits an inverse function

$$F^{-1} :]0,1[\to \mathbb{R}$$

Example 107 *The exponential law of parameter $\lambda > 0$ is the density law $\lambda \exp(-\lambda x)$ if $x > 0$. The associated distribution function is*

$$F(x) = \int_{-\infty}^{x} \lambda \exp(-\lambda t)\, dt = (1 - \exp(-\lambda x)) \text{ if } x > 0.$$

For $u \in]0,1[$, we have

$$F(x) = u \Leftrightarrow x = F^{-1}(u) = -\frac{1}{\lambda} \ln(1-u).$$

We deduce that if

$$U \to \mathcal{U}([0,1]), \quad -\frac{1}{\lambda} \ln(1-U) \to \exp(\lambda).$$

and since we have $1-U$ and U have the same probability distribution such that UU ($1-U \to \mathcal{U}([0,1])$ so in this case

$$-\frac{1}{\lambda} \ln(1-U) \overset{loi}{=} -\frac{1}{\lambda} \ln(U) \to \exp(\lambda).$$

References

[1] Bruno Saussereau, Cours de théorie des probabilités avec exercices corrigés et devoirs, Année universitaire 2013-2014.

[2] Brigitte Bonnet, Cours de problème de convergence et approximations en probabilité. Lycée International de Valbanne. Avril 2011.

[3] Benjamin JOURDAIN. Probabilités et statistique pour l'ingénieur 10 janvier 2018.

[4] JULIEN JACQUES. Introduction aux série temporelle, Harner, New York 1965

[5] Jim Pitman, Probability. Springer, New York, 1993.

[6] Khaled Khaldi, Méthodes statistiques et probabilités. CASBAH Editions.

[7] Kaci Redjal, Cours de probabilités, L'office des publications universitaires, Algérie 1988.

[8] Marie-Pierre Béal-Nicolas Sendrier. Théorie de l'information et codag (Notes des coure) 21 novembre 2012.

[9] Michel Lejeune, Statistique La Théorie et ses applications. Springer, Paris, 2004.

[10] Sheldon M. Ross, Initiation aux probabilités. Presses polytechniques romandes, Lausanne, 1987.

[11] Stephan Morgenthaler, Introduction à la statistique. Presses polytechniques et universitaires romandes, Lausanne, 2ème édition, 2001.

[12] Sharon Goldwater, Basic probability theory University of Edinburgh 10 Sep 2018.

[13] OLIVIER GAUDOIN. Principes et Méthodes Statistiques.

[14] Yadolah Dodge, Premiers pas en statistique. Springer, Paris, 2003.