Statistiques descriptives et inférentielles

Abdallah Benaissa [1]

14 Avril 2020

[1] Professeur en mathématiques, chargé du module de biostatistiques à la faculté de médecine de l'université de Batna - Algérie

Table des matières

Preface

Ce polycopié est adressé aux étudiants de la première année en médecine. Il est conçu selon le nouveau programme officielle élaboré en Juin 2017.

Première partie

Rappels mathématiques.

Chapitre 1

Théorie des ensembles.

1.1 Ensemble, Elément.

1.1.1 Définitions

On appelle ensemble, toute liste où collection d'objets bien définis, explicitement où implicitement ; on appelle éléments ou membres de l'ensemble, les objets appartenant à l'ensemble et on note :

- $p \in A$ si p est un élément de l'ensemble A.
- $B \subset A$ ou $A \supset B$, si $x \in B \Rightarrow x \in A$.
- Si $B \subset A$, on dit que B est une partie de A, ou que B est un sous ensemble de A.

On définit un ensemble, soit en listant ses éléments : $A = \{1, 2, 3\}$, soit en donnant la définition de ses éléments : $X = \{x : x \text{ est un entier pair}\}$.

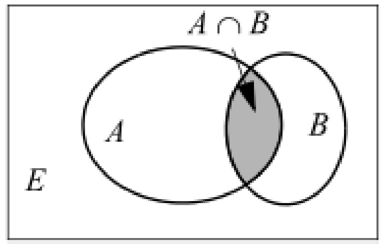
1.1.2 Notation

- la négation de $x \in A$ est $x \notin A$.
- $\emptyset$ est l'ensemble vide.
- E est l'ensemble universel.

1.1.3 Opération sur les ensembles

Soient A et B deux ensembles quelconques.

Intersection



L'intersection de A et B, notée $A \cap B$, est l'ensemble des éléments x tels que $x \in A$ et $x \in B$,

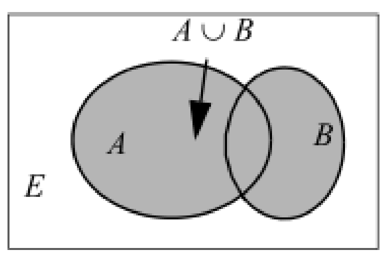
soit :

$$A \cap B = \{x : x \in A \text{ et } x \in B\}.$$

Le terme « et » est employé au sens, $x \in A$ et B, si x appartient à la fois à A et à B.

Cas particulier. Si $A \cap B = \emptyset$, on dit que A et B sont disjoints, on dit aussi en language probabiliste incompatible

Reunion

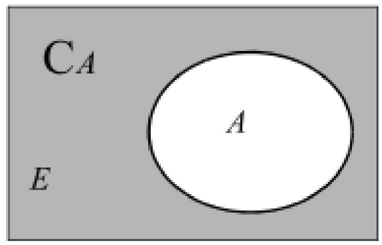


La réunion de A et B, notée $A \cup B$, est l'ensemble des éléments x tels que $x \in A$ ou $x \in B$, soit:

$A \cup B = \{x : x \in A \text{ ou } x \in B\}$.

le terme « ou » est employé au sens, $x \in A$ ou $x \in B$, si x appartient à A (uniquement), ou x appartient à B (uniquement), ou à la fois à A et à B.

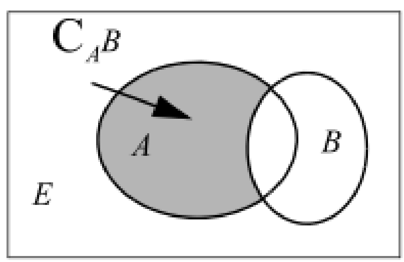
Complémentaire



Le complémentaire de A, noté CA ou $\overline{A}$, est l'ensemble des éléments de E qui n'appartiennent pas à A :

$$CA = \{x : x \notin A\}$$

Différence



La différence entre A et B, ou complémentaire de B relatif à A, noté $A - B$, ou C_A^B , est l'ensemble des éléments de A qui n'appartiennent pas à B :

$$A - B = \{x : x \in A \text{ et } x \notin B\}$$

1.1.4 Propriétés des opérations sur les ensembles

Loi idempotente

$$A \cup A = A; \qquad A \cap A = A$$

Loi associative

$$(A \cup B) \cup C = A \cup (B \cup C)\,; \qquad (A \cap B) \cap C = A \cap (B \cap C)$$

Loi commutative

$$A \cup B = B \cup A; \qquad A \cap B = B \cap A$$

Loi de distributivité

$$(A \cup B) \cap C = \quad (A \cap C) \cup (B \cap C)\,; \quad (A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

Loi d'identité

$$A \cup \emptyset = A; \qquad A \cap \emptyset = \emptyset$$
$$A \cup E = E; \qquad A \cap E = A$$

Loi de complémentarité

$$A \cup C\ A = E; \qquad A \cap C\ A = \emptyset$$
$$C\ C\ A = A; \qquad C\ E = \emptyset; \qquad C\ \emptyset = E$$

Loi de De Morgan

$$C\ (A \cup B) = C\ A \cap C\ B; \qquad C\ (A \cap B) = C\ A \cup C\ B$$

1.2 Ensembles particuliers

1.2.1 Ensemble fini

Un ensemble E est fini s'il est vide ou s'il contient un nombre fini d'éléments, sinon il est infini.

Exemple. L'ensemble E des cinq lettres a, b, c, d, e est fini

Exemple. l'ensemble des habitants de l'algérie est un ensemble fini (même s'il est relativement grand plus de trente millions).

1.2.2 Ensemble dénombrable

– Un ensemble est dénombrable s'il existe une bijection entre une partie de l'ensemble des entiers naturel $\mathbb{N}$ et cet ensemble . On a trois genre d'ensemble dénombrable :

a- L'ensemble vide $\emptyset$ est dénombrable par convention.

b- Un ensemble fini A, s'écrit évidemment sous la forme $\{a_1, a_2, ...a_n\}$, n étant un élément de $\mathbb{N}$.

c- Un ensemble dénombrable non fini A est un ensemble qui s'écrit sous la forme $\{a_1, a_2, ...\}$.

L'ensemble des entiers naturels N est évidemment un ensemble dénombrable non fini.

1.2.3 Ensemble infini non dénombrable.

Un ensemble A est non dénombrable s'il n'existe pas de bijection entre une partie de $\mathbb{N}$ et cet ensemble. Intuitivement çà signifie qu'il est plus grand que $\mathbb{N}$.

Exemple. Les ensembles infinis non dénombrales qu'on rencontrera dans notre contexte sont surtout, les intervalles de $\mathbb{R}$.

1.2.4 Ensembles produits

L'ensemble produit de deux ensembles A et B est l'ensemble noté $A \times B$ formé des couples ordonnés (a, b) où a est un élément de A et b est un élément de B :

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

Exemple 1. Si $A = \{1, 2, d\}$ et $B = \{a, d\}$, alors

$$A \times B = \{(1, a), (1, d), (2, a), (2, d), (d, a), (d, d)\} .$$

Exemple 2. Si $A = B = \mathbb{R}$, alors

$$A \times B = \mathbb{R}^2 = \{(x, y) : x \in \mathbb{R} \text{ et } y \in \mathbb{R}\}$$

1.2.5 Famille d'ensembles

Il peut arriver qu'on considère un ensemble dont les éléments sont eux mêmes des ensembles. Un tel ensemble est appelé une famille d'ensembles. On réprésente généralement une famille d'ensembles en fixant un ensemble I appelé ensemble d'indices et on associe à chaque indice $i \in I$, un élément de la famille qu'on note A_i, la famille d'ensembles est alors représentée par $A_i, i \in I$. Parties d'un ensemble

Soit A un ensemble quelconque. On note l'ensemble des parties de l'ensemble A, par $P(A)$.

Exemple 1. Si $A = \{a, b, d\}$ alors

$$P(A) = \{\emptyset, \{a\}, \{b\}, \{d\}, \{a, b\}, \{a, d\}, \{b, d\}, \{a, b, d\}\}$$

Union et intersection infinies

Soient $A_i, i \in I$ une famille infinie d'ensembles.

a- Union L'union des éléments de la famille $A_i, i \in I$, noté $\underset{i\in I}{\cup} A_i$, est l'ensemble des éléments qui appartiennent au moins à un certain A_i :

$$\underset{i\in I}{\cup} A_i = \{x : \exists i \in I, x \in A_i\}.$$

Si l'ensemble des indices est une partie $I = \{n \in \mathbb{N} : n \geq k\}$ de l'ensemble $\mathbb{N}$ des entiers naturels on utilise aussi la notation $\overset{\infty}{\underset{i=k}{\cup}} A_i$ ou $\underset{i\geq k}{\cup} A_i$.

Exemple 1. Soient pour tout entier naturel n, l'intervalle $A_n = [0, n[$. On a $\overset{\infty}{\underset{n=1}{\cup}} A_n = \mathbb{R}$.

Exemple 2. Soient pour tout entier naturel $k \geq 1$, l'intervalle $B_k = \left[\frac{1}{k}, k\right[$. on peut vérifier que $\underset{K\geq 1}{\cup} B_k = \mathbb{R}^* =]0, +\infty[$.

b- Intersection l'intersection des éléments de la famille $A_i, i \in I$, est l'ensemble des éléments qui appartiennent à tout élément A_i de la famille d'ensembles en question.

Exemples. Reprenons les ensembles $A_n = [0, n[$ et $B_k = \left[\frac{1}{k}, k\right[$, considérés précédemment.

a- $\overset{\infty}{\underset{n=1}{\cap}} A_n = [0, 1[$, car $A_1 = [0, 1[$ est contenu dans tous les autres ensembles $A_n, n > 1$.

b- $\bigcap\limits_{n=1}^{\infty} B_n = \emptyset$, car $B_1 = [1, 1[= \emptyset$.

1.3 Partition d'un ensemble

Une partition d'un ensemble A est une subdivision de l'ensemble A. Rigoureusement parlant, une partition de A est une famille de parties de A telle que l'union des éléments de cette famille est égale à A, et l'intersection de deux élements (quelconques) différents de cette famille est égale à l'ensemble vide.

Exemple 1. Soient $A = \{a, b, c, 4\}$, $A_1 = \{a\}$, $A_2 = \{a, b\}$, $A_3 = \{c, 4\}$, $A_4 = \{b, c\}$, $A_5 = \{4\}$.

A_1, A_2, A_4 est une famille de parties de A, qui ne forme pas une partition de A, car $A_1 \cap A_2 = \{a\} \neq \emptyset$.

A_2, A_5 est une famille de parties de A, qui n'est pas une partition de A, car $A_2 \cup A_3 = \{a, b, 4\} \neq A$.

A_2, A_3 est une partition de A.

Exemple 2. Soient A l'ensemble des êtres humains, A_1 l'ensemble des êtres humains de sexe masculin, A_2 l'ensemble des êtres humains de sexe féminin.

La famille A_1, A_2 est une partition de A.

Chapitre 2

Sommation de nombres et de fonctions.

2.1 Le symbole $\sum$.

2.1.1 Cas fini

Si $a_1, a_2, ..., a_n$ est une suite finie de nombres, on désigne la somme $a_1 + a_2 + ... + a_n$ par $\sum\limits_{i=1}^{n} a_i$:

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + ... + a_n. \tag{2.1}$$

Exemple 1.

$$\begin{aligned}
\sum_{i=1}^{5} \frac{1}{2^i} &= \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^4} + \frac{1}{2^5} \\
&= \frac{1}{2} \frac{1 - \left(\frac{1}{2}\right)^5}{\frac{1}{2}} \text{ (somme des cinq premiers termes d'une suite géométrique)} \\
&= 1 - \left(\frac{1}{2}\right)^5
\end{aligned}$$

Exemple 2.

$$\sum_{i=1}^{n} 2^i = 2^1 + 2^2 + ...2^n.$$

2.1.2 Cas infini

Si $a_1, a_2, ...$ est une suite infinie de nombres, on designe par $\sum\limits_{i=1}^{\infty} a_i$ la limite de la suite $S_n = \sum\limits_{i=1}^{n} a_i$:

$$\sum_{i=1}^{+\infty} a_i = \lim_n \sum_{i=1}^{n} a_i.$$

Exemple .

$$\begin{aligned} \sum_{i=1}^{+\infty} \frac{1}{2^i} &= \lim_n \left[\frac{1}{2^1} + \frac{1}{2^2} + ... + \frac{1}{2^n}\right] \\ &= \lim_n \frac{1}{2} \frac{1 - \left(\frac{1}{2}\right)^n}{1 - \frac{1}{2}} \text{ (somme des } n \text{ premiers termes d'une suite géométrique)} \\ &= 1. \end{aligned}$$

2.2 Le symbole $\int$.

2.2.1 Cas de fonction à valeurs positives.

Soit $f(x)$ une fonction à valeurs réelles non négatives $(f(x) \geq 0)$ définie sur un intervalle $[a, b]$. On désigne par $\int_a^b f(x)\,dx$, l'air sous le graphe de la fonction $f(x)$ entre les droites $x = a$ et $x = b$.

2.2.2 Exemple

$\int_1^3 (2x + 1)\,dx$ est l'air d'un trapèze. Cette aire est égale à $\int_1^3 (2x + 1)\,dx = \left(\frac{7+3}{2} \times 2\right) = 10$

2.2.3 Cas de fonction à valeurs non positives.

Si $f(x) \leq 0$, $\int_a^b f(x)\,dx$ designe l'air au dessus du graphe de la fonction $f(x)$, située entre les deux droites $y = a$ et $y = b$, avec le signe $-$.

2.2.4 Exemple

$\int_1^3 (-2x-1)\,dx$ est l'air du même trapèze avec le signe $-$. on a donc

$$\int_1^3 (-2x-1)\,dx = -10.$$

2.2.5 Cas général.

Si $f(x)$ a des valeurs positives et aussi des valeurs négatives sur l'intervalle $[a,b]$, on divise l'intervalle $[a,b]$ en n sous intervalles $[a,a_2[\,,[a_2,a_3[\,,\ldots[a_n,b[$ de sorte que le signe de la fonction $f(x)$ est, soit $\leq$ ou $\geq$ sur chaque sous intervalle de la subdivision. L'intégrale de $f(x)$ sur l'intervalle $[a,b]$ est la somme des integrales de $f(x)$ sur chaque sous intervalle de la subdivision.

2.2.6 Exemple

Soit la fonction $f(x)$ définie sur l'intervalle $[-2,7]$ comme suit

$$f(x) = \left\{ \begin{array}{c} \frac{x}{2};\ \text{pour } -2 \leq x < 2 \\ -x,\ \text{pour } 2 \leq x < 4 \\ x+b,\ \text{pour } 4 \leq x \leq 7 \end{array} \right\}$$

On a

$$\begin{aligned} \int_{-2}^{7} f(x)dx &= \int_{-2}^{0} f(x)dx + \int_{0}^{3} f(x)dx + \int_{3}^{5} f(x)dx + \int_{5}^{7} f(x)dx \\ &= -\frac{2}{2} + \frac{3\times 1}{2} - \frac{2\times 1}{2} + \frac{2\times 2}{2} = 1,5 \end{aligned}$$

2.2.7 Intégrale sur un intervalle non borné.

Soit une fonction $f(x)$ définie sur l'intervalle $[a,+\infty[$. Si la suite $S_n = \int_a^n f(x)\,dx$, $n > a$, converge vers une limite $S \in R$, On appelle cette limite l'intégrale indéfinie de la fonction $f(x)$ sur l'intervalle $[a,+\infty[$ et on note

$$\int_a^{+\infty} f(x)\,dx = \lim_n \int_a^n f(x)\,dx. \tag{2.2}$$

On définie de la même façon les intégrales indéfinies

$$\int_{-\infty}^{b} f(x)\,dx = \lim_{n} \int_{-n}^{+\infty} f(x)\,dx, (n \in \mathbb{N}, \quad -n < b) \tag{2.3}$$

$$\int_{-\infty}^{+\infty} f(x)\,dx = \lim_{n} \int_{-n}^{0} f(x)\,dx + \lim_{n} \int_{0}^{n} f(x)\,dx, (n \in \mathbb{N}, n \geq 1\ ) \tag{2.4}$$

2.2.8 Exemple

Considérons la fonction $f(x)$ définie sur l'intervalle $[1, +\infty[$ comme suit.

$$f(x) = \begin{cases} \frac{1}{2^1}, \text{ pour } \quad 1 \leq x < 2, \\ \frac{1}{2^2}, \text{ pour } \quad 2 \leq x < 3, \\ \qquad\qquad ... \\ \frac{1}{2^n}, \text{ pour } n \leq x < n+1. \end{cases}$$

$f(x)$ est constante sur tout intervalle $[n, n+1[\,,\ n = 1, 2, ...$

$$\begin{aligned} \int_{1}^{+\infty} f(x)\,dx &= \lim_{n} \int_{1}^{n} f(x)\,dx \\ &= \lim_{n} \sum_{k=1}^{n} \frac{1}{2^n} \\ &= \lim_{n} \frac{1}{2} \frac{1 - \left(\frac{1}{2}\right)^n}{1 - \frac{1}{2}} \\ &= 1. \end{aligned}$$

2.3 Primitive d'une fonction.

Soit une fonction $f(x)$ définie un intervalle $I = [a, b]$, $(a < b$, l'un où les deux pouvant être infinis), ou définie sur l'union d'intervalles.

On dit qu'une fonction $F(x)$, définie sur le même ensemble, est une primitive de la fonction $f(x)$, si la dérivée $F'(x)$ de la fonction F au point $x \in I$, est égale à $f(x)$, valeur de f au point x.

2.3.1 Notation

Si une fonction $F(x)$ est une primitive de la fonction $f(x)$, on écrit

$$F(x) = \int f(x)dx.$$

2.3.2 Remarque

Faites attention ! Il faut bien distinguer entre l'intégrale définie $\int_a^b f(x)dx$ et l'intégrale indéfinie $\int f(x)dx$. L'intégrale définie $\int_a^b f(x)dx$ est un nombre lié à l'intervalle $[a,b]$. l'intégrale indéfinie est une fonction qui à tout nombre $x \in [a,b]$ associe un nombre noté usuellement par $\int f(x)dx$.

2.3.3 Exemples.

1-

$$f(x) = x^2 + 1,$$

$$F(x) = \frac{x^3}{3} + x$$

est une primitive de $f(x)$. Et, en général si C est un nombre quelconque la fonction $\frac{x^3}{3} + x + C$ est une primitive de la même fonction

$$f(x) = x^2 + 1.$$

2- Si n est un entier différent de -1, alors

$$\int x^n dx = \frac{1}{n+1} x^{n+1}$$

3-

$$\int \sin x dx = \cos x + C.$$

4- Si $a \neq 0$,

$$\int \exp(ax)dx = \frac{1}{a}\exp(ax) + C.$$

2.3.4 Propriétés des primitives

1- En ajoutant un nombre quelconque à une primitive d'une fonction, on obtient une autre primitive de la même fonction : Si $F(x) = \int f(x)dx$ et C est un nombre, alors $F(x) + C = \int f(x)dx$.

2- Linéarité :

$$\begin{aligned} \int (f(x) + g(x))\, dx &= \int f(x)dx + \int g(x)dx, \text{ et} \\ \int \lambda f(x)dx &= \lambda \int f(x)dx \text{ (pour tout nombre } \lambda) \end{aligned} \tag{2.5}$$

2.3.5 Primitive et calcul d'intégrales

a- Si elle existe, alors la fonction $F(x) = \int_{-\infty}^{x} f(t)dt$ est une primitive de la fonction $f(x)$.

Ici x est fixé, c'est la borne supérieure de l'intervalle $]-\infty, x[$ sur laquelle on calcule l'intégrale, t est une lettre muette, elle n'a pas de sens toute seule.

b- Si une fonction $F(x)$ est une primitive de la fonction $f(x)$, alors on la formule suivante pour calculer l'intégrale définie de $f(x)$ sur l'intervalle $[a, b]$.

$$\int_a^b f(x)\, dx = F(b) - F(a) \tag{2.6}$$

Notation Dans les calculs, on exprime la quantité $F(b) - F(a)$ par $[F(x)]_a^b$.

2.3.6 Exemples

1-

$$\begin{aligned} \int_0^2 \left(x^2 + 1\right) dx &= \left[x^3 + x\right]_0^2 = \left(2^2 + 2\right) - \left(0^3 + 0\right) \\ &= 6 - 0 = 6. \end{aligned}$$

2-

$$\begin{aligned}\int_0^{\frac{\pi}{2}} \sin x dx &= [\cos x]_0^{\frac{\pi}{2}} \\ &= \cos\frac{\pi}{2} - \cos 0 \\ &= 0 - 1 = -1.\end{aligned}$$

3-

$$\begin{aligned}\int_1^5 Exp(-2x)dx &= \left[-\frac{1}{2}Exp(-2x)\right]_1^5 \\ &= \left(-\frac{1}{2}Exp(-2\times 5)\right) - \left(-\frac{1}{2}Exp(-2\times 1)\right) \\ &= -\frac{1}{2}\left(e^{-10} - e^{-2}\right).\end{aligned}$$

Remarque La formule (2.6) est valable même dans le cas où l'une ou les deux bornes a et b sont infinies, par exemple

$$\begin{aligned}\int_a^{+\infty} f(x)\,dx &= F(+\infty) - F(a). \qquad (2.7) \\ &= \left(\lim_n F(n)\right) - F(a).\end{aligned}$$

2.3.7 propriétés de l'intégrale.

soit $f(x)$ et $g(x)$ deux fonctions à valeurs réelles, λ un nombre réel.

Linéarité a- $\int_a^b \lambda f(x)\,dx = \lambda \int_a^b f(x)\,dx$
b- $\int_a^b (f(x) + g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx.$

Exemple 1. $\int_{-1}^1 (-x^3 + 2x^2 - x^5)\,dx = -\int_{-1}^1 x^3dx + 2\int_{-1}^1 x^2dx - \int_{-1}^1 x^5dx =$
$= -\left[\frac{x^4}{4}\right]_{-1}^{+1} + 2\left[\frac{x^3}{3}\right]_{-1}^{+1} - \left[\frac{x^6}{6}\right]_{-1}^{+1} =$
$= -\left(\frac{1^4}{4} - \frac{(-1)^4}{4}\right) + 2\left(\frac{1^3}{3} - \frac{(-1)^3}{3}\right) - \left(\frac{1^6}{6} - \frac{(-1)^6}{6}\right) =$
$= -(0) + 2\left(\frac{2}{3}\right) - (0) = \frac{4}{3}.$

Exemple 2. $\int_{-\pi}^{\pi}(-2\cos x+e^{-x}-x^2)\,dx=-2\int_{-\pi}^{\pi}(\cos x)\,dx+\int_{-\pi}^{\pi}e^{-x}dx-\int_{-\pi}^{\pi}x^2dx=$
$=-2\left[-\sin x\right]_{-\pi}^{\pi}+\left[-e^{-x}\right]_{-\pi}^{\pi}-\left[\frac{1}{3}x^3\right]_{-\pi}^{\pi}=$
$=-2\left\{(-\sin\pi)-(-\sin(-\pi)\right\}+\left\{(-e^{\pi})-(-e^{-\pi})\right\}-\left\{\frac{1}{3}\pi^3-\frac{1}{3}(-\pi)^3\right\}=$
$=-e^{\pi}+e^{-\pi}-\frac{2}{3}\pi^3.$

Relation de Chasles. Si c est un nombre quelconque et $f(x)$ est définie sur les trois intervalles $[a,c]$, $[c,b]$ et $[a,b]$, alors on a la relation suivante dite relation de chales.

$$\int_a^b f(x)\,dx=\int_a^c f(x)\,dx+\int_c^b f(x)\,dx. \tag{2.8}$$

Deuxième partie

Statistiques descriptives

La statistique est la science dont l'objet est de recueillir, de traiter et d'analyser des données issues de l'observation de phénomènes aléatoires, c'est-à-dire dans lesquels le hasard intervient. Une étude statistique passe généralement par trois étapes.

1)- Elle commence par l'identification du phénomène ou problème à étudier, par exemple pour étudier le problème du diabète en Algérie, on doit évaluer le taux de la glycémie, l'effet de ce fléau sur le fonctionnement du corps humain, la relation entre ce fléau et le rendement économique,...

2)- La deuxième étape concerne la collecte des données, généralement, sur un échantillon représentatif de la population à étudier. La collecte des données se fait, soit par des questionnaires (par exemple, le nombre d'enfants de familles, le niveau d'instruction,...), soit par des manipulations (la mesure de la glycémie, la mesure de la tension artérielle,...) .

3)- La troisième étape concerne l'analyse et l'interprétation des résultats. L'analyse des données est utilisée pour décrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet.

Les trois démarches qu'on vient de décrire concernent les statistiques descriptives. Il est important de noter que les statistiques descriptives utilisées il y a quelques siècles, se sont beaucoup développées aujourd'hui en donnant naissance à ce qu'on appelle les statistiques inférentielles.

En effet, on peut répartir la théorie des statistiques en deux grandes classes de statistiques, la deuxième étant un prolongement de la première :

Statistique descriptive, a pour but de recuillir les données et les résumer de façon synthétique et efficace. Elle utilise pour cela des représentations de données sous forme de graphiques, de tableaux et d'indicateurs numériques (par exemple des moyennes). Elle permet de dégager les caractéristiques essentielles du phénomène étudié et de suggérer des hypothèses pour une étude ultérieure plus sophistiquée.

Statistique inférentielle Le rôle de la statistique inférencielle est de fournir des outils mathématiques basés sur la théorie des probabilités, servant à généraliser les propriétés et les caractéristiques calculées sur l'échantillon à la population toute entière.

La statistique descriptive est dite univariée ou " à une dimension ", lorsqu'il s'agit de l'étude d'un seul caractère, sinon elle est dite multidimensionnelle.

Chapitre 3

Statistique unidimensionnelle

3.1 Définition

3.1.1 Population, unité et caractère statistiques.

Une population statistique est un ensemble d'êtres dont on s'intéresse à un de ses caractères, par exempe la population de la ville de Batna, l'ensemble des tremblements de terre en Algérie durant le $20^{ème}$ siècle.

L'unité statistique est un élément quelconque de la population statistique, par exemple l'unité statistique dans le cas de la population statistique des habitants de la ville de Batna est un habitant de Batna.

Le caractère statistique est la propriété des unités statistiques à étudier, par exemple la couleur des yeux, la taille pour la population des habitants de Batna. Le caractère statistique peut avoir plusieurs modalités (ou valeurs).

Le caractère étudié est quantitatif si ses modalités (ou ses valeurs possibles) sont mesurables, c'est-à-dire exprimées par des valeurs numériques, et il est qualitatif si ses modalités ne sont pas mesurables. Un caractère quantitatif est discret si ses modalités sont en nombre fini (ou dénombrables) et il est continu si ses valeurs possibles forment un ou plusieurs intervalles. Un caractère statistique qualitatif est ordinal si ses modalités peuvent être ordonnées (par exemple le stade d'une maladie), sinon il est nominal.

3.1.2 Recencement - Echantillonage

Pour étudier un caractère d'une population statistique, on ne peut pas en général, relever la valeur (ou la modalité) de ce caractère sur chaque individu de cette population. Alors, on considère une partie de taille raisonnable de cette population. Cette partie, sélectionnée à priori, s'appelle échantillon de la population et elle doit être représentative de la population. Les techniques utilisées pour assurer cette représentativité s'appellent "théorie de l'échantillonage".

Mais, parfois on doit considérer tous les éléments de la population statistique, on dit dans ce cas qu'il s'agit d'un recencement, par exemple le recencement de la population de l'Algerie qui se fait tous les dix ans (nombre d'habitants, repartition selon le sexe, selon les tranches d'âges, selon le niveau d'instruction...)

3.1.3 Série statistique simple

Une série statistique simple est une suite de nombres représentant les mesures d'un caractère quantitatif prises sur un échantillon d'une population statistique

3.2 Natures des variables statistiques et échelles de mesures.

Les données qualitatives définissent des échelles soit nominales soit ordinales. L'échelle nominale comporte un certain nombre de catégories, dont la seule propriété est qu'elles sont toutes différentes les unes des autres (sexe, nationalité, type de diplôme, etc.). Dans le cas d'une échelle ordinale, en revanche, les catégories qui la composent sont munies d'une structure d'ordre, établie en fonction d'un critère donné (de moins à plus "quelque chose" : origine sociale, opinion plus ou moins favorable, stade de développement d'une maladie,...).

Les données quantitatives se réfèrent à des variables d'intervalle ou de rapport. Dans le premier cas, on dispose d'échelles au sein desquelles la comparaison d'intervalles est possible (il est possible de déterminer si deux intervalles sont ou ne sont pas de même étendue). Dans le deuxième cas, l'échelle permet non seulement la comparaison d'intervalles, mais également la comparaison de rapports (il est possible de déterminer si deux rapports sont ou ne sont pas égaux). La différence fondamentale entre ces deux types d'échelles est liée au statut de la valeur nulle : sur une échelle d'intervalle, le zéro est situé de manière arbitraire, comme pour la mesure des températures par exemple (échelles Celsius et Fahrenheit). Sur une échelle de rapport, en revanche, le zéro a une signification précise, puisqu'il désigne l'absence du caractère considéré (âge, salaire, taille, vitesse, etc.).

3.2.1 Exemples.

Caractère quantitatif discret

On s'intéresse au nombre de modules obtenus par chaque étudiant de la premère année universitaire écoulée de la faculté de médecine. La population statistique est l'ensemble des étudiants de la première année de la faculté de médecine, un étudiant de la première année est une unité statistique, le caractère statistique étudié est le nombre de modules obtenus, c'est un caractère quantitatif discret.

Caractère quantitatif continu

Si au lieu du nombre de modules obtenus, on s'intérèsse à la taille de chaque étudiant, alors le caractère statistique "la taille" est un caractère quantitatif continu.

Caractère qualitatif nominal

Si au lieu du nombre de modules obtenus, on s'intérèsse à la couleur des yeux de chaque étudiant, alors le caractère statistique "couleur des yeux" est un caractère qualitatif nominal.

Caractère qualitatif ordinal

Si on considère le stade de la maladie (niveau de gravité de la maladie) chez un ensemble de cancéreux, alors on a affaire à un caractère statistique ordinal.

3.2.2 Echelles de mesures : Recapitulation par des exemples

Variables qualitatives nominales, **échelle nominale,** exemples : le nom des journaux, le métier ou le nom**.**

Variables qualitatives ordinales, **échelle ordinale,** exemples **:** le rang, stade d'une maladie.

Variables quantitatives discrètes, **échelle de rapport,** exemples : nombre d'enfants, nombre d'habitants.

Variables quantitatives discrètes, **échelle d'intervalle,** exemples **:** date de naissance, heure d'arrivée

Variables quantitatives continues, **échelle de rapport,** exemples : distance, durée.

Variables quantitatives continues, **échelle d'intervalle,** exemple : température.

3.3 La variabilité et l'incertain en biologie

3.3.1 Variabilité biologique et variabilité métrologique

La statistique constitue, en médecine, l'outil permettant de répondre à de nombreuses questions qui se posent en permanence au médecin :

Quelle est la valeur normale d'une grandeur biologique, taille, poids, glycémie ?

Quelle est la fiabilité d'un examen complémentaire ?

Quel est le risque de complication d'un état pathologique, et quel est le risque d'un traitement ?

Le traitement A est-il plus efficace que le traitement B ?

Toutes ces questions, proprement médicales, reflètent une propriété fondamentale des systèmes biologiques qui est leur variabilité. Cette variabilité est la somme d'une variabilité expérimentale (liée au protocole de mesure et aussi aux appareils utilisés) et d'une variabilité proprement biologique. On peut ainsi décomposer la variabilité d'une grandeur mesurée en deux grandes composantes :

$$\boxed{\text{variabilité totale} = \text{variabilité biologique} + \text{variabilité métrologique.}}$$

La variabilité biologique peut être elle-même décomposée en deux termes : d'une part la variabilité intra-individuelle, qui fait que la même grandeur mesurée chez un sujet donné peut être soumise à des variations aléatoires ; et d'autre part la variabilité inter-individuelle qui fait que cette même grandeur varie d'un individu à l'autre.

$$\boxed{\text{variabilité biologique} = \text{variabilité intra-individuelle} + \text{variabilité inter-individuelle.}}$$

La variabilité intra-individuelle peut être observée lors de la mesure de la performance d'un athlète qui n'est pas capable des mêmes performances à chaque essai, mais qui se différencie des autres athlètes (variabilité inter-individuelle). En général, la variabilité intra est moindre que la variabilité inter.

La variabilité métrologique peut être elle aussi décomposée en deux termes : d'une part les conditions expérimentales dont les variations entraînent un facteur d'aléas ; et d'autre part les erreurs induites par l'appareil de mesure utilisé.

$$\boxed{\text{variabilité métrologique} = \text{variabilité expérimentale} + \text{variabilité appareil de mesure}}$$

3.3.2 La décision dans l'incertain

Pour prendre une décision diagnostique ou thérapeutique, le médecin ne peut pas être sûr, cent pour cent d'avoir fait le bon choix, mais il y a des méthodes statistiques pour minimiser le risque de faire le mauvaix choix. Dans sa démarche, le médecin doit en particulier, prendre en compte la variabilité naturelle des grandeurs qu'il manipule, pour distinguer ce qui est normal de ce qui est pathologique (décision à propos d'un patient) et pour évaluer la qualité d'un nouvel examen, ou d'une nouvelle thérapeutique (décision thérapeutique). La compréhension des méthodes statistiques, de leur puissance et de leurs limites, est essentielle pour un médecin de nos jours. Tout résultat de recherche médicale résulte d'une expérimentation (clinique ou biologique) qui s'appuie sur une méthodologie statistique rigoureuse, et dont les résultats sont analysés en termes statistiques.

3.4 Représentation des données

Il y a trois principales façons de représenter les données concernant un caractère statistique : représentation brute (ou en vrac), dans un tableau statistique et la représentation graphique.

On développe dans ce qui suit, par des exemples, les différentes façons de représenter des données, selon la nature du caractère statistique.

3.4.1 Caractère qualitatif

Pour un caractère qualitatif, on présente les données dans un tableau à plusieurs lignes. Dans la première ligne, on reporte les différentes modalités, et dans la deuxième ligne les effectifs corespondants. On peut également ajouter d'autres lignes , par exemple une ligne pour les fréquences et une autre pour les poucentages correspondants.

Il y a plusieurs façons de représenter graphiquement un caractère qualitatif, mais les plus utilisées sont les diagrammes en secteurs circulaires ou les diagrammes en bandes.

Exemple

On reprend l'exemple de la couleur des yeux et on suppose que le nombre des étudiants est 200 et qu'il y a quatre couleurs différentes.

Les données sont représentées dans un tableau statistique simple

Couleur des yeux : x_i	noire	noisette	bleu	verte	Total
effectif : n_i	120	30	30	20	200

Ce tableau peut être développé, en y ajoutant, par exemple, des lignes pour les fréquences relatives et les poucentages relatifs. On obtient alors le tableau statistique suivant.

Couleur des yeux : x_i	noire	noisette	bleue	verte	Total
effectif : n_i	120	30	30	20	200
fréquences relatives	$\frac{120}{200}$	$\frac{30}{200}$	$\frac{30}{200}$	$\frac{20}{200}$	1
pourcentages relatifs	60	15	15	10	100

La représentation en bandes (en TD)
La représentation en secteurs circulaires (en TD)

3.4.2 Caractère quantitatif discret

Dans le cas d'un caractère discret, on collecte la valeur du caractère chez chaque individu de la population et on représente ces valeurs de deux façons :

La représentation brute des données consiste à écrire les valeurs des observations une aprés une sans aucune manipulation.

La représentation des données groupées par valeurs consiste à représenter les données dans un tableau avec au moins deux lignes, dans la première ligne on inscrit les valeurs et dans la deuxième les effectifs correspondants. Parfois on ajoute d'autres lignes, par exemple pour les effectifs cummulés, les fréquences,... .

La représentation graphique se fait par le diagramme en bâtons **(en TD).**

Exemple.

Le recueil d'informations concernant le nombres d'enfants pour vingt famille à donné la suite de nombres suivante

$$3\ 2\ 1\ 0\ 2\ 3\ 4\ 0\ \ 2\ 3\ 1\ 0\ 3\ 5\ 6\ 6\ 5\ 6\ 5\ 4.$$

Cette suite est une réprésentation brute de la série des nombres d'enfants des vingt familles.

La représentation groupée par valeurs est

valeurs (x_i)	0	1	2	3	4	5	6	Total
effectifs (n_i)	3	2	3	4	2	3	3	20

On peut ajouter à ce tableau d'autres lignes, par exemple, une ligne pour les effectifs cummulés (croissants) et une ligne pour les fréquences. On obtient alors le tableau suivant :

valeurs (x_i)	0	1	2	3	4	5	6	Total
effectifs (n_i)	3	2	3	4	2	3	3	20
effectifs cummulés (n_icum)	3	5	8	12	14	17	20	
fréquences (f_i)	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{4}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{3}{20}$	1

On peut également ajouter d'autres lignes pour y faire des calculs concernant le calcul de paramètres comme la moyenne m ou la variance σ^2.

La représentation des données groupées par valeurs peut aussi se faire par une suite de couples $(x_1; n_1)$, $(x_2; n_2)$, ..., $(x_k; n_k)$, où les x_i sont les valeurs de la série et les n_i les effectifs. Néamoins la représentation dans un tableau est la plus commode, soit pour les calculs, soit pour éviter les subtilités.

3.4.3 Caractère quantitatif représenté en classes de valeurs.

Exemple

Les notes en mathématiques des élèves d'une classe de vingt élèves sont comme suit :

les notes de 4 élèves sont ≥ 6 et < 9, les notes de 7 élèves ≥ 9 et < 12, les notes de 6 élèves sont ≥ 12 et < 15, et 3 élèves avaient une note ≥ 15 et < 18.

La représentation de la série des notes de ces élèves en classes de valeurs d'amplitude 3 est

classes (C_i)	$[6;9[$	$[9;12[$	$[12;15[$	$[15;18[$	total
effectifs (n_i)	4	7	6	3	20

La représentation graphique d'un caractère, dont les valeurs sont représentées en classes et non en valeurs discrètes, se fait par un histogramme **(en TD)**. Dans un histogramme on représente les classes par des intervalles sur l'axe des abscisses, et on représente l'effectif (ou la fréquence de chaque classe) par un rectangle dont la superficie est proportionnelle à l'effectif (ou à la fréquence).

3.5 Les différents types de tableaux statistiques

Il y a trois types de tableaux en statistiques.

3.5.1 Tableaux de données.

Ces tableaux sont confectionnés pour collecter les données d'un ou de plusieurs caractères sur les sujets, d'une population ou d'un échantillon d'une population.

Exemple

Dans une classe de 15 élèves on a relevé sur le tableau suivant, les données concernant le sexe, le niveau de la langue française et la nationalité.

N°	sexe		français			moyenne générale			
	M	F	A	B	C	[7; 10[	[10; 13[	[13; 16[	[16; 19[
1	x			x		x			
2		x		x			x		
3	x		x		x		x		
4		x			x			x	
5	x				x		x		
6		x		x				x	
7		x	x				x		
8	x		x				x		
9		x		x			x		
10	x		x					x	
11		x	x						x
12		x		x			x		
13		x		x				x	
14		x	x			x			
15	x			x					x

dans ce tableau, on a utilisé les conventions suivantes.
Pour le sexe, M = sexe masculin, F = sexe féminin.
Pour la langue française, A = bon, B = moyen, C= faible.

3.5.2 Tableaux de distributions statistiques.

Ce type de tableaux est le plus connu. Il utilise généralement un tableau des données, pour exprimer la distribution statistique (c'est-à-dire la distibution des effectifs ou des fréquences) d'un ou de plusieurs caractères statistiques. Les tableaux statistiques déja rencontrés dans ce cours sont de ce type.

Exemple.

L'utilisation du tableau des données de l'exemple précédent donne le tableau suivant de la distribution statistique des moyennes générales des quinze élèves.

moyennes générales (x_i)	$[7;10[$	$[10;13[$	$[13;16[$	$[16;19[$	total
effectifs (n_i)	2	7	4	2	15

3.5.3 Tableau de contingence.

Ces tableaux sont constitués par croisement de deux variables statistiques. Leur élaboration nécessite de revenir au tableau initial de données et le dénombrement des sujets présentant simultanément deux valeurs des variables considérées. La fonction d'un tableau de contingence est de permettre de tester l'indépendance ou le lien entre deux variables.

Exemple.

Pour étudier la relation entre le sexe et la réussite scolaire des quinze élèves de l'exemple précédent, l'utilisation du tableau des données dans le même exemple donne le tableau de contingence suivant.

sexe\M.G	$[7;10[$	$[10;13[$	$[13;16[$	$[16;19[$	totaux
masculin	1	3	1	1	6
feminin	1	4	3	1	9
totaux	2	7	4	2	15

3.5.4 Utilisation des effectifs cummulés.

La ligne des effectifs cummulés dans un tableau statistique est utilisée pour trouver la valeur d'une observation dont on fixe le numéro : x_1 est la valeur des observations numéro $1, 2, ..., n_1cum$,

x_2 est la valeur des observations numéro $n_1cum + 1, n_1cum + 2, ..., n_2cum$,

...

x_r est la valeur des observations numéro $n_{r-1}cum + 1, n_{r-1}cum + 2, ..., n_rcum$.

Exemple.

Les données concernant le nombre d'enfants pour 200 familles sont présentées dans le tableau suivant

nombres d'enfants x_i	3	4	5	6	Total
effectif n_i de familles ayant x_i enfants	120	30	30	20	200

On ajoute à ce tableau une ligne pour les effectif cummulés, on obtient alors le tableau statistique suivant.

nombre d'enfants : x_i	3	4	5	6	Total
effectif : n_i	120	30	30	20	200
effectifs cummulés	120	150	180	200	

Dans cet exemple, on a
Le nombre des valeurs (ou modalités) est $r = 4$
Le nombre total d''observations ou l'effectif total est

$$\begin{aligned} N &= \sum_{1}^{4} n_i \\ &= 120 + 30 + 30 + 20 \\ & \quad 200. \end{aligned}$$

Les quatres modalités sont $x_1 = 3$, $x_2 = 4$, $x_3 = 5$, et $x_4 = 6$.
$n_1 cum = n_1 = 120$,
$n_2 cum = n_1 cum + n_2 = 120 + 30 = 150$,
$n_3 cum = n_2 cum + n_3 = 150 + 30 = 180$,
$n_4 cum = n_3 cum + n_4 = 180 + 20 = 200$.

le numéro 61 est compris entre 1 et $n_1 cum = 120$, donc la valeur de l'observation numéro 61 est égale à 3.

La valeur de l'observation numéro 140 est égale à 4, car 140 est dans l'intervalle des entiers $[n_1 cum + 1; n_2 cum[= [121; 150[$.

3.6 Caractéristiques de tendance centrale

3.6.1 La moyenne arithmétique.

Définition

La moyenne arithétique noté $\overline{x}$ ou m, d'une série statistique simple représentée en données groupées en valeurs " $(x_1, n_1), (x_2, n_2), ..., (x_p, n_p)$ " est donnée par la formule

$$\overline{x} = \frac{\sum_{i=1}^{p} n_i x_i}{\sum_{i=1}^{p} n_i}. \tag{3.1}$$

Ici, p est le nombre des valeurs de la série statistique X, n_i est le nombre de repétition de la valeur x_i.

Remarque.

Si chaque valeur apparait une seul fois, ou si on répète dans le numérateur chaque x_i, n_i fois, la formule (3.1) devient

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}, \tag{3.2}$$

où N est le nombre total des termes de la série statistique, c'est-à- dire le nombre d'observations.

Exemple.

Soit les deux séries statistiques

$$X : 1,5; -1; 2; -1; 1,5; 2; 3; 1,5; 1,5; 1; -1$$
$$Y : 3; 2; 4; 1; 1; -2; 8; 7; 0; 10; -3$$

Les deux séries statistiques sont exprimées en représentation brute.

Pour la série X, puisque plusieurs valeurs de la série se répète, il vaut mieux utiliser la représentation des données groupées par valeurs.

x_i	-1	1	2	3	1.5
n_i	3	1	2	1	4

Pour la série Y, on ordonne les valeurs dans un ordre non décroissant,

$$Y : -3; -2; 0; 1; 2; 3; 4; 7; 10.$$

La moyenne arithmétique de X est donnée par la formule

$$\begin{aligned}\overline{x} &= \frac{\sum_{i=1}^{5} n_i x_i}{\sum_{i=1}^{5} n_i} \\ &= \frac{(3\times -1)+(1\times 1)+(2\times 2)+(1\times 3)+(4\times 1.5)}{3+1+2+1+4} \\ &= \frac{11}{11} = 1.\end{aligned}$$

La formule (3.2) s'écrit ici

$$\begin{aligned}\overline{x} &= \frac{\sum_{i=1}^{N} x_i}{N} \\ &= \frac{(-1)+(-1)+(-1)+1+2+2+3+1.5+1.5+1.5+1.5}{11} \\ &= \frac{11}{11} = 1.\end{aligned}$$

Pour calculer la moyenne $\overline{y}$ de la série Y, puisque chaque valeur de la série apparait uniquement une seule fois, on utilise la formule (3.2) pour avoir

$$\begin{aligned}\overline{y} &= \frac{\sum_{i=1}^{N} y_i}{N} \\ &= \frac{(-3)+(-2)+0+1+2+3+4+7+10}{9} \\ &= \frac{22}{9} = 2.44.\end{aligned}$$

Cas de données groupées en classes de valeurs. Si la série est représentée en classes de valeurs, on accepte une approximation de la moyenne par la même formule (3.1) , où

$$x_i = \frac{e_{i-1}+e_i}{2}$$

est le centre de la classe $[e_{i-1}, e_i[$.

3.6.2 Exemple

Soit une série statistique X, dont les données sont réprésentées par le tableau suivant, où C_i représente la classe (i.e. l'intervalle) $[e_{i-1}, e_i[$

c_i	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$
n_i	3	2	2	1	3

Pour calculer la moyenne on ajoute une ligne dans laquelle on exprime les centres de classe, qu'on note x_i et on utilise la formule (3.1).

C_i	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$
n_i	3	2	2	1	3
x_i	-1	1	3	5	7

$$\begin{aligned}
\overline{x} &= \frac{\sum_{i=1}^{5} n_i x_i}{\sum_{i=1}^{5} n_i} \\
&= \frac{(3\times -1)+(2\times 1)+(2\times 3)+(1\times 5)+(3\times 7)}{3+2+2+1+3} \\
&= \frac{31}{11}.
\end{aligned}$$

3.6.3 La médiane

Définition. La médiane d'une série est une valeur qui partage cette série, préalablement ordonnée, en deux sous séries ayant le même effectif. La mediane peut être une valeur de la série ou non.

Calcul de la médiane.

1)- Cas de série représentée par valeurs discrètes Si l'effectif N est impair la médiane M_e est égale à l'observation numéro $\frac{N-1}{2}+1$:

$$M_e = \left(\frac{N-1}{2}+1\right)^{\grave{e}me} obs. \tag{3.3}$$

Si l'effectif est pair la médiane M_e est la moyenne des deux observations numéro $\frac{N}{2}$ et $\frac{N}{2}+1$:

$$M_e = \frac{1}{2}\left[\left(\frac{N}{2}\right)^{\grave{e}me} obs. + \left(\frac{N}{2}+1\right)^{\grave{e}me} obs.\right] \tag{3.4}$$

2)- Cas de série représentée par classes de valeurs On définit d'abord la classe médiane.

Définition de la classe médiane. La classe médiane est la classe contenant l'observation numéro $\frac{N}{2}$ si l'effectif total N est pair, ou l'oservation numéro $\frac{N+1}{2}$ si N est impair.

La médiane est un nombre qu'on calcule par La méthode d'interpolation linéaire. Cette méthode suppose que les observations de la classe médiane sont réparties d'une façon uniforme.

Définition de la valeur médiane La médiane M_e est définie par

$$M_e = e_{i-1} + k\frac{\frac{N}{2} - n_{i-1}cum}{n_i cum - n_{i-1}cum}, \tag{3.5}$$

où

e_{i-1} est la borne inférieure de la classe mediane,

k est l'amplitude de la classe médiane.

$n_{i-1}cum$ est l'effectif cummulé de la classe qui précède la classe médiane.

$n_i cum$ est l'effectif cummulé de la classe médiane.

On détermine la classe médiane de la façon suivante.

On ajoute une ligne pour les effectifs cummulés, pour l'utiliser dans la détermination de l'observation numéro $\frac{N}{2}$ ou $\frac{N+1}{2}$.

3.6.4 Exemple

On considère la série suivante, représentée en classes de valeurs.

c_i	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$
n_i	3	2	2	1	3

Pour déterminer la médiane on commence par ajouter au tableau une ligne des effectifs cummulés :

c_i	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$
n_i	3	2	2	1	3
$n_i cum$	3	5	7	8	11

L'effectif total est $N = 11$, c'est un entier impair.

$$\frac{N+1}{2} = \frac{11+1}{2} = 6$$

La classe médiane est celle contenant l'observation numéro 6, la classe médiane est donc $[2;4[$.

La médiane M_e par interpolation linéaire est

$$\begin{aligned} M_e &= e_{i-1} + k\frac{\frac{N}{2} - n_{i-1}cum}{n_i cum - n_{i-1}cum} \\ &= 2 + 2\frac{5.5-5}{7-5} = 2.5. \end{aligned}$$

3.6.5 Le mode

Définition Le mode d'une série statistique simple est la valeur qui a le plus grand effectif. Une série peut avoir plusieurs modes.

Cas des données groupées en classes de valeurs Contrairement au cas des données discrètes, dans le cas des données groupées en classes de valeurs, le mode n'est pas toujours l'une des valeurs de la série, on se satisfait dans cette situation à définir la classe modale qui est la classe ayant le plus grand effectif.

3.6.6 Exemple.

Considérons l'exemple

c_i	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$
n_i	3	2	5	1	3

La classe modale est $[2;4[$.

3.7 Caractéristiques de dispersion

3.7.1 L'étendue

L'étendue E est la longueur de l'intervalle sur laquelle sont réparties les valeurs de la série, c'est donc la différence entre la plus grande et la plus petite valeur de la série.

$$E = Max\,(x_i) - Min\,(x_i) \tag{3.6}$$

.

Si la série est représentée en classes de valeurs, on prend comme étendue la différence entre la borne supérieure e_p de la dernière classe $[e_{p-1}; e_p[$ et la borne inférieure de la première classe $[e_1; e_2[$

$$E = e_p - e_1 \tag{3.7}$$

3.7.2 Quantile

Un quantile d'ordre $0 < \alpha < 1$, noté x_α, d'une série quantitative, formée de N observations, est un nombre noté x_α, tel que, à peu prés, αN des observations formant la suite sont inférieures à ce quantile x_α, et le reste, c'est-à-dire à peu prés $(1-\alpha)\,N$ observations, sont supérieures à x_α.

1er quartile

Le premier quartile, noté Q_1,est le quantile d'ordre $\alpha = \frac{1}{4}$.

3eme quartile

Le troisième quartile noté Q_3 est un nombre tel que 25% des observations lui sont supérieures (donc situées à sa droite).

L'étendue interquartile

L'étendu interquartile noté I_Q est la différence entre le troisième quartile Q_3 et le premier quartile Q_1 :

$$I_Q = Q_3 - Q_1 \tag{3.8}$$

Décile

Le premier décile est le quantile d'ordre $\alpha = \frac{1}{10}$, le deuxième décile est le quantile d'ordre, $\frac{2}{10}, \ldots.$

Centiles

Le premier centile est le quantile d'ordre $\alpha = \frac{1}{100}$, le deuxième décile est le quantile d'ordre, $\frac{2}{100}, \ldots.$

3.7.3 Calcul des quartiles : Données discrètes

L'effectif total N n'est pas toujours un multiple de quatre, et ainsi le un quart de l'effectif total n'existe pas toujours. Ainsi, on calcule approximativement le premier et le troisième quartile de la façon suivante.

Calcul de Q_1.

Si $\frac{N}{4}$ est un entier naturel, Q_1 est l'observation numéro $\frac{N}{4}$

$$Q_1 = \left(\frac{N}{4}\right)^{ème} \quad obs. \tag{3.9}$$

Si $\frac{N}{4}$ n'est pas un entier naturel, Q_1 est l'observation correspondant au numéro qui vient juste aprés le nombre décimal $\frac{N}{4}$.

Calcul de Q_3.

Si $\frac{3N}{4}$ est un entier naturel, Q_3 est l'observation numéro $\frac{3N}{4}$

$$Q_3 = \left(\frac{3N}{4}\right)^{ème} \quad obs. \tag{3.10}$$

Si $\frac{3N}{4}$ n'est pas un entier naturel, Q_3 est l'observation correspondant au numéro qui vient juste aprés le nombre décimal $\frac{3N}{4}$.

3.7.4 Exemple 1. (données en vrac).

1)- Soit la série

$$2\ 2\ 3\ 4\ 5\ 6\ 6\ 6\ 7\ 7\ 7\ 8\ 8\ 10$$

L'effectif total est

$$N = 14 \text{ est pair.}$$

La médiane M_e est la moyenne des deux observations $\left(\frac{N}{2}\right)^{ème}$ et $\left(\frac{N}{2}+1\right)^{ème}$:

$$M_e = \frac{6+6}{2} = 6.$$

Puisque $\frac{N}{4} = \frac{14}{4} = 3,5$ n'est pas un nombre entier et également $\frac{3N}{4}$, le premier quartile Q_1 et le troisième quartile Q_3 sont

$$Q_1 = 4^{ème} \text{ obs.} = 4. \text{ et } Q_3 = 11^{ème} \text{ obs.} = 7$$

et ainsi

$$I_Q = 7 - 4 = 3.$$

3.7.5 Exemple 2. (données groupées en valeurs).

Soit la série

Valeurs (x_i)	-2	0	1,5	4	5	10	11	15
effectifs (n_i)	10	2	40	8	5	15	5	15

Pour trouver les valeurs d'observations des rangs concernés, on doit calculer les effectifs cummulés, en ajoutant une autre ligne au tableau des données :

Valeurs (x_i)	-2	0	1,5	4	5	10	11	15	total
effectifs (n_i)	10	2	40	8	5	15	5	15	100
effectifs cummulés (n_i cum)	10	12	52	60	65	80	85	100	

L'effectif total est

$$N = 100, \text{ il est pair.}$$

la mediane M_e est

$$M_e = \frac{1,5 + 1,5}{2} = 1,5.$$

Le premier quartile Q_1 est

$$Q_1 = \left(\frac{100}{4} = 25\right)^{ème} obs. = 1,5.$$

Le troisième quartile Q_3 est

$$Q_3 = \left(\frac{3 \times 100}{4} = 75\right)^{ème} obs. = 10.$$

L'intervalle interquartile I_Q est

$$I_Q = Q_3 - Q_1 = 10 - 1,5 = 8,5.$$

3.7.6 Calcul des quartiles : Données groupées en classes de valeurs.

Le premier et le troisième quartiles se calcule, comme la médiane, par la méthode d'interpolation linéaire.

Calcul de Q_1.

On détermine la classe $[e_{i-1}, e_i[$ du premier quartile Q_1, qui est la classe contenant l'observation numéro $\frac{N}{4}$ (ou le numéro qui vient juste aprés le nombre décimal $\frac{N}{4}$) et on applique la formule

$$Q_1 = e_{i-1} + k\frac{\frac{N}{4} - n_{i-1}cum}{n_icum - n_{i-1}cum}, \tag{3.11}$$

où

e_{i-1} est la borne inférieure de la classe du premier quartile,
k est l'amplitude de la même classe.

Calcul de Q_3.

On détermine la classe $[e_{i-1}, e_i[$ du troisième quartile Q_3, qui est la classe contenant l'observation numéro $\frac{3N}{4}$ (ou le numéro qui vient juste aprés le nombre dicimal $\frac{3N}{4}$) et on applique la formule

$$Q_3 = e_{i-1} + k\frac{\frac{3N}{4} - n_{i-1}cum}{n_icum - n_{i-1}cum}, \tag{3.12}$$

où

e_{i-1} est la borne inférieure de la classe du troisième quartile,
k est l'amplitude de la même classe.

3.7.7 Exemple 3.

On considère la série définie par le tableau suivant.

Classe C_i	$[-4;-2[$	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$	total
effectif n_i	6	4	8	5	10	1	34

L'effectif total est N=34.
L'etendu E est

$$E = 8 - (-4) = 12.$$

Pour les quartiles, et les quantiles en général, on ajoute au tableau une ligne pour les effectifs cummulés :

Classe C_i	$[-4;-2[$	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6;8[$	total
effectif n_i	6	4	8	5	10	1	34
effectifs cummulés n_icum	6	10	18	23	33	34	

a)-La médiane :
$\frac{N}{2} = \frac{34}{2} = 17$, la classe médiane est la classe qui contient l'observation numéro 17, c'est donc la classe $[e_{i-1}; e_i[= [0;2[$. L'amplitude de classe est $k = 2$.

La médiane M_e est

$$\begin{aligned} M_e &= e_{i-1} + k\frac{\frac{N}{2} - n_{i-1}cum}{n_icum - n_{i-1}cum} \\ &= 0 + 2\frac{17-10}{18-10} \\ &= 1.75. \end{aligned}$$

b)- Le premier quartile $\mathbf{Q}_1$:

$\frac{N}{4} = 8.5$, la classe du premier quartile est celle contenant l'observation numéro 9, c'est la classe $[-2;0[$. Le premier quartile Q_1 est

$$\begin{aligned} Q_1 &= e_{i-1} + k\frac{\frac{N}{4} - n_{i-1}cum}{n_icum - n_{i-1}cum} \\ &= -2 + 2\frac{8.5-6}{10-6} \\ &= -0.75. \end{aligned}$$

c)- Le troisième quartile $\mathbf{Q}_3$:

$\frac{3N}{4} = 25.5$, la classe du troisième quartile est celle contenant l'observation numéro 26, c'est la classe $[4;6[$. Le premier quartile Q_3 est

$$\begin{aligned} Q_3 &= e_{i-1} + k\frac{\frac{3N}{4} - n_{i-1}cum}{n_icum - n_{i-1}cum} \\ &= 4 + 2\frac{25.5-23}{33-23} \\ &= 4.5. \end{aligned}$$

d)- L'intervalle interquartile I_Q

$$\begin{aligned} I_Q &= Q_3 - Q_1 \\ &= 4.5 - (-0.75) \\ &= 5.25. \end{aligned}$$

3.7.8 Boite à moustache

Définition.

La boite à moustache (ou box plot) est un graphique qui résume la dispersion d'une série statistique. Elle est définie par cinq nombres, la valeur minimale et la valeur maximale (les moustaches), les deux quartiles Q_1, Q_2 et la médiane M_e (la moustache).

Exemple.

Considérons les trois séries statistiques A, B et C suivantes.
A : 1, 1.5, 1.5, 2, 2.5, 3, 3, 3, 3, 3.4, 4.5, 4, 4.8, 5, 5, 5.
B : 1, 1, 1, 1, 1, 1.2, 1.2, 1.3, 2, 2.3, 2.4, 2.5, 3, 4, 4, 5, 5.
C : 1, 3, 3, 3, 3.2, 3.2, 3.4, 3.5, 4.5, 4.6, 4.7, 5, 5, 5, 5, 5.
Les boites à moustaches des trois séries sont représentées dans (Figure 1)
.

Utilité de la boite à moustache.

La boite à moustache peut être utilisée pour comparer la dispersion de séries, relativement à la médiane.

On obtient de la boite à moustaches des trois séries A, B, C, représenté dans Figure 1, une idée sur la dispersion des observations de chaque série, relativement à la médiane.

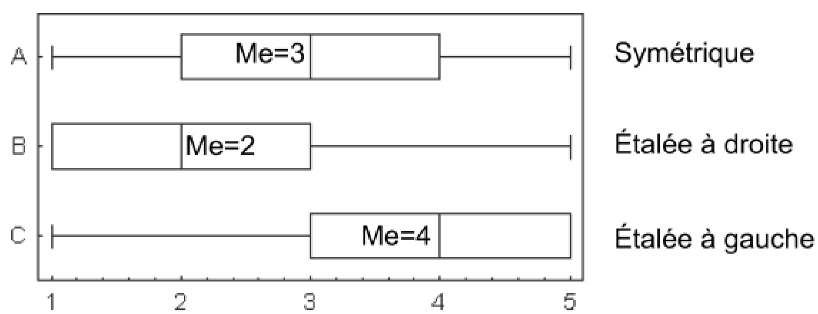


Fig. 3.1 – Les boites à moustaches des trois séries statistiques A, B et C.

3.7.9 Variance et Ecart-type

La variance et l'écart-type donne une idée sur l'éloignement des valeurs par rapport à la valeur moyenne.

Définition. La variance σ^2 d'une série statistique $\{x_1, ...x_N\}$ de moyenne m est définie par

$$\sigma^2 = \frac{\sum_{i-1}^{N} (x_i - m)^2}{N}. \tag{3.13}$$

L'écart-type σ est défini par

$$\sigma = \sqrt{\sigma^2}. \tag{3.14}$$

Si les données de la série (on dit aussi la distribution statistique) sont groupées en valeurs $\{(n_{1;}x_1), ..., (n_p; x_p)\}$, la définition (3.13) devient

$$\sigma^2 = \frac{\sum_{i-1}^{p} n_i (x_i - m)^2}{\sum_{i-1}^{p} n_i}. \tag{3.15}$$

Si les données de la série (on dit aussi la distribution statistique) sont groupées en classes de valeurs, la variance et l'écart-type sont définies par les mêmes formules où x_i est le centre de la classe d'effectif $n_{i.}$.

3.7.10 Théorème de Konig-Huygens

On peut démontrer que l'expression (3.15) définissant la variance est algébriquement équivalente à l'expression suivante, appelée formule de Konig-Huygens.

$$\sigma^2 = \frac{\sum_{i-1}^{p} n_i x_i^2}{\sum_{i-1}^{p} n_i} - m^2. \tag{3.16}$$

Coefficient de variation

Le coefficient de variation C_v est le rapport entre l'écart-type et la moyenne.

$$C_v = \frac{\sigma}{m}. \tag{3.17}$$

Le coefficient de variation donne une meilleure idée de la dispersion des valeurs de la série que l'écart-type, car c'est une valeur relative et non une valeur absolue.

3.7.11 Exemples

Exemple 1.

Soit la série statistique

1.5 ; 2.0 ; 3 ; 4 ; 5 ;8 ; 8 ; 9 ; 10 :12 ; 12.5 ; 15.

L'effectif total est $N = 12$.

Puisque les valeurs des mêmes observations ne se répète pas beaucoup, il n'est pas necéssaire d'utiliser la représentation des données groupées en valeurs.

$$\begin{aligned} m &= \frac{\sum_{i=1}^{12} x_i}{N} \\ &= \frac{1.5 + 2.0 + ... + 12.5 + 15}{12} \\ &= \frac{90}{12} = 7.5. \end{aligned}$$

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^{12} x_i^2}{N} - m^2 \\ &= \frac{1.5^2 + 2.0^2 + ... + 12.5^2 + 15^2}{12} - (7.5)^2 \\ &\simeq 17.96. \end{aligned}$$

$$\begin{aligned} \sigma &= \sqrt{\sigma^2} \\ &= \sqrt{17.96} \\ &\simeq 4.24. \end{aligned}$$

$$\begin{aligned} C_V &= \frac{\sigma}{m} \simeq \frac{4.24}{7.5} \\ &\simeq 0.56. \end{aligned}$$

Exemple 2.

Soit la série

-5 ; -8 ; -3 ; 0 ; 1 ; 0 ; 1 ; 2 ; 2 ; -3 ; 4 ; 0 ; 1 ; 2 ; -3 ; -3 ; 1 ;0 ; 0 ; 0 ; 0 ; 0 ; 0

-8 ; -8 ; -3 ; 2 ; 4 ; 4 ; 2 ; 3 ; -5 ; -5 ; 2 ; 2 ; 2 ; 4 ; 2 ; 3 ; 1 ; 5 ; 0 ; 0

Puisque beaucoup d'observations ont la même valeur, on utilise la représentations des données groupées en valeurs

x_i	-8	-5	-3	0	1	2	3	4	5	total
n_i	3	3	5	11	5	9	2	4	1	43

Le nombre des valeurs est $p = 9$.
Le nombre d'observations (c'est-à-dire l'effectif total) est $N = 43$.
La moyenne m est donnée par

$$\begin{aligned} m &= \frac{\sum_{i=1}^{9} n_i x_i}{\sum_{i=1}^{9} n_i} \\ &= \frac{3(-8) + 3(-5) + ... + 4(4) + 1(5)}{3 + 3 + ... + 4 + 1} \\ &= -\frac{4}{43} \\ &\simeq -0.093. \end{aligned}$$

La variance σ^2 et l'écart type σ sont donnés par

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^{9} n_i x_i^2}{\sum_{i=1}^{9} n_i} - m^2 \\ &= \frac{3(-8)^2 + 3(-5)^2 + ... + 4(4)^2 + 1(5)^2}{3 + 3 + ... + 4 + 1} - (-0.093)^2 \\ &\simeq 10.689. \end{aligned}$$

$$\sigma = \sqrt{\sigma^2} \simeq \sqrt{10.689} \simeq 3.269$$

Le coefficient de variation C_V est égale à

$$C_V = \frac{\sigma}{m} \simeq \frac{3.269}{-0.093} = -35.146.$$

On peut utiliser un tableau statistique pour y faire les calculs, de la façon suivante.

x_i	-8	-5	-3	0	1	2	3	4	5	*total*
n_i	3	3	5	11	5	9	2	4	1	43
$n_i x_i$	-24	-15	-15	0	5	18	6	16	5	-4
$n_i x_i^2$	192	75	45	0	5	36	18	64	25	460

Et ainsi,

$$m = \frac{-4}{43} \simeq -0.093,$$
$$\sigma^2 \simeq \frac{460}{43} - (-0.093)^2 \simeq 10.689.$$

Exemple 3.
Soit la série suivante, dont les données sont groupées en classes de valeurs.

classe C_i	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6,8[$	
effectif n_i	5	0	6	2	4	

Pour calculer la moyenne m, la variance σ^2, l'écart type σ et le coefficient de variation C_V, on calcule les centres x_i des classes C_i et on procède de la même façon que dans le cas des données groupées en valeurs, c'est-à-dire comme dans l'exemple précédent. On peut également faire les calculs sur le tableau statstique :

classe C_i	$[-2;0[$	$[0;2[$	$[2;4[$	$[4;6[$	$[6,8[$	total
effectif n_i	5	0	6	2	4	17
centre de classe x_i	-1	1	3	5	7	
$n_i x_i$	-5	0	18	10	28	51
$n_i x_i^2$	5	0	54	50	196	305

Ainsi, on a

$$m = \frac{\sum_{i=1}^{5} n_i x_i}{\sum_{i=1}^{5} n_i} = \frac{51}{17} = 3.$$
$$\sigma^2 = \frac{\sum_{i=1}^{5} n_i x_i^2}{\sum_{i=1}^{5} n_i} - m^2 = \frac{305}{17} - (3)^2 \simeq 8.94.$$

Chapitre 4

Séries statistiques Doubles

4.1 Vocabulaire

4.1.1 Definition

une série statistique double intervient quand on considère deux variables statistiques X et Y en même temps dans le but de chercher une relation les liant. Les éléments d'une série statistique double sont des binômes (x, y), où x et y sont des valeurs de X et Y respectivement. Une série double est quantitative si les deux variables statistiques X et Y sont quantitatives, elles est qualitative si X et Y sont qualitatives et elle est mixte si l'une des deux variables est quantitative et l'autre et qualitative.

Pratiquement, on considère un echantillon de N individus et on mesure simultanément la valeur x du caractère X et la valeur y du caractère Y sur chaque individu de l'échantillon, on obtient ainsi N couples (x_1, y_1), (x_2, y_2), ..., (x_N, y_N). On note qu'un couple peut être obtenu plus qu'une fois.

4.1.2 Notation

On note $x_1, x_2, ..., x_k$ les k valeurs ou modalités du caractère X et y_1, y_2, ... , y_l les l valeurs ou modalités du caractère Y.

On note n_{ij} l'effectif du couple (x_i, y_j).

4.1.3 Exemple

Dans le but d'étudier la relation entre la taille et le poids d'un groupe d'adolescents, on a pris un échantillon de 18 adolescents et on a mésuré la taille et le poids de chacun. On a obtenu la série quantitative suivante, où la première composante est la taille et la deuxième est le poids.

$\{(155; 46,1)$;(140 ; 38,2) ; (161 ; 44,3) ; $(148; 38,2)$; $(155; 50,5)$; $(123; 22,4)$; $(160; 40,4)$;

$(140; 34,7)$; $(165; 50,5)$; $(172; 50,5)$; $(155; 38,1)$; $(160; 57,3)$; $(142; 39,3)$; $(157; 46,1)$;

$(142; 37,1)$; $(167; 60)$;$(148; 45,9)$; $(165; 50,5)\}$.

Cette série double est quantitative, puisque ses deux composantes sont quantitatives.

4.1.4 Exemple

Les données concernant le sexe et l'activité professionnelle de 20 personnes sont présentées en couples. La première composante est le sexe avec deux modalités mâle (M) ou femelle (F), la deuxième est l'activité professionnelle avec trois modalités chomeur (C), actif occupé (AO) et inactif (I). Ces données obtenues par questionnaire sont comme suit

{{F ; AO} ; {M ; I} ; {F ; C} ; {F ; C} ; {M ; AO} ; {M ; AO} ; {M ; C} ; {F ; I} ; {F ;I } ; {F ;I } ; {M ; C} ;

{F ; AO} ; {F ; AO} ; {F ; AO} ; {M ; AO} ; {M ; C} ; {M ; AO} ; {F ; I} ; {F , C} ; {M , AO}}.

Cette série double est qualitative, puisque ses deux composantes sont qualitatives.

4.2 Variable indépendante Versus Variable dépendante

4.2.1 Définition.

La notion de variable statistique indépendante et variable statistique dépendante est évoquée lorsqu'on veut étudier l'influence d'une variable (ou caractère) statistique sur une autre variable statistique. Généralement, l'expérimentateur agit sur les variations d'une variable statistique dite indépendante et mesure l'effet de ces variations sur une deuxième variable statistique, dite dépendante. Dans une série statistique double (X, Y), généralement, la première composante X est la variable indépendante et la deuxième composante est la variable dépendante.

4.2.2 Exemple.

Si on compare les hommes et les femmes quant à leur satisfaction au travail dans une usine, la variable indépendante c'est le sexe et la variable dépendante c'est la satisfaction au travail.

La question de recherche statistique est : Quele est l'effet de l'appartenance à un sexe sur la satisfaction au travail ?

Le sexe et la satisfaction au travail sont deux variables qui existent par elles-mêmes. Le rôle de variable indépendante et variable dépendante sera fixé selon la question de recherche statistique.

4.2.3 Remarque.

Une variable dépendante dans une situation peut être une variable indépendante dans une autre situatiation. Par exemple la satisfaction au travail (dépendante) peut être analysée selon le sexe (indépendante). Il serait aussi possible d'analyser l'absentéisme (dependante) selon la satisfaction au travail (indépendante).

4.3 Représentations

On représente une série double de plusieurs façons :

4.3.1 Représentation brute

Les couples (de valeurs) de la série double sont écrites, côte à côte. Notons qu'un couple de valeurs peut apparaitre plusieurs fois.

Dans l'exemple de la taille et du poid, mesurés sur 18 adolescents, on a utilisé une représentation brute.

4.3.2 Représentation dans un tableau à deux lignes

On dessine un tableau à deux lignes, la première ligne est réservée à la première composante noté x_i et la deuxième est réservée à la deuxième composante noté y_i. Généralement on commence par les paires dont les premières composantes sont les plus petites. Ce mode convient pour les séries quantitatives à petit effectif.

Ce mode de représentation donne pour la série double

$\{(155; 46,1)$;$(140 ; 38,2)$; $(161 ; 44,3)$; $(148; 38,2)$; $(155; 50,5)$; $(123; 22,4)$; $(160; 40,4)$;
$(140; 34,7)$; $(165; 50,5)$; $(148; 38,2)\}$, le tableau suivant.

x$_i$	123	140	140	148	148	155	155	160	161	165
y$_i$	22,4	34,7	38,2	38,2	38,2	46,1	50,5	40,4	44,3	50,5

4.3.3 Représentation dans un tableau à trois lignes

S'il y a des couples (x_i, y_i) de la série double se répètant plus qu'une fois, on représente la série par un tableau à trois lignes, en ajoutant une troixièème ligne pour les effectifs.

Par exemple, la série double

$$(1.1) ; (1,1) ; (1,3) ; (3;2) ; (1;3) ; (1,5) ; (3,2) ; (3;2) ; (4;2) ; (1;5) ; (7,8)$$

est représentée par le tableau à trois lignes

x_i	1	1	1	3	4	7
y_i	1	3	5	2	2	8
n_i	2	2	2	3	1	1

4.3.4 Représentation dans un tableau de contingence

Ce mode de représentation est plus pratique.

La série est représentée par un tableau dont, le nombre de lignes est égale au nombre de valeurs (ou de classes de valeurs) du caractère X et le nombre de colonnes est égale au nombre de valeurs (ou de classes de valeurs) du caractère Y. L'effectif de chaque couple est inscrit dans la cellule du tableau, intersection de la colonne contenant la premiere composante et la ligne contenant la deuxième composante.

Exemple 1 la série qualitative de l'exemple 2.1.3 est représentée dans le tableau de contigence suivant.

Sexe$\diagdown$ *Statut*	AO	C	I	Totaux
M	5	3	1	9
F	4	3	4	11
Totaux	9	6	5	20

L'effectif du couple (M ;AO) est 5, puisque ce couple apparait 5 fois dans cette série. Cet effectif est inscrit dans la cellule intersection de la ligne contenant la première composante (ici M) et de La colonne contenant la deuxième composante (ici OA).

Exemple 2 Une étude sur 5761 femmes de la survenue d'accouchement prématuré et de l'exposition à des évènements stressants a donné les résultats suivants.

X : type d'accouchement, variable qualitative à 2 modalités

Y : score sur une échelle allant de 0 à 3, variable quantitative discrète à 4 valeurs.

La représentation de cette série double dans un tableau de contingence a donné le tableau suivant.

$X\diagdown Y$	0	1	2	3	totaux
A terme	4698	413	250	197	5558
Prématuré	165	16	12	10	203
Totaux	4863	429	262	207	5761

4.3.5 Notation et complétition d'un tableau de contingence

On utilise les notations suivantes pour la représentation d'une série double dans un tableau de contingence

$X \diagdown Y$	y_1	...	y_j	...	y_l	Totaux
x_1	n_{11}	...	n_{1j}	...	n_{1l}	$n_{1\bullet}$
...	...					
x_i	n_{i1}	...	n_ij		n_{il}	$n_{i\bullet}$
...	...					
x_k	n_{k1}		n_{kj}		n_{kl}	$n_{k\bullet}$
Totaux	$n_{\bullet 1}$		$n_{\bullet j}$		$n_{\bullet l}$	$n_{\bullet\bullet}$

4.4 Fréquences associées à une série double

4.4.1 Définition

1)- Pour chaque i et j, le nombre n_{ij} est l'effectif du couple (x_i, y_j) , dont la première composante est x_i, se trouvant dans la ième ligne, et la deuxième composante est y_j se trouvant dans la jème colonne. L'effectif n_{ij} s'appelle aussi fréquence absolue du couple (x_i, y_j) , ou également fréquence absolue jointe de x_i et y_j.

Le nombre

$$f_{ij} = \frac{n_{ij}}{N} \tag{4.1}$$

s'appelle fréquence (ou fréquence relative, s'il ya ambiguité) du couple (x_i, y_j) , ou également fréquence jointe de x_i et y_j..

2)- Pour chaque i, le nombre $n_{i\bullet}$ est l'effectif de x_i, c'est-à-dire, le nombre de couples dont la première composante est x_i, ou en d'autres termes, le nombre d'observations de x_i. Leffectif $n_{i\bullet}$ s'appelle aussi fréquence absolue marginale de x_i.

Le nombre

$$f_{i\bullet} = \frac{n_{i\bullet}}{N} \tag{4.2}$$

s'appelle fréquence marginale de x_i.

3)-Pour chaque j, le nombre $n_{\bullet j}$ est l'effectif de y_j, c'est-à-dire, le nombre de couples dont la deuxième composante est y_j, ou en d'autres termes, le nombre d'observation de y_j. Leffectif $n_{\bullet j}$ s'appelle aussi fréquence absolue marginale de y_j.

Le nombre

$$f_{\bullet j} = \frac{n_{\bullet j}}{N} \tag{4.3}$$

s'appelle fréquence marginale de y_j.

4.4.2 Remarques.

1)- Ici on a utilisé les notations suivantes

$$n_{i\bullet} = \sum_{j=1}^{l} n_{ij}, \qquad n_{\bullet j} = \sum_{i=1}^{k} n_{ij}, \quad n_{\bullet\bullet} = \sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij} \tag{4.4}$$

2)-$n_{\bullet\bullet}$ est l'effectif total de la série double, c'est-à-dire

$$n_{\bullet\bullet} = \sum_{n=1}^{k}\sum_{m=1}^{l} n_{nm} = N \tag{4.5}$$

3)-

$$f_{\bullet\bullet} = \sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij} = 1, \quad \sum_{i=1}^{k} f_{i\bullet} = 1, \quad \sum_{j=1}^{l} f_{\bullet j} = 1 \tag{4.6}$$

4.4.3 Tableau de contingence des fréquences.

Si on remplace dans le tableau de contigence les effectifs par les fréquences correspondantes, on obtient le tableau de contingence (des fréquences) :

$X \diagdown Y$	y_1	...	y_j	...	y_l	Totaux
x_1	f_{11}	...	f_{1j}	...	f_{1l}	$f_{1\bullet}$
...	...					
x_i	f_{i1}	...	f_ij		f_{il}	$f_{i\bullet}$
...	...					
x_k	f_{k1}		f_{kj}		f_{kl}	$f_{k\bullet}$
Totaux	$f_{\bullet 1}$		$f_{\bullet j}$		$f_{\bullet l}$	$f_{\bullet\bullet} = 1$

4.4.4 Exemple

Considérons le tableau de contingence

$X \diagdown Y$	0	1	2	3	Totaux
A terme	4698	413	250	197	5558
Prématuré	165	16	12	10	203
Totaux	4863	429	262	207	5761

On a

$$\begin{aligned}
n_{11} &= 4698 \text{ est l'effectif du couple (A terme ;0),} \\
n_{21} &= 165 \text{ est l'effectif du couple (Prématuré ;0),} \\
n_{23} &= 12 \text{ est l'effectif du couple (Prématuré ;2),} \\
n_{1\bullet} &= 5558 \text{ est l'effectif marginale de "A terme",} \\
n_{\bullet 2} &= 429 \text{ est l'effectif marginale de "1"} \\
n_{\bullet\bullet} &= 5761 \text{ est l'effectif total de la série double,} \\
f_{11} &= \frac{4698}{5761} \text{ est la fréquence du couple (A terme ;0),} \\
f_{1\bullet} &= \frac{5558}{5761} \text{ est la fréquence marginale de "A terme",} \\
f_{\bullet 2} &= \frac{429}{5761} \text{ est la fréquence marginale de "1".}
\end{aligned}$$

4.5 Différentes distributions

4.5.1 Distribution jointe des effectifs de X et Y

On appellera distribution jointe des effectifs de X et Y l'ensemble des informations (x_i, y_j, n_{ij}) pour $i = 1, ..., k$ et
$j = 1, ..., \ell$.

Dans l'exemple ci-dessus, la distribution jointe est représentée par le tableau de contingence, elle est représentée aussi par la suite des huit triplets :
(A terme, 0, 4698), (A terme, 1, 413), (A terme, 2, 250), (A terme, 3, 197),
(Prématuré, 0, 165), (Prématuré,1 , 16), (Prématuré, 2,12), (Prématuré, 3,10).

4.5.2 Distributions marginales

Considérons le tableau de contingence

$X \diagdown Y$	y_1	...	y_j	...	y_l	Totaux
x_1	n_{11}	...	n_{1j}	...	n_{1l}	$n_{1\bullet}$
...	...					
x_i	n_{i1}	...	n_{ij}		n_{il}	$n_{i\bullet}$
...	...					
x_k	n_{k1}		n_{kj}		n_{kl}	$n_{k\bullet}$
Totaux	$n_{\bullet 1}$		$n_{\bullet j}$		$n_{\bullet l}$	$n_{\bullet\bullet}$

A)- En marge à droite (totaux en ligne), pour chaque indice i, l'effectif n$_{i\bullet}$

est le nombre total d'observations de la modalité x_i de X, quelle que soit la modalité de Y . C'est-'a-dire

$$n_{i\bullet} = \sum_{j=1}^{l} n_{ij} = \text{ total de la ligne } i. \tag{4.7}$$

Definition. Les k couples $(x_i, n_{i\bullet})$ définissent la distribution marginale des effectifs de la variable X.

Remarque.

$$\sum_{i=1}^{k} n_{i\bullet} = N \tag{4.8}$$

B)- En marge en bas (totaux en colonne), pour chaque indice j, l'effectif $n_{\bullet j}$ est le nombre total d'observations de la modalité y_j de Y quelle que soit la modalité de X. C'est-à-dire

$$n_{\bullet j} = \sum_{i=1}^{k} n_{ij} = \text{ total de la colonne } j.$$

Définition. Les l couples $(y_j, n_{\bullet j})$ définissent la distribution marginale des effectifs de la variable Y.

Remarque.

$$\sum_{j=1}^{l} n_{\bullet j} = N \tag{4.9}$$

4.5.3 Exemple.

Considérons la série double représentée par le tableau de contingence

$X \diagdown Y$	0	1	2	3	Totaux
A terme	4698	413	250	197	5558
Prématuré	165	16	12	10	203
Totaux	4863	429	262	207	5761

Alors,

A)- la distribution marginale des effectifs de la variable X est

X	à terme	prématuré	effectif total
effectifs	5558	203	5761

B)- la distribution marginale des effectifs de la variable Y est

Y	0	1	2	3	Totaux
effectifs	4863	429	262	207	5761

4.5.4 Distributions conditionnelles

Soit une série statistique double dont les données sont représentées par le tableau de contingence, à k lignes et l colonnes, suivant :

$X \diagdown Y$	y_1	...	y_j	...	y_l	Totaux
x_1	n_{11}	...	n_{1j}	...	n_{1l}	$n_{1\bullet}$
...	...					
x_i	n_{i1}	...	n_{ij}		n_{il}	$n_{i\bullet}$
...	...					
x_k	n_{k1}		n_{kj}		n_{kl}	$n_{k\bullet}$
Totaux	$n_{\bullet 1}$		$n_{\bullet j}$		$n_{\bullet l}$	$n_{\bullet\bullet}$

A)- pour chaque $1 \leq j \leq l$, la distribution conditionnelle des fréquences du caractère X sachant $Y = y_j$ est représentée par le tableau à 2 lignes

X	x_1	...	x_i	...	x_k	$total$
$f(X/Y = y_j)$	$\frac{n_{1j}}{n_{\bullet j}}$	...	$\frac{n_{ij}}{n_{\bullet j}}$	...	$\frac{n_{kj}}{n_{\bullet j}}$	1

On utilise la notation

$$f(X = x_i/Y = y_j) = \frac{n_{ij}}{n_{\bullet j}}. \tag{4.10}$$

et on lit la fréquence de $X = x_i$, sachant $Y = y_j$, est égale à $\frac{n_{ij}}{n_{\bullet j}}$.

B)- pour chaque $1 \leq i \leq k$, la ditribution conditionnelle (des fréquences) du caractère Y sachant $X = x_i$ est représentée par le tableau à 2 lignes

Y	y_1	...	y_j	...	y_l	Total
$f(Y/X = x_i)$	$\frac{n_{i1}}{n_{i\bullet}}$	...	$\frac{n_{ij}}{n_{i\bullet}}$	...	$\frac{n_{il}}{n_{i\bullet}}$	1

On note
$$f(Y = y_j \diagup X = x_i) = \frac{n_{ij}}{n_{i\bullet}} \tag{4.11}$$
et on lit, la fréquence de $Y = y_j$, sachant $X = x_i$, est égale à $\frac{n_{ij}}{n_{i\bullet}}$.

Exemple
Soit le tableau de contingence suivant

$X \diagdown Y$	-2	1	4	totaux
a	2	1	0	3
b	0	0	2	2
c	4	5	3	12
totaux	6	6	5	17

A)- Les distributions conditionnelles des fréquences du caractère X sont :
1)- La distribution de X sachant $Y = -2$, représentée par le tableau à deux lignes suivant

X	a	b	c	total
$f(X \diagup Y = -2)$	$\frac{n_{11}}{n_{\bullet 1}} = \frac{2}{6}$	$\frac{n_{21}}{n_{\bullet 1}} = \frac{0}{6}$	$\frac{n_{31}}{n_{\bullet 1}} = \frac{4}{6}$	1

La distribution de X sachant $Y = 1$, représentée par le tableau à deux lignes suivant

X	a	b	c	total
$f(X \diagup Y = 1)$	$\frac{n_{12}}{n_{\bullet 2}} = \frac{1}{6}$	$\frac{n_{22}}{n_{\bullet 2}} = \frac{0}{6}$	$\frac{n_{32}}{n_{\bullet 2}} = \frac{5}{6}$	1

La distribution de X sachant $Y = 4$, représentée par le tableau à deux lignes suivant

X	a	b	c	total
$f(X \diagup Y = 4)$	$\frac{n_{13}}{n_{\bullet 3}} = \frac{0}{5}$	$\frac{n_{23}}{n_{\bullet 3}} = \frac{2}{5}$	$\frac{n_{33}}{n_{\bullet 3}} = \frac{3}{5}$	1

B)- Les distributions conditionnelles des fréquences du caractère Y sont :
1)- La distribution de Y sachant $X = a$, représentée par le tableau à deux lignes suivant

X	a	b	c	total
$f(Y \diagup X = a)$	$\frac{n_{11}}{n_{1\bullet}} = \frac{2}{3}$	$\frac{n_{12}}{n_{1\bullet}} = \frac{1}{3}$	$\frac{n_{13}}{n_{1\bullet}} = \frac{3}{3}$	1

La distribution de Y sachant $X = b$, représentée par le tableau à deux lignes suivant

X	a	b	c	total
$f(Y/X = b)$	$\frac{n}{n_{2\bullet}} = \frac{0}{2}$	$\frac{n_{22}}{n_{2\bullet}} = \frac{0}{2}$	$\frac{n_{23}}{n_{2\bullet}} = \frac{2}{2}$	1

La distribution de Y sachant $X = c$, représentée par le tableau à deux lignes suivant

X	a	b	c	total
$f(X/Y = 4)$	$\frac{n_{31}}{n_{\bullet 3}} = \frac{4}{12}$	$\frac{n_{32}}{n_{\bullet 3}} = \frac{5}{12}$	$\frac{n_{33}}{n_{\bullet 3}} = \frac{3}{12}$	1

Remarque.

Les distributions conditionnelles des fréquences du caractère X peuvent être représentées dans un même tableau comme suit :

X	a	b	c	total
$f(X/Y = -2)$	$\frac{n_{11}}{n_{\bullet 1}} = \frac{2}{6}$	$\frac{n_{21}}{n_{\bullet 1}} = \frac{0}{6}$	$\frac{n_{31}}{n_{\bullet 1}} = \frac{4}{6}$	1
$f(X/Y = 1)$	$\frac{n_{12}}{n_{\bullet 2}} = \frac{1}{6}$	$\frac{n_{22}}{n_{\bullet 2}} = \frac{0}{6}$	$\frac{n_{32}}{n_{\bullet 2}} = \frac{5}{6}$	1
$f(X/Y = 4)$	$\frac{n_{13}}{n_{\bullet 3}} = \frac{0}{5}$	$\frac{n_{23}}{n_{\bullet 3}} = \frac{2}{5}$	$\frac{n_{33}}{n_{\bullet 3}} = \frac{3}{5}$	1

Ce tableau s'appelle aussi tableau de la distribution (au lieu de tableau des distributions) de X sachant Y.

De la même façon on construit le tableau de la distribution de Y sachant X.

4.5.5 Indépendance de variables statistiques

Définition.

Les deux variables statistiques X et Y d'une série statistiques doubles, représentée par un tableau de contingence, sont dites indépendantes si

$$n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n_{\bullet\bullet}}, \text{ pour tous } i \text{ et } j. \tag{4.12}$$

Conséquence.

L'indépendance des caratères statistiques X et Y signifie que les distributions conditionnelles des fréquences sont identiques aux distributions marginales des fréquences, c'est-à-dire

$$\frac{n_{ij}}{n_{i\bullet}} = \frac{n_{\bullet j}}{n_{\bullet\bullet}} \text{ et } \frac{n_{ij}}{n_{\bullet j}} = \frac{n_{i\bullet}}{n_{\bullet\bullet}}, \text{ pour tous } i \text{ et } j. \tag{4.13}$$

Exemple 1.

Soit une série double représentée par le tableau de contingence suivant

$X\diagdown Y$	y_1	y_2	y_3	totaux
x_1	2	5	6	13
x_2	6	15	18	39
Totaux	8	20	24	52

Les deux caractères X et Y sont indépendants, car la condition

$$n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{n_{\bullet\bullet}},$$

est vérifiée pour tous $i = 1, 2$ et $j = 1, 2, 3$. En effet,

$i = 1$ et $j = 1$ donne

$$2 = \frac{13 \times 8}{52},$$

$i = 1$ et $j = 2$ donne

$$5 = \frac{13 \times 20}{52},$$

$i = 1$ et $j = 3$ donne

$$6 = \frac{13 \times 24}{52},$$

$i = 2$ et $j = 1$ donne

$$6 = \frac{39 \times 8}{52},$$

$i = 2$ et $j = 2$ donne

$$15 = \frac{39 \times 20}{52},$$

$i = 2$ et $j = 3$ donne

$$18 = \frac{39 \times 24}{52}.$$

Exemple 2.

Soit une série double représentée par le tableau de contingence suivant

$X \diagdown Y$	y_1	y_1	totaux
x_1	3	4	7
x_2	6	8	14
x_3	9	10	19
totaux	18	22	40

Les caractères X et Y ne sont pas indépendants. Pour le prouver, il suffit de montrer qu'il existe au moins un indice i et un indice j, tels que

$$n_{ij} \neq \frac{n_{i\bullet} \times n_{\bullet j}}{n_{\bullet\bullet}}.$$

C'est le cas, car, par exemple pour $i = 1$ et $j = 1$, on a

$$n_{11} = 3 \text{ et } \frac{n_{1\bullet} \times n_{\bullet 1}}{n_{\bullet\bullet}} = \frac{7 \times 18}{40} = 3.15.$$

4.6 Séries doubles quantitatives

4.6.1 Passage d'une représentation à une autre

Exemple 1.

Soit une série quantitative double dont les données sont représentées par le tableau à deux lignes suivant :

x_i	2	4	5	5	6	6	7	7	10	10	10	
y_i	3	3	6	6	2	5	1	7	8	8	8	

Ces données donnent le tableau à trois lignes suivant :

x_i	2	4	5	6	6	7	7	10
y_i	3	3	6	2	5	1	7	8
n_i	1	1	2	1	1	1	1	3

Ces mêmes données donnent le tableau de contingence suivant :

$X \diagdown Y$	1	2	3	5	6	7	8
2	0	0	1	0	0	0	0
4	0	0	1	0	0	0	0
5	0	0	0	0	2	0	0
6	0	1	0	1	0	0	0
7	1	0	0	0	0	1	0
10	0	0	0	0	0	0	3

Exemple 2.
Le tableau de contingence

$X \diagdown Y$	4	8	10
-1	2	0	3
0	1	2	1

Equivaut au tableau à deux lignes suivant

x_i	-1	-1	-1	-1	-1	0	0	0	0
y_i	4	4	10	10	10	4	8	8	10

Ce même tableau équivaut au tableau à trois lignes

x_i	-1	-1	0	0	0
y_i	4	10	4	8	10
n_i	2	3	1	2	1

Remarque.

1)- Si les valeurs des caractères quantitatives X et Y sont données individuellement, alors $x_1, x_2, ...x_k$ et $y_1, y_2, ...y_l$ désigneront les valeurs de X et Y respectivement. Ces valeurs sont représentées dans le tableau de contingence dans l'ordre croissant, c'est-à-dire du plus petit au plus grand.

2)- Si les valeurs des caractères quantitatives X et Y sont données en classes, alors $x_1, x_2, ...x_k$ et $y_1, y_2, ...y_l$ désigneront les centres de classes des valeurs de X et Y respectivement.

4.6.2 Exemple

Une entreprise employant 100 femmes relève pour chaque femme son âge, noté X, et le nombre de journées d'absence durant le mois de janvier, noté Y. Le résultat de cette opération est représenté dans le tableau de contingence suivant.

$X \diagdown Y$	0	1	2	3
$[20; 30[$	0	0	5	15
$[30; 40[$	0	15	20	0
$[40; 50[$	15	10	5	0
$[50; 60[$	0	5	5	5

Pour faire des calculs sur ce tableau, on remplace les classes par leurs centres et on le complète par les totaux marginaux. On obtient alors le tableau suivant.

$X \diagdown Y$	0	1	2	3	Totaux
25	0	0	5	15	20
35	0	15	20	0	35
45	15	10	5	0	30
55	0	5	5	5	15
Totaux	15	30	35	20	100

4.6.3 Moyennes des distributions marginales :

A)- Moyenne de X, noté $\mu(X)$ ou $\bar{X}$:

$$\mu(X) = \frac{1}{N}\sum_{i=1}^{k} n_{i\bullet}x_i = \sum_{i=1}^{k} p_{i\bullet}x_i \tag{4.14}$$

B)- Moyenne de Y, noté $\mu(Y)$ ou $\bar{Y}$:

$$\mu(Y) = \frac{1}{N}\sum_{j=1}^{l} n_{\bullet j}y_j = \sum_{j=1}^{l} p_{\bullet j}y_j \tag{4.15}$$

L'application de cette definition à notre exemple donne

$$\mu(X) = \frac{1}{100}(20 \times 25 + 35 \times 35 + 30 \times 45 + 15 \times 55) = 39$$
$$\mu(Y) = \frac{1}{100}(15 \times\ 0 + 30 \times\ 1 + 35 \times\ 2 + 20 \times\ 3) = 1.6$$

4.6.4 Variances des distributions marginales :

A)-Variance (notée V_X ou σ_X^2) et écart-type (noté σ_X) de X :

$$V_X = \frac{1}{N}\sum_{i=1}^{k} n_{i\bullet}x_i^2 - \mu(X)^2 \tag{4.16}$$

$$\sigma_X = \sqrt{V_X}$$

B)-Variance (notée V_Y ou σ_Y^2) et écart-type (noté σ_Y) de Y :

$$V_Y = \frac{1}{N}\sum_{j=1}^{l} n_{j\bullet}y_j^2 - \mu(Y)^2 \tag{4.17}$$

$$\sigma_Y = \sqrt{V_Y} \tag{4.18}$$

L'application de cette définition à notre exemple donne

$$V_X = \frac{1}{100}(20 \times 25^2 + 35 \times 35^2 + 30 \times 45^2 + 15 \times 55^2) - 39^2 = 94$$

$$\sigma_X = \sqrt{94} = 9,70$$

$$V_Y = \frac{1}{100}(15 \times\ 0^2 + 30 \times\ 1^2 + 35 \times\ 2^2 + 20 \times\ 3^2) - 1,6^2 = 0,94$$

$$\sigma_Y = \sqrt{0,94} = 0,97$$

Covariance - Corrélation

Definition.

La covariance d'une série quantitative double $(x_i, y_j)\,, 1 \leq i \leq l, 1 \leq j \leq k$, qu'on note ici $Cov\,(X, Y)$ ou V_{XY} , est le nombre réel, défini par

$$Cov\,(X, Y) = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}\,(x_i - \overline{x})\,(y_j - \overline{y})\,. \tag{4.19}$$

Propriétés.

On a

$$Cov\,(X, Y) = \frac{1}{N}\sum_{i=1}^{l}\sum_{j=1}^{k} n_{ij}x_iy_j - \overline{x}\ \overline{y} \tag{4.20}$$

$$Cov\,(X, Y) = \sum_{i=1}^{l}\sum_{j=1}^{k} f_{ij}x_iy_j - \overline{x}\ \overline{y}, \tag{4.21}$$

où f_{ij} est la fréquence du couple $(x_i, y_j)\,.$

$$Cov(X, Y) = Cov(Y, X) \ \text{ et } \ cov(X, X) = V_X.$$

L'application de la définition de la covariance à l'exemple précédent donne

$$\begin{aligned} Cov\,(X, Y) &= \frac{1}{N}\sum_{i=1}^{l}\sum_{j=1}^{k} n_{ij}x_iy_j - \overline{x}\ \overline{y} \\ &= \frac{1}{100}\left[(0 \times 25 \times 0) + (0 \times 25 \times 1) + ...\,(5 \times 55 \times 2) + (5 \times 55 \times 3)\right] - (39 \times 1.6) \\ &= \frac{5850}{100} - 39 \times 1.6 = -3.9. \end{aligned}$$

Dans le calcul de $Cov\,(X, Y)$ ci-dessus, on a commencé par la première ligne jusqu'à la quatrième ligne

Definition.

Le coefficient de corrélation, qu'on note ici $Cor\,(X, Y)$ ou r, est le nombre réel, défini par

$$Cor\,(X, Y) = \frac{Cov\,(X, Y)}{\sigma_X \sigma_Y}. \tag{4.22}$$

L'application de cette définition à l'exemple précédent donne

$$\begin{aligned} Cor\,(X, Y) &= \frac{-3.9}{9.70 \times 0.97} \\ &= 0.41. \end{aligned}$$

Propriétés.

1)-

$$-1 \leq Cor\,(X, Y) \leq 1. \tag{4.23}$$

2)-

$$Cor\,(X, Y) = Cor\,(Y, X)\,. \tag{4.24}$$

3)-

$$Cor\,(X, X) = 1. \tag{4.25}$$

Utilité du coefficient de corrélation

Le coefficient de corrélation mesure l'existance d'une liaison linéaire entre X et Y et le degré de cette liaison:

1.
$$Cor(X, Y) = 1 : \text{liaison linéaire exacte, } Y = aX + b, \text{ avec } a > 0.$$

2.
$$Cor(X, Y) = -1 : \text{ liaison linéaire exacte, } Y = aX + b, \text{ avec } a < 0.$$

3.
$$Cor(X, Y) = 0 : \text{ pas de corrélation.}$$

4.
$$Cor(X, Y) > 0 : \text{ } X \text{ et } Y \text{ ont tendance à varier dans le même sens.}$$

5.
$$Cor(X, Y) < 0 : \text{ } X \text{ et } Y \text{ ont tendance à varier dans le sens contraire.}$$

6.
$$|Cor(X, Y)| > 0,8 : \text{ } X \text{ et } Y \text{ ont une forte corrélation (liaison) linéaire.}$$

4.6.5 Ajustement linéaire d'une série quantitative double

L'ajustement linéaire signifie la recherche d'une droite qui représente le mieux, une liaison linéaire entre X et Y.

Ajustement graphique.

L'ajustement graphique d'une série double (x_i, y_i) se fait par la visualisation du nuage de points (x_i, y_i) de la série et le traçage d'une droite, appelée droite d'ajustement graphique, de sorte que les points (x_i, y_i) du nuage soient le plus proches possible de cette droite. L'ajustement graphique sert à donner rapidement une première idée sur la relation entre l'évolution du caractère X et l'évolution du caractère Y.

L'avantage de l'ajustement graphique réside dans sa simplicité et son utilisation immédiate. L'inconvénient vient de sa subjectivité : il n'est pas unique car il dépend de la personne qui l'a réalisée.

Nuage de points.

Pour représenter le nuage de points, on dessine un repère orthogonal. Le nuage de points associé à la série double est l'ensemble des points (x_i, y_i) représentés dans ce repère.

Ajustement par la méthode des moindres carrés.

Principe.

Soit une série double représentée par un tableau de contingence.

A)- On montre qu'il existe une droite unique D_y d'équation

$$y = ax + b$$

qui minimise la somme

$$\sum_{j=1}^{l} \sum_{i=1}^{k} n_{ij} \left(y_j - (ax_i + b)\right)^2 .$$

Les coefficients a et b se calculent par les formules

$$a = \frac{cov\,(X, Y)}{\sigma_X^2} \text{ et } b = \overline{y} - a\overline{x}. \tag{4.26}$$

La droite D_y s'appelle droite de regression de y en x. Cette droite est utilisée pour prévoir les valeurs de Y pour des valeurs de X, non trouvées expérimentalement,...

B)- On montre qu'il existe une droite unique D_x d'équation

$$x = a'y + b'$$

qui minimise la somme

$$\sum_{j=1}^{l} \sum_{i=1}^{k} n_{ij} \left(x_i - \left(a'y_j + b' \right) \right)^2 .$$

Les coefficient a et b se calculent par les formules

$$a' = \frac{cov\,(X,Y)}{\sigma_y^2} \text{ et } b' = \overline{x} - a'\overline{y}. \tag{4.27}$$

La droite D_x s'appelle droite de regression de x en y. Cette droite est utilisée pour prévoir les valeurs de X pour des valeurs de Y, non données expérimentalement,...

Autres formules pour a, a′et Cor (X,Y) .

A partir des formules précédentes, on obtient les formules suivantes.

$$\begin{aligned}
a &= \frac{\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j - \overline{x}\sum_{j=1}^{l} n_{\bullet j}y_j}{\sum_{i=1}^{k} n_{i\bullet}x_i^2 - \overline{x}\sum_{i=1}^{k} n_{i\bullet}x_i} \\
&\frac{\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j - N\overline{x}\;\overline{y}}{\sum_{i=1}^{k} n_{i\bullet}x_i^2 - N\overline{x}^2} \\
a' &= \frac{\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j - \overline{x}\sum_{j=1}^{l} n_{\bullet j}y_j}{\sum_{j=1}^{l} n_{\bullet j}y_j^2 - \overline{y}\sum_{j=1}^{l} n_{\bullet j}y_j} \\
&= \frac{\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j - N\overline{x}\;\overline{y}}{\sum_{j=1}^{l} n_{\bullet j}y_j^2 - N\overline{y}^2} \\
Cor\,(X,Y) &= \frac{\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j - \overline{x}\sum_{j=1}^{l} n_{\bullet j}y_j}{\sqrt{\left(\sum_{i=1}^{k} n_{i\bullet}x_i^2 - \overline{x}\sum_{i=1}^{k} n_{i\bullet}x_i\right)\left(\sum_{j=1}^{l} n_{\bullet j}y_j^2 - \overline{y}\sum_{j=1}^{l} n_{\bullet j}y_j\right)}}
\end{aligned} \tag{4.28}$$

Le coefficient de corrélation $\mathbf{Cor}\,(X,Y)$, $\mathbf{a}$ et $\mathbf{a}'$ ont même signe et on a

$$\mathbf{Cor}\,(X,Y)^2 = \mathbf{aa}'. \tag{4.29}$$

Exemple.

On considère la série double représentée par le tableau de contingence suivant

$X \diagdown Y$	2	3	Totaux
25	5	15	20
35	20	5	25
45	5	0	5
Totaux	30	20	50

On a

$$\begin{aligned}
\overline{x} &= \frac{\sum_{i=1}^{3} n_{i\bullet} x_i}{N} \\
&= \frac{20 \times 25 + 25 \times 35 + 5 \times 45}{50} \\
&\frac{1600}{50} \\
&= 32.
\end{aligned}$$

$$\begin{aligned}
\overline{y} &= \frac{\sum_{j=1}^{2} n_{\bullet j} x_j}{N} \\
&= \frac{30 \times 2 + 20 \times 3}{50} \\
&\frac{120}{50} \\
&= 2.4
\end{aligned}$$

$$\begin{aligned}
&\sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} x_i y_j \\
&= (5 \times 25 \times 2) + (15 \times 25 \times 3) + (20 \times 35 \times 2) \\
&\quad + (5 \times 35 \times 3) + (5 \times 45 \times 2) + (0 \times 45 \times 3) \\
&= 3750
\end{aligned}$$

$$\sum_{i=1}^{k} n_{i\bullet} x_i^2 = \left(20 \times 25^2\right) + \left(25 \times 35^2\right) + \left(5 \times 45^2\right)$$
$$= 53250.$$

$$\sum_{j=1}^{2} n_{\bullet j} y_j^2 = 30 \times 2^2 + 20 \times 3^2$$
$$= 300.$$

$$a = \frac{3750 - 50 \times 32 \times 2.4}{53250 - 50 \times 32^2}$$
$$= -0.044.$$

$$a' = \frac{3750 - 50 \times 32 \times 2.4}{300 - 50 \times 2.4^2}$$
$$= -7.5.$$

$$Cor\left(X, Y\right) = \frac{3750 - 50 \times 32 \times 2.4}{\sqrt{\left(53250 - 50 \times 32^2\right)\left(300 - 50 \times 2.4^2\right)}}$$
$$= -0.57.$$

$$b = \overline{y} - a\overline{x}$$
$$= 2.4 - (-0.044) \times 32$$
$$= 3.808.$$

$$b' = \overline{x} - a'\overline{y}$$
$$= 32 - (-7.5) \times 2.4$$
$$= 50.$$

L'équation de la droite de regression de y en x est

$$y = -0.044x + 3.808.$$

L'équation de la droite de regression de x en y est

$$x = -7.5y + 50.$$

Puisque la valeur absolue $(= 0.57)$ du coefficient de corrélation n'est pas proche de un, la corrélation linéaire entre x et y n'est pas bonne.

Cas de représentation par un tableau à deux lignes

Pour une série double représentée par un tableau à deux lignes.

x_1	x_2	...	x_N
y_1	y_2	...	y_N

On a les formules simples suivantes

$$\begin{aligned}\overline{x} &= \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (4.30)\\ \overline{y} &= \frac{1}{N}\sum_{i=1}^{N} y_i\\ V(X) &= \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \overline{x}^2\\ V(Y) &= \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \overline{y}^2\\ Cov(X,Y) &= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \overline{x}\,\overline{y}\end{aligned}$$

A partir de ces formules on peut également vérifier les formules suivantes.

$$\begin{aligned}a &= \frac{\sum_{i=1}^{N} x_i y_i - \overline{x}\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i^2 - \overline{x}\sum_{i=1}^{N} x_i} \qquad (4.31)\\ a' &= \frac{\sum_{i=1}^{N} x_i y_i - \overline{x}\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} y_i^2 - \overline{y}\sum_{i=1}^{N} y_i}\\ Cor(X,Y) &= \frac{\sum_{i=1}^{N} x_i y_i - \overline{x}\sum_{i=1}^{N} y_i}{\sqrt{\left(\sum_{i=1}^{N} x_i^2 - \overline{x}\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i^2 - \overline{y}\sum_{i=1}^{N} y_i\right)}}\end{aligned}$$

Exemple.
Le tableau ci-dessous liste les classements de salaires et de stress pour des emplois sélectionnés aléatoirement. Le rang 1 pour les salaires correspond au salaire le plus bas et le rang 1 pour le stress correspond au stress le plus faible.

Emploi	Rang du salaire	Rang du stress
agent de change	9	9
zoologiste	5	4
ingénieur en électricité	8	5
CPE	6	7
gérant d'hôtel	4	6
employé de banque	1	3
inspecteur de sécurité	2	2
économiste	3	1
psycologue	7	8
pilote de l'air	10	11
trader à wall street	11	10

1)- Calculez les moyennes marginales $\overline{x}$, $\overline{y}$, de X et Y.

2)- Calculer les sommes des carrés x_i^2, les sommes des carrés y_i^2 et les sommes des produits $x_i y_i$

3)- Calculez le coefficient de corrrélation linéaire $Cor\,(X,Y)$ et estimez la qualité de la corrélation linéaire entre le rang X du salaire et le rang Y du stress.

4)- Calculez les coefficients a et b de la droite de regression de Y en X : $y = ax + b$.

On peut faire les calculs sur un tableau statistique :

	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
	9	9	81	81	81
	5	4	20	25	16
	8	5	40	64	25
	6	7	42	36	49
	4	6	24	16	36
	1	3	3	1	9
	2	2	4	4	4
	3	1	3	9	1
	7	8	56	49	64
	10	11	110	100	121
	11	10	110	121	100
Total	66	66	493	506	506

1)-

$$\overline{x} = \frac{\sum_{i=}^{N} x_i}{N} = \frac{\sum_{i=}^{11} x_i}{11} = \frac{66}{11} = 6$$

$$\overline{y} = \frac{\sum_{i=}^{N} y_i}{N} = \frac{\sum_{i=}^{11} y_i}{11} = \frac{66}{11} = 6$$

2)-

$$\sum_{i=}^{N} x_i^2 = 506, \;\; \sum_{i=1}^{11} y_i^2 = 506, \;\; \sum_{i=}^{11} x_i y_i = 493.$$

3)-

$$
\begin{aligned}
Cor\left(X,Y\right) &= \frac{\sum_{i=1}^{11} x_i y_i - \overline{x}\sum_{i=1}^{11} y_i}{\sqrt{\left(\sum_{i=1}^{11} x_i^2 - \overline{x}\sum_{i=1}^{11} x_i\right)\left(\sum_{i=1}^{11} y_i^2 - \overline{y}\sum_{i=1}^{11} y_i\right)}} \\
&= \frac{493-(6\times 66)}{\sqrt{[506-(6\times 66)]\,[506-(6\times 66)]}} \\
& \frac{97}{\sqrt{12100}} \\
&= 0.88
\end{aligned}
$$

Il y a une forte corrélation linéaire entre les deux caractères X et Y.

4)-

$$
\begin{aligned}
a &= \frac{\sum_{i=1}^{11} x_i y_i - \overline{x}\sum_{i=1}^{11} y_i}{\sum_{i=1}^{11} x_i^2 - \overline{x}\sum_{i=1}^{11} x_i} \\
&= \frac{97}{[506-(6\times 66)]} \\
&= 0.88
\end{aligned}
$$

$$
\begin{aligned}
b &= \overline{y} - a\overline{x} \\
&= 6-(0.88\times 6) \\
&= 0.72.
\end{aligned}
$$

Cas de représentation par un tableau à trois lignes

Pour une série double représentée par un tableau à trois lignes.

x_1	x_2	...	x_r
y_1	y_2	...	y_r
n_1	n_2	...	n_r

On a les formules simples suivants

$$\begin{aligned}
\overline{x} &= \frac{1}{N}\sum_{i=1}^{r} n_i x_i \qquad (4.32)\\
\overline{y} &= \frac{1}{N}\sum_{i=1}^{r} n_i y_i\\
V(X) &= \frac{1}{N}\sum_{i=1}^{r} n_i x_i^2 - \overline{x}^2\\
V(Y) &= \frac{1}{N}\sum_{i=1}^{r} n_i y_i^2 - \overline{y}^2\\
Cov(X,Y) &= \frac{1}{N}\sum_{i=1}^{r} n_i x_i y_i - \overline{x}\ \overline{y}
\end{aligned}$$

Il faut noter que dans ce cas, l'entier r est le nombre des couples (x_i, y_i), n_i est le nombre de répétition du couple (x_i, y_i) et l'effectif total N est égal à la somme des n_i.

Pour cette représentation, on peut également vérifier les formules suivantes.

$$\begin{aligned}
a &= \frac{\sum_{i=1}^{r} n_i x_i y_i - \overline{x}\sum_{i=1}^{r} n_i y_i}{\sum_{i=1}^{N} n_i x_i^2 - \overline{x}\sum_{i=1}^{N} n_i x_i} \qquad (4.33)\\
a' &= \frac{\sum_{i=1}^{r} x_i y_i - \overline{x}\sum_{i=1}^{r} n_i y_i}{\sum_{i=1}^{r} n_i y_i^2 - \overline{y}\sum_{i=1}^{r} n_i y_i}\\
Cor(X,Y) &= \frac{\sum_{i=1}^{r} n_i x_i y_i - \overline{x}\sum_{i=1}^{r} n_i y_i}{\sqrt{\left(\sum_{i=1}^{r} n_i x_i^2 - \overline{x}\sum_{i=1}^{r} n_i x_i\right)\left(\sum_{i=1}^{r} n_i y_i^2 - \overline{y}\sum_{i=1}^{r} n_i y_i\right)}}
\end{aligned}$$

Exemple.

on considère la série double suivante

$$\begin{aligned}
&\{(-1;0)\,, \quad (1;0)\,, \quad (2;1)\,, \quad (-1;0)\,, \quad (2;-1)\,,\\
&(2;5) \quad (3;2)\,, \quad (-1;0)\,, \quad (2;5)\,, \quad (2;-1)\,,\\
&(-1;0)\,, \quad (-1;0)\,, \quad (3;2)\,, \quad (-1;3)\,, \quad (-1;0)\\
&(2;5)\,, \quad (2;-1)\,, \quad (2;-1)\,, \quad (-1;3)\,, \quad (0;0)\}.
\end{aligned}$$

1)-Représenter les données de cette série double dans un tableau statistique.

2)- Calculer les paramètres suivants : $\overline{x}, \overline{y}, V_X, V_Y, V_{XY}, Cor(X,Y)$.

Solution.

1)- On représente cette série double dans le tableau statistique (à trois lignes) suivant :

x_i	-1	-1	0	1	2	2	2	3
y_i	0	3	0	0	-1	1	5	2
n$_i$	6	2	1	1	4	1	3	2

2)- Pour faire des calculs sur le tableau, on ajoute des lignes :

									Total
x_i	-1	-1	0	1	2	2	2	3	
y_i	0	3	0	0	-1	1	5	2	
n_i	6	2	1	1	4	1	3	2	20
n_ix_i	-6	-2	0	1	8	2	6	6	15
n_iy_i	0	6	0	0	-4	1	15	4	22
$n_ix_iy_i$	0	-6	0	0	-8	2	30	12	30
$n_ix_i^2$	6	2	0	1	16	4	12	18	59
$n_iy_i^2$	0	18	0	0	4	1	75	8	106

D'aprés le tableau, on a

$$\begin{aligned}
\overline{x} &= \frac{\sum_{i=1}^{8} n_i x_i}{\sum_{i=1}^{8} n_i} = \frac{15}{20} = \mathbf{0.75}, \\
\overline{y} &= \frac{\sum_{i=1}^{8} n_i y_i}{\sum_{i=1}^{8} n_i} = \frac{22}{20} = \mathbf{1.1}, \\
V_X &= \frac{\sum_{i=1}^{8} n_i x_i^2}{\sum_{i=1}^{8} n_i} - \overline{x}^2 = \frac{59}{20} - 0.75^2 = \mathbf{2.3875}, \\
V_Y &= \frac{\sum_{i=1}^{8} n_i y_i^2}{\sum_{i=1}^{8} n_i} - \overline{y}^2 = \frac{106}{20} - 1.1^2 = 4.\mathbf{09}, \\
V_{XY} &= \frac{\sum_{i=1}^{8} n_i x_i y_i}{\sum_{i=1}^{8} n_i} - \overline{xy} = \frac{30}{20} - 0.75 \times 1.1 = \mathbf{0.675}, \\
Cor\,(X,Y) &= \frac{V_{XY}}{V_X V_Y} = \frac{0.675}{2.3875 \times 6.19} = \mathbf{0.216}.
\end{aligned}$$

Remarque importante

Dans les exemples précédents, on a utilisé des applications disponibles dans les calculatrices ordinaires (KENKO,...). Ces applications donnent les valeurs des paramètres directement. Si on fait les calculs à la main, les résultats obtenus peuvent être légèrement différents, à cause des arrondis.

Troisième partie

Analyse combinatoire et probabilités

Chapitre 5

Analyse combinatoire

le terme "Analyse combinatoire", désigne par convention, les techniques de dénombrer des ensembles finis, c'est-à-dire les techniques pour calculer le nombre des éléments de ces ensembles. En probabilités les ensembles à dénombrer sont souvent déterminés par une certaine expérience aléatoire (c'est-à-dire dont le résultat ne peut pas être déterminé d'avance).

5.1 Arrangements et permutatations

5.1.1 Arrangements.

Arrangement (sans répétition).

Définition On appelle arrangement de r éléments d'un ensemble E, à n éléments ($r \leq n$) l'opération suivante :

choisir dans l'ordre r éléments parmi les n éléments de l'ensemble E. Cela est possible parce que $r \leq n$.

Question.

Combien y a-t-il d'arrangements de r éléments d'un ensemble à n éléments ?

Réponse.

On procède comme suit.

1- On choisit un premier élément de l'ensemble E : on a n façons de le faire.

2- Il reste $n-1$ éléments, on en choisit un deuxième élément : on a $n-1$ façons de le faire.

.
.
.

r- Il reste $(n-r+1)$ éléments, on en choisit un $r^{ème}$ élément : on a $(n-r+1)$ de le faire.

En résumé on a réalisé r expériences, la première a r résultats possibles, la deuxième a $r-1$ résultats possibles,..., la $r^{ème}$ a $(n-r+1)$ résultats possibles.

Ainsi, le nombre d'arrangements de r éléments d'un ensemble à n éléments $(r \leq n)$, noté A_n^r est donné par la formule

$$A_n^r{=}n \times (n-1) \times ... \times (n-r+1) = \frac{n!}{(n-r)!} \tag{5.1}$$

Exemple. 15 chevaux ont participé à une course. Quels est le nombre de quartés dans l'ordre possibles ?

Un quarté dans l'ordre est l'ensemble des 4 numéros ayant gagné la course, placés dans l'ordre selon le résultat de la course.

La première place sera gagnée par 1 cheval parmi les 15 participants, il y a 15 résultats possibles.

Pour chaque résultat de la première place, la deuxième place sera gagnée par 1 cheval parmi les 14 chevaux restants, il y a 14 résultats possibles.

Pour chaque résultat des deux premières places, la troisième place sera gagnée par un cheval parmi les 13 qui restent, il y a 13 résultats possibles.

Enfin, pour chaque resultats des trois premieres places, la quatrième et dernière place sera gagnée par un cheval parmi les 12 restants.

Ainsi, le nombre total des quartés dans l'ordre, dans une course de 15 chevaux est égal à

$$A_{15}^4 = 15 \times 14 \times 13 \times 12 = \frac{15!}{(15-4)!} = 32760.$$

5.1.2 Arrangement avec répétition.

Définition. Un arrangement avec répétition de r éléments d'un ensemble $\grave{a}$ n éléments est l'opération de choisir dans l'ordre, r éléments parmi les n éléments de l'ensemble E, où un élément peut être choisi plus d'une fois. Ici

r peut être plus grand que n. parce que un même élément peut être choisi plusieurs fois, et c'est là la différence avec l'arrangement (sans répétition).

Question.

Quel est le nombre d'arrangements avec répétition, de r éléments d'un ensemble à n éléments.

Réponse.

On choisit un élément parmi les n éléments pour la première place, on a n façons de le faire.

On choisit ensuite un élément pour la deuxième place, toujours parmi les n éléments, parce que l'élément choisi pour la première place peut aussi être choisi pour la deuxième place, on a aussi n façons de le faire.

On répète cette opération r fois et chaque fois on aura n façons possibles de le faire. Ainsi, le nombre d'arrangements avec répétition de r éléments d'un ensemble à n éléments, noté R_n^r, est donné par la formule

$$R_n^r = n^r \tag{5.2}$$

Exemple. Un groupe de 10 élèves sont en compétition pour avoir la meilleure note dans chacune des quatre matières : mathématiques, physique, arabe, français.

on élimine les cas où deux élèves ou plus auront la même note dans la même matière.

Ici un élève peut avoir la meilleure note en une, en deux, en trois, ou dans les quatre matières.

Quel est le nombre de situations possibles ?

La meilleure note en mathématiques sera acquise par l'un des 10 étudiants, on a 10 possibilités.

La meilleure note en physique sera acquise aussi par l'un des mêmes 10 étudiants, y compris celui qui aura la meilleure note en mathématiques, on a donc aussi 10 possibilités.

Le même nombre de possibilités pour l'arabe et aussi le français.

Ainsi le nombre total de résultats possible est égal à

$$R_{10}^4 = 10^4$$

5.1.3 Permutation.

Une permutation d'un ensemble E à n éléments est un arrangement des n éléments de l'ensemble E.

Donc la permutation est un arrangement particulier et ainsi, on calcule le nombre de permutations d'un ensemble ayant n éléments par la même formule. Ainsi, le nombre de permutations d'un ensemble à n éléments est égal à

$$n(n-1) \times ... \times 1 = n! \tag{5.3}$$

Exemple. Dix candidats ont participé à un concours et ils attendent l'annonce du résultat du concours, où les 10 étudiants seront classés de 1 à 10. Quel est le nombre de résultats possibles du concours.

Le nombre de résultats possibles est $10!$

5.2 Combinaisons.

5.2.1 Combinaison (sans répétition)

Définition.

Une combinaison de p éléments d'un ensemble E à n éléments $(p \leq n)$ est un sous ensemble contenant p éléments, de l'ensemble E à n éléments.

On désigne par C_n^p le nombre des combinaisons de p éléments d'un ensemble à n éléments.

Exemple.

Soit $E = \{a, b, c, d, e, f, g\}$.

$A = \{a, d, e\}$ est une combinaison de 3 éléments de l'ensemble E à 7 éléments.

$\{a\}$ est une combinaison de 1 élément de l'ensemble E.

L'ensemble vide $\emptyset$ est une combinaison de zéro élément de l'ensemble E.

E est une combinaison de n éléments de l'ensemble E.

5.2.2 Combinaisons et arrangements.

On montrera ici comment calculer le nombre de combinaisons à partir du nombre d' arrangements, déjà donné par la formule (5.1) . On expliquera çà dans le cas particulier $(p = 3, n = 7)$.

Dans l'exemple précédent les deux représentations $\{a, d, e\}$ et $\{d, a, e\}$ représentent la même combinaison, parce qu' elles sont formées des mêmes éléments. Mais, elles représentent deux arrangements différents, parce que, bien qu'elles soient formées des mêmes éléments, ces éléments ne sont pas dans le même ordre.

Les arrangements formées des mêmes éléments a, d, e de E sont :
$\{a, d, e\}, \{a, e, d\}, \{e, d, a\}, \{e, a, d\}, \{d, a, e\}\ \{d, e, a\}$
Ils sont en nombre de 6.

On a construit ces 6 arrangements en considérant tous les positions possibles des trois éléments a, d, e. C'est-à-dire en considérant toutes les permutations de l'ensemble de ces trois éléments. Ainsi, le nombres d'arrangements formés de ces mêmes trois éléments a, d, e est égal au nombre des permutations d'un ensemble à 3 éléments, c'est-dire $3! = 6$. Donc le nombre A_7^3 des arrangements de 3 éléments d'un ensemble à 7 éléments est donné par la formule $A_7^3 = 3! C_7^3$.

Puisque la valeur de A_7^3 a déjà été calculée, on en déduit le nombre C_7^3 des combinaisons de 3 éléments d'un ensemble à 7 éléments, par la formule suivante.

$$C_7^3 = \frac{A_7^3}{3!} = \frac{7!}{(7-3)!3!} = \frac{7 \times 6 \times 5}{2 \times 3} = 35$$

Dans le cas général, ce raisonnement donne la formule

$$C_n^p = \frac{n!}{(n-p)!p!} \tag{5.4}$$

Exemple

Dans un championnat de football, il y a 20 équipes, où pendant la saison chaque deux équipes doivent se rencontrer.

Quelle est le nombre total de matchs dans les deux cas suivants.

1)- Chaque deux équipes s'affrontent une fois.

2)- chaque deux équipes font deux matchs : un match "aller" et un match "retour".

Solution 1)- Chaque deux équipes des 20 équipes font un match, donc le nombre de matchs est égal au nombre de combinaisons

$$C_{20}^{2} = \frac{20}{18!2!} = 190 \text{ matchs.}$$

2)– Dans cette situation, pour programmer un match on choisit les deux équipes qui vont s'affronter et ensuite on les ordonne en commençant par exemple par l'équipe qui reçoit, la programmation d'un match correspond donc au choix d'un arrangement de deux des vingt équipes : le nombre de matchs est égal au nombre d'arrangements

$$A_{20}^{2} = 20 \times 19 = 380 \text{ matchs.}$$

5.2.3 Combinaison avec répétition.

Définition

On appelle combinaison avec répétition de p éléments d'un ensemble E à n éléments une liste de p éléments de l'ensemble E, où un élément peut apparaitre plus qu'une fois sur cette liste, et où l'ordre d'écriture des p éléments n'est pas pris en considération.

Par exemple $\{1,1,2,2\}$ et $\{1,2,2,1\}$ représente la même combinaison avec répétition de 4 éléments de l'ensemble E des 10 numéros de 1 à 10 .

Par contre, ils représentent deux arrangements avec répétitions différents de 4 éléments de l'ensemble E des dix numéros de 1 à 10.

Remarque importante. Pour former des combinaison avec répétition de p éléments d'un ensemble à n éléments, il n'est pas nécessaire que p ne soit pas plus grand que n, par exemple $\{1,1,1,1,2\}$ est une combinaison avec répétition de 5 éléments de l'ensemble $\{1,2\}$ à deux éléments.

5.2.4 Formule.

On accepte sans démonstration le résultat suivant. le nombre des combinaisons avec répétitions de p éléments d'un ensemble à n éléments, noté K_n^p, est donné par la formule

$$K_n^p = C_{n+p-1}^p = \frac{(n+p-1)!}{(n-1)!p!}. \tag{5.5}$$

Exemple. Une association a l'intention de distribuer vingt lits médicaux à cinq centres médicaux. De combien de façons peut-elle répartir ces vingt lits sur les cinq centres médicaux.

Réponse.

Le nombre de façons possibles est égal à

$$K_5^{20} = C_{24}^{20} = \frac{24!}{19!4!} = 212520 \text{ façons}$$

Combinaison - combinaison avec répétition

Une combinaison est déterminée par les éléments qui la composent, mais une combinaison avec répétition est déterminée par "les éléments qui la composent et le nombre de répétition de chacun de ces éléments".

5.2.5 Permutation avec répétition.

Au lieu de donner directement la définition de la Permutation avec répétition, avec ses notations encombrantes, il vaut mieux l'introduire par des exemples.

Exemple 1.

Dans le cadre d'une décoration, on a 8 pots de même forme mais de couleurs différentes : 3 rouges 4 bleus et 1 jaune. On place ces pots côte à côte sur une étagère. Quel est le nombre de dispositions possibles ?

Solution. Supposons qu'on place ces pots de droite à gauche. On peut par

exemple commencer par les pots rouges, en les plaçant côte à côte, ils vont occuper les places 1,2,3, on place ensuite les 4 pots bleus, ils occuperont les places 4,5,6,7, on met enfin le pot jaune après les 7 pots déjà placés, il occupera la dernière place, c'est-à-dire la $8^{ème}$, il y a une seule façon de le placer.

On va maintenant construire un raisonnement logique qui nous permettra de trouver le nombre total de façons à placer ces 8 pots côte à côte (sans faire de distinction entre les pots de même couleur).

De droite à gauche il y a 8 positions.

On commence par choisir 3 positions parmi ces 8 positions pour les 3 pots rouges, pratiquement çà veut dire choisir 3 numéros parmi les 8, ça revient à choisir une combinaison de 3 numéros parmi les 8 numéros : on a C_8^3 façons de le faire.

Après avoir choisi les positions des pots rouges, on choisit 4 positions pour les pots bleus, ça veut dire choisir 4 numéros parmi les 5 numéros qui restent, ça revient à choisir une combinaison de 4 numéros parmi les cinq qui restent, on a C_5^4 façons de le faire.

Enfin, on attribue le numéro qui reste au pot jaune, on a une seule façon de le faire, c'est-à-dire $1 = C_1^1$.

Le nombre total de façons de placer ces pots côte à côte est égal à

$$C_8^3 \times C_5^4 \times C_1^1 = \frac{8!}{3!4!1!}.$$

Ce raisonnement est valable dans le cas général, qu'on va exposer tout de suite.

Définition

Soit r ensembles $E_1, E_2, ...E_r$ ayant respectivement $n_1, n_2, ..., n_r$ éléments. On considère l'ensemble E réunion de ces r ensembles : $E = E_1 \cup ... \cup E_r$. les ensembles $E_1, E_2, ...E_r$ sont des parties de L'ensemble E et le nombre n des éléments de E est $n = n_1 + n_2 + ... + n_r$.

L'emplacement des n éléments de l'ensemble E côte à côte, sans faire de distinction entre les éléments d'une même partie E_i s'appelle une permutation avec répétition de l'ensemble E à n éléments, relativement à la partition $(E_1, n_1), ... (E_r, n_r)$. Le nombre de ces permutations avec répétitions est noté $P_{n,n_1,n_2,...,n_r}$.

En calquant le raisonnement fait dans l'exemple précédent, on trouve la formule

$$P_{n,n_1,n_2,...,n_r} = C_n^{n_1} C_{n-n_1}^{n_2} C_{n-(n_1+n_2)}^{n_2} ... C_{n_r}^{n_r} = \frac{n!}{n_1!n_2!...n_r!} \tag{5.6}$$

Exemple 2. Tahar veut préparer à son fils un programme de révision pour l'examen de sixième. La durée de révision est de 7 jours (de vendredi à jeudi), il veut les répartir sur les trois matières comme suit : 3 jours pour le calcul, 3 jours pour l'arabe et un jour pour le français.

1)- De combien de façons peut-il répartir les 7 jours de la semaine sur les trois matières ?

2)- On suppose que, l'arabe est composée de trois parties A_1, A_2, A_3, le calcul est également composé de trois parties C_1, C_2, C_2 et le français est composé d'une seule partie F.

Tahar a pensé qu'il vaut mieux, fixer à chaque partie des trois matières un jour de révision. De combien de façon peut-il programmer, dans ce cas, la révision de son fils ?

Solution

1)- Il choisit 3 jours de la semaine pour l'arabe. Parmi les 4 jours qui restent, il en choisi 3 pour le calcul, et le jour qui reste sera forcément attribué au français : le nombre de façons est égal au nombre $P_{7,3,3,1}$ de permutations avec répétitions donné par la formule

$$P_{7,3,3,1} = \frac{7!}{3!3!1!} = 140 \text{ façons.}$$

2)- Dans ce cas, chaque partie (des 7 parties) sera attribuée à un jour de la semaine, le nombre de façons est égal au nombre de permutations

$$P_8 = 7!. = 5040 \text{ façons}$$

Chapitre 6

Notions de probabilités.

6.1 Introduction

Considérons le jeu suivant. on met dans une boite 20 boules de même forme, mais de couleurs différentes, 8 blanches et 12 noires.

On se pose la question suivante. Si on tire de la boite une boule au hasard, laquelle des deux couleurs (blanche ou noire) a plus de chances de sortir ?

Intuitivement et en suivant " le bon sens " la réponse dépendra du rapport entre le nombre des blanches et celui des noires.

a- si le nombre des blanches est plus grand que celui des noires, la couleur blanche a plus de chance de sortir.

b- si le nombre des blanches est plus petit que celui des noires, la couleur noire a plus de chance de sortir.

c- si le nombre des blanches est égal à celui des noires, la couleur blanche a les mêmes chances de sortir que la couleur noire.

Dans d'autres expériences, l'intuition ne suffit pas de choisir une réponse évidente. Alors, en se basant sur le bon sens, on construit des modèles mathématiques qui nous permettent de donner une réponse logique. C'est le cas dans certains jeux de hasard. Au début du dix septième siècle le problème de choisir des issues favorables (ayant plus de chances de sortir) dans les jeux de hasard a posé problème aux notables d'Europe qui participaient à ces jeux. Ces notables avaient alors sollicité l'aide des mathématiciens de l'époque pour les aider à prévoir les issues les plus favorables dans ces jeux. On estime que c'est l'origine de la théorie mathématique des probabilités. Plusieurs éminents mathématiciens avaient alors participé à asseoir les fondements de la théorie, comme Pascal, Fermat, Laplace,...,

La formulation axiomatique de la théorie des probabilités à permis l'utilisation d'outils mathématiques puissants, comme la théorie des ensembles et la théorie de la mesure. La théorie de probabilité est devenue aujourd'hui un outil incontournable dans toutes les sciences : économie, physique, biomédical, ...

6.1.1 Exemple

Un dé est un cube, sur chacun de ses 6 faces est inscrit un numéro, de 1 à 6. On considère l'expérience suivante. On jette un dé bien équilibré, et on note le numéro sur la face d'en haut. l'ensemble des résultats possibles est l'ensemble $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Cet ensemble s'appelle l'espace fondamental lié à cette expérience.

Chaque partie de l'ensemble Ω s'appelle événement, par exemple l'ensemble $\{1, 4, 6\}$ est un événement.

L'événement Ω composé de tous les résultats possibles est appelé l'événement certain.

Un événement composé d'un seul résultat possible est appelé événement élémentaire, par exemple $\{1\}$ est un événement élémentaire.

Intuitivement, on peut dire que chaque numéro des 6 a les mêmes chances d'apparaitre. Le nombre total des numéros étant égal à 6, chaque numéro a une chance sur 6 d'apparaitre. Selon la même logique, on peut dire que les chances d'apparaître d'un évènement sont estimées par le rapport entre le nombre des résultats possibles qui le compose et le nombre 6 de tous les résultats possibles. On estime donc les chances d'apparaitre d'un évènement lié à l'expérience par un nombre positif compris entre 0 et 1.

A partir de tel raisonnement basé sur le bon sens, les mathématiciens ont progressivement construit les fondements mathématiques de la théorie des probabilités jusqu'à l'aboutissement au modèle final.

6.2 Définitions

Considérons une certaine expérience (pour se fixer les idées, vous pouvez faire la liaison avec l'expérience du dé décrite ci-dessous).

6.2.1 Ensemble fondamental, événement.

L'ensemble fondamental lié à une certaine expérience est l'ensemble formé de tous les résultats possibles de l'expérience. On le note généralement Ω. Dans l'expérience du dé, on a

$$\Omega = \{1, 2, 3, 4, 5, 6\} .$$

On appelle événement toute partie de l'ensemble fondamental Ω. Dans l'expérience du dé $A = \{3, 4, 5\}$ est un événement. L'ensemble des événements est donc l'ensemble des parties de Ω, noté $P(\Omega)$.

Un événement élémentaire est un événement, formé d'un seul élément de l'ensemble fondamental Ω, dans l'expérience du dé, $\{3\}$ est un événement élémentaire.

L'événement formé de tous les éléments de l'ensemble fondamental Ω (c'est-à-dire, Ω lui même), s'appelle l'événement sûr, ou l'évènement certain.

L'ensemble vide $\emptyset$, c'est-à-dire ne contenant aucun élément de Ω s'appelle l'événement impossible.

6.3 Opérations sur les événements.

Puisque les événements sont des parties d'un ensemble, toutes les opérations sur les ensembles (union, intersection, complémentaire,...) sont valables pour les événements, et toutes les propriétés de ces opérations sur les ensembles sont valables pour ces opérations sur les événements.

6.3.1 Négation d'un événement.

La négation d'un événement A, noté $\overline{A}$ (on lit A barre) ou CA (on lit complémentaire de A), est l'événement formé des résultats possibles qui ne sont pas dans A. Donc, en langage ensembliste, $\overline{A}$ est le complémentaire de A.

6.3.2 Intersection de deux événements.

L'intersection de deux événements A et B, est l'événement noté $A \cap B$ (on lit A inter B), formé des résultats possibles qui sont à la fois dans A et dans B.

6.3.3 Union de deux événements.

L'union de deux événements A et B est l'événement $A \cup B$ (on lit A union B), formé des résultats possibles qui sont, soit dans A et non dans B, soit dans B et non dans A, soit à la fois dans A et dans B.

6.3.4 Exemple.

Considérons l'expérience du jet du dé décrite précédemment.

L'évènement $A = \{4, 5\}$ correspond à la face 4 ou la face 5. La réalisation de l'évènement A veut dire "après le jet du dé, la face du dé obtenue est l'une des deux faces, 4 ou 5".

L'évènement $B = \{2, 4, 3\}$ correspond à l'une des trois faces, 2, 4 ou 3.La réalisation de l'évènement B veut dire "après le jet du dé, la face du dé obtenue est l'une des trois faces, 2, 4 ou 3".

La négation de l'évènement A est l'évènement

$$\overline{A} = \{1, 2, 3, 6\} .$$

La réalisation de l'évènement $\overline{A}$ signifie la non réalisation de l'évènement A. Dans cet exemple la réalisation de $\overline{A}$, signifie que la face du dé obtenue porte l'un des numéros $1, 2, 3, 6$.

L'union des évènements A et B est l'évènement

$$A \cup B = \{2, 3, 4, 5\} .$$

La réalisation de l'évènement $A \cup B$ signifie la réalisation de l'un des deux évènements A ou B au moins. Dans cet exemple la réalisation de $A \cup B$, signifie que la face du dé obtenue porte l'un des numéros $2, 3, 4, 5$.

L'intersection des évènements A et B est l'évènement

$$A \cap B = \{4\} .$$

La réalisation de l'évènement $A \cap B$ signifie la réalisation des deux évènements A ou B en même temps. Dans cet exemple la réalisation de $A \cap B$ signifie que la face obtenue porte le numéro 4.

6.4 Probabilité sur un ensemble.

Dans l'expérience du dé, si on associe à chaque événement le nombre positif compris entre 0 et 1, égal au rapport du nombre des résultats possibles composant cet événement et le nombre 6 égal au nombre total de tous les résultats possibles de l'expérience, on dit qu'on a défini une probabilité sur l'ensemble fondamental Ω.

Cette procédure est un cas particulier qui ne couvre pas toutes les situations importantes dans lesquelles s'impliquent les probabilités. On pense donc qu'il est plus judicieux de présenter la théorie en utilisant l'approche axiomatique.

6.4.1 Ensemble probabilisable.

Soit Ω un ensemble non vide (contenant au moins un élément), et soit $\mathcal{A}$ une partie non vide de $\mathcal{P}(\Omega)$ $(\emptyset \neq \mathcal{A} \subset \mathcal{P}(\Omega))$.

A- On dit que $\mathcal{A}$ est une algèbre de $\mathcal{P}(\Omega)$ si $\mathcal{A}$ est stable par les deux

opérations ensemblistes, l'union et le complémentaire. C'est-à-dire, si les deux conditions suivantes sont vérifiées.

a- $A \in \mathcal{A}$ et $B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$.

b- $A \in \mathcal{A} \Rightarrow CA \in A$.

B- On dit que $\mathcal{A}$ est une σ-algèbre ou tribu de $\mathcal{P}(\Omega)$ si la condition b- dans **A-** est remplacée par la condition plus forte suivante, de la stabilité par union dénombrable.

c- Si $A_1, A_2, ...$ est une suite d'éléments de $\mathcal{A}$ alors $\overset{\infty}{\underset{i=1}{\cup}} A_i \in \mathcal{A}$.

C- Le couple $(\Omega, \mathcal{A})$, formé d'un ensemble Ω et une tribu $\mathcal{A}$ de l'ensemble $\mathcal{P}(\Omega)$ des parties de l'ensemble Ω, s'appelle ensemble **probabilisable.**

6.4.2 Ensemble probabilisé.

Définition d'un ensemble probabilisé.

On appelle une probabilité $\mathbf{P}$ sur un ensemble probabilisable $(\Omega, \mathcal{A})$, une application qui associe à chaque élément A de $\mathcal{A}$, un nombre $\mathbf{P}(A)$ compris entre 0 et 1, et qui vérifie les propriétés suivantes.

a- $\mathbf{P}(\Omega) = 1$

b- Si $A_1, A_2, ...$ est une suite d'éléments de $\mathcal{A}$ disjoints deux à deux, c'est-à-dire telle que $A_n \cap A_m = \emptyset$ pour tous $m \neq n$, alors on a
$\mathbf{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} \mathbf{P}(A_i)$.

Dans ce cas, on appelle le triplet $(\Omega, \mathcal{A}, \mathbf{P})$ un ensemble probabilisé.

6.4.3 Propriétés d'une probabilité.

1- Si $A_1, ..., A_n$ est une suite finie de n événements disjoints deux à deux, alors on a

$$\mathbf{P}(A_1 \cup ... \cup A_n) = \mathbf{P}(A_1) + ... + \mathbf{P}(A_n). \tag{6.1}$$

2-

$$\mathbf{P}(\emptyset) = 0. \tag{6.2}$$

3-

$$\mathbf{P}(A) = 1 - \mathbf{P}(CA). \tag{6.3}$$

4-

$$A \subset B \Rightarrow \mathbf{P}(A) \leq \mathbf{P}(B) \tag{6.4}$$

5-

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B) \tag{6.5}$$

.

5 Bis- La propriété 5- peut se généraliser à l'union d'un nombre fini quelconque d'évènements. Dans le cas de l'union de trois évènements on a

$$\begin{aligned} \mathbf{P}(A \cup B \cup C) &= \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) + \mathbf{P}(A \cap B \cap C) \\ &\quad - [\mathbf{P}(A \cap B) + \mathbf{P}(A \cap C) + \mathbf{P}(B \cap C)] \end{aligned} \tag{6.6}$$

Preuve.

Pour prouver 1-, on complète la suite par des ensemble vides et on applique la propriété b- :

$$\mathbf{P}(A_1 \cup ... \cup A_n) = \mathbf{P}(A_1 \cup ... \cup A_n \cup \emptyset \cup ...) =$$
$$= \mathbf{P}(A_1) + ... + \mathbf{P}(A_n) + \mathbf{P}(\emptyset) + ... = \mathbf{P}(A_1) + ... + \mathbf{P}(A_n).$$

Pour prouver 2- on utilise la propriété 1- juste prouvée,

$$\mathbf{P}(\emptyset) = \mathbf{P}(\emptyset \cup \emptyset) = \mathbf{P}(\emptyset) + \mathbf{P}(\emptyset).$$

Donc on a $\mathbf{P}(\emptyset) = \mathbf{P}(\emptyset) + \mathbf{P}(\emptyset)$. Retranchant $\mathbf{P}(\emptyset)$ des deux membres de cette égalité, on obtient $\mathbf{P}(\emptyset) = 0$.

Pour prouver 3-, écrivant Ω sous la forme $\Omega = A \cup CA$ et appliquant 1- ; on obtient

$$1 = \mathbf{P}(\Omega) = \mathbf{P}(A \cup CA) = \mathbf{P}(A) + \mathbf{P}(CA).$$

4- Puisque $A \subset B$, on a $B = A \cup (B \cap CA)$. Des deux propriétés de la tribu $\mathcal{A}$ découle l'appartenance de l'ensemble $B \cap CA$ à $\mathcal{A}$. On peut donc appliquer la propriété 1- et écrire

$$\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \cap CA) \geq \mathbf{P}(A).$$

5- On remarque que

$$\begin{aligned} A &= (A \cap B) \cup (A \cap CB) \\ B &= (B \cap A) \cup (B \cap CA) \\ A \cup B &= (A \cap CB) \cup (B \cap CA) \cup (A \cap B), \end{aligned}$$

où dans chacune des trois égalités, les ensembles à droite sont deux à deux disjoints. Il suffit donc d'appliquer la propriété 1-.

6.4.4 Cas d'ensemble finis.

Si $\Omega = \{a_1, a_2, ..., a_N\}$ est un ensemble fini composé de N éléments $(N \geq 1)$. On définit une probabilité $\mathbf{P}$ sur l'ensemble probabilisé $(\Omega, \mathcal{P}(\Omega))$ de la façon suivante. On associe à chaque événement élémentaire a_i de l'ensemble Ω un nombre p_i compris entre 0 et 1, tels que

$$p_1 + p_2 + ... + p_N = 1. \tag{6.7}$$

Et, on associe à chaque événement (c'est-à-dire à chaque partie de l'ensemble Ω) la somme des probabilités des événements élémentaires le composant.

6.4.5 Ensemble Equiprobable.

Si dans la définition précédente,

$$p_1 = p_2 = ... = p_N = \frac{1}{N}, \tag{6.8}$$

on dit que l'ensemble probabilisé $(\Omega, \mathcal{P}(\Omega), \mathbf{P})$ est équiprobable.

Dans le cas d'ensembles équiprobables la probabilité d'un évènement A est donnée par la formule

$$\mathbf{P}(A) = \frac{nombre\ de\ cas\ favorables}{nombre\ de\ cas\ possibles}, \tag{6.9}$$

où le nombre de cas favorables est le nombres des éléments de l'événement A (c'est-à-dire le nombre des événements élémentaires le composant) et le nombre de cas possibles est égal au nombre total des éléments de l'espace fondamental Ω (c'est-à- dire le nombre de tous les événements élémentaires).

C'est surtout dans le cas d'ensembles équiprobables qu'on utilise les techniques d'analyse combinatoire.

6.4.6 Incompatibilité- indépendance d'événements.

Deux événements A et B d'un ensemble probabilisé sont dits incompatibles si

$$A \cap B = \emptyset. \tag{6.10}$$

Deux événements A et B d'un ensemble probabilisé sont dits indépendants si

$$\mathbf{P}(A \cap B) = \mathbf{P}(A).\mathbf{P}(B). \tag{6.11}$$

6.5 Exemples.

6.5.1 Exemple.

On considère l'expérience du jet de la pièce de monnaie trois fois de suite. Un résultat possible de cette expérience est exprimé par une suite de 3 lettres, chaque lettre est P ou F, selon le résultat obtenue. Par exemple, si le premier jet donne pile, le second donne face et le troisième donne pile, le résultat de l'expérience est PFP.

L'ensemble fondamental est

$$\Omega = \{PPP,\ PPF,\ PFP,\ PFF,\ FPP,\ FPF,\ FFP,\ FFF\ \}$$

.

Posons

$$\begin{array}{c} \mathbf{P}\left(PPP\right) = \mathbf{P}\left(PPF\right) = \mathbf{P}\left(PFP\right) = \mathbf{P}\left(PFF\right) = \\ = \mathbf{P}\left(FPP\right) = \mathbf{P}\left(FPF\right) = \mathbf{P}\left(FFP\right) = \mathbf{P}\left(FFF\right) = \frac{1}{8}. \end{array}$$

On définit ainsi un ensemble équiprobable. Cela est en accord avec le bon sens, parce que après les trois jets de la pièce de monnaie, tous les résultats possibles ont les mêmes chances de sortir.

Considérons les événements

A : aucun jet des trois ne donne pile,

B : au moins l'un des trois jets donne pile,

C : Il y a exactement un pile dans les trois jets,

D : le premier jet donne pile,

E : le deuxième jet donne face.

On a

$$A = \{FFF\},$$

$$B = \{PPP,\ PPF,\ PFP,\ PFF,\ FPP,\ FPF,\ FFP\},$$

$$C = \{PFF, FPF, FFP\}.$$

$$D = \{PPP, PPF, PFP, PFF\}$$

$$E = \{PFP, PFF, FFP, FFF\},$$

d'où

$$\begin{aligned}
\mathbf{P}(A) &= \frac{1}{8}, \\
\mathbf{P}(B) &= \frac{7}{8}, \\
\mathbf{P}(C) &= \frac{3}{8}, \\
A \cup B &= \Omega, \\
\mathbf{P}(D) &= \mathbf{P}(E) = \frac{4}{8} = \frac{1}{2}
\end{aligned}$$

Puisque

$$A \cap B = \emptyset,$$

alors les événements A et B sont incompatibles. On a

$$D \cap E = \{PFP, PFF\},$$

d'où

$$\mathbf{P}(D \cap E) = \frac{2}{8} = \frac{1}{4},$$

et

$$\mathbf{P}(D \cap E) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = \mathbf{P}(D) \times \mathbf{P}(E).$$

Donc, les événements D et E sont indépendants.

6.5.2 Exemple.

On lance simultanément une pièce de monnaie et un dé. Le dé a 6 faces numérotés de 1 à 6 et la pièce de monnaie en a 2, pile et face.

Déterminer l'ensemble fondamental Ω dans les cas suivants.

a- On s'intéresse à la fois, au résultat de la pièce de monnaie (pile ou face) et au résultat du dé (un numéro des six).

b- on s'intéresse uniquement au résultat de la pièce de monnaie.

c- on s'intéresse uniquement au résultat du dé.

Solution.

a- On a deux résultats possibles pour la pièce de monnaie et chaque résultat de la pièce de monnaie lui correspond 6 résultats du dé. Donc, on a au total 2×6 résultats qui sont

$\Omega = \{(p,1),(p,2),(p,3),(p,4),(p,5),(p,6),(f,1),(f,2),(f,3),(f,4),(f,5),(f,6)\}.$

b- $\Omega = \{p, f\}$.

c- $\Omega = \{1,2,3,4,5,6\}$.

6.5.3 Exemple.

Dans le cas fini, il est plus convenable de travailler sur des ensembles équiprobables. Car, on peut dans cette situation utiliser les techniques d'analyse combinatoire. Soit l'exemple suivant.

On a 7 enveloppes fermées, contenant chacune un bout de papier sur lequel est inscrit un nombre. Ces nombres sont répartis comme suit. Une enveloppe contient le nombre 0, deux enveloppes contiennent chacune le numéro 10, deux enveloppes contiennent chacune le numéro 20 et deux enveloppes contiennent chacune le numéro 30. On prend 2 enveloppes au hasard, et on mentionne la somme des deux numéros dans les deux enveloppes.

- Quelle est la probabilité que la somme des deux nombres est égale à 0 ?
- Quelle est la probabilité que la somme est égale à 50 ?
- Quelle est la probabilité que la somme est égale à 10 ?
- Quelle est la probabilité que la somme est égale à 60 ?
- Quelle est la probabilité que la somme soit au moins égale à 50 ?

Solution.

On a 7 enveloppes, qu'on imagine numérotées de 1 à 7. Un résultat possible de cette éxpérience est une combinaison de deux enveloppes (donc de deux numéros) parmi les 7. L'ensemble fondamental Ω est l'ensemble de toutes les combinaisons de deux numéros parmi les 7 numéros, Ω contient

$$C_7^2 = \frac{7!}{5!2!} = 21$$

éléments. Chaque combinaison de deux numéros a les mêmes chances d'être tirée. Il s'agit donc d'un espace équiprobable, où la probabilité d'un évènement élémentaire est égale à

$$\frac{1}{21},$$

et la probabilité d'un évènement quelconque est le rapport du nombre des combinaisons qui la composent et le nombre 21. On note par 1 l'enveloppe contenant 0, par 2 une enveloppe contenant 10, par 3 la deuxième enveloppe contenant 10, par 4 une enveloppe contenant 20, par 5 la deuxième enveloppe contenant 20, par 6 une enveloppe contenant le nombre 30 et enfin par 7 la dernière enveloppe qui va forcément contenir le nombre 30. L'opération de prendre deux enveloppes parmi les 7 signifie le prise d'une combinaison de deux numéros parmi les 7.

- Aucune combinaison ne donne zéro, donc notre évènement ne contient aucun élément de Ω, c'est l'ensemble vide,

$$\text{(La probabilité que la somme soit zéro)} = \mathbf{P}(\emptyset) = 0$$

- Les éléments de Ω qui correspondent à l'événement " la somme est 50" sont $(4,6)$, $(4,7)$, $(5,6)$, $(5,7)$. Leur nombre est 4. Donc,

$$\text{La probabilité que "la somme est 50" est } \frac{4}{21}$$

- Les éléments de Ω qui correspondent à l'évènement "la somme est 10" sont $(1,2)$, $(1,3)$. Leur nombre est 2. Donc,

$$\text{La probabilité que "la somme est 10" est } \frac{2}{21}$$

- Il y a un élément et un seul qui correspond à l'événement "la somme est 60", c'est la combinaison $(6,7)$. Donc,

$$\text{La probabilité que "la somme est 60" est } \frac{1}{21}$$

- Ici la somme doit être égale à 50 ou égale à 60. Les éléments de cet évènement sont $(4,6)$, $(4,7)$, $(5,6)$, $(5,7)$, $(6,7)$. Leur nombre est 5. Donc

$$\text{La probabilité que "la somme soit au moins 50" est } \frac{5}{21}$$

6.5.4 Exemple.

On met dans une boite 20 boules de même forme, mais de couleurs différentes : 5 blanches, 8 noires et 7 jaunes.

A. Tirage sans remise.

On tire une première boule on constate sa couleur et on la met de coté, on tire ensuite une deuxième boule on constate sa couleur et on la met de coté, on tire enfin une troisième boule et on constate sa couleur.

a- Quelle est la probabilité que la première boule soit Jaune ?

b- Quelle la probabilité que la première boule et aussi la deuxième soient jaunes.

c- Quelle est la probabilité que les trois boules soient de même couleur ?

B. Tirage avec remise.

On refait la même expérience, mais cette fois on remet chaque boule tirée dans la boite après avoir constaté sa couleur, et on fait ensuite le tirage suivant.

On répond aux mêmes questions a-,b- et c- considérées dans **A.**

Solution.

A. a- Les tirages deux et trois n'interviennent pas. Le nombre de cas possibles est égal à 20. On a 7 boules jaunes, le nombre de cas favorables est égal à 7. La probabilité que le premier tirage donne une boule jaune est égale

$$\frac{7}{20}$$

.

b- On considère les deux premiers tirages à la fois. Chaque résultat possible est un arrangement de deux boules parmi les 20, le nombre de cas possible est égal à $A_{20}^2 = 20 \times 19 = 380$. Chaque résultat favorable correspond à un arrangement de deux boules parmi les 7 boules jaunes, le nombre de cas favorables est donc égal à $A_7^2 = 7 \times 6 = 42$. La probabilité est donc égale à

$$\frac{A_7^2}{A_{20}^2} = \frac{42}{380} = \frac{21}{190}$$

.

c- Le nombre de cas possibles est égal à A_{20}^3. Pour avoir trois boules de même couleur, il faut que les 3 boules soient toutes de couleur blanche, de couleur noire ou de couleur jaune. Donc le nombre de cas favorables est égal à la somme des cas favorables correspondants à chacune des trois couleurs, donc égal à A_5^3+ A_8^3+ A_7^3. La probabilité est donc égale à

$$\frac{A_5^3 + A_8^3 + A_7^3}{A_{20}^3} = \frac{606}{6840}.$$

B. Puisque avant le tirage suivant, on remet dans la boite la boule tirée, la même boule peut apparaitre dans deux tirages ou même, dans les trois tirages. On a dans ce cas affaire à des arrangements avec répétition.

Les probabilités en questions seront égales à

a- $$\frac{7}{20},$$

b- $$\frac{R_7^2}{R_{20}^2} = \frac{7^2}{20^2},$$

c- $$\frac{R_5^3 + R_8^3 + R_7^3}{R_{20}^3} = \frac{5^3 + 8^3 + 7^3}{20^3}.$$

6.5.5 Exemple.

Dans certaines situations, on ne connait pas l'ensemble fondamental. Dans de telles situations, les méthodes d'analyse combinatoire n'interviennent pas et on utilise uniquement les définitions et les données spécifiques au problème considéré.

1- Soit A et B deux événements liés à une même expérience tels que

$$P(A) = \frac{1}{5},\ P(B) = \frac{1}{3},\ P(A \cap B) = \frac{2}{15}.$$

Calculer les probabilités des évènements

E : au moins l'un de ces événements se produit, et

F : un seul et un seul de ces deux événements se produit.

Solution. a-

$$E = A \cup B$$

, d'où

$$\begin{aligned} P(E) &= P(A \cup B) \\ &= P(A) + P(B) - P(A \cap B) \\ &= \frac{1}{5} + \frac{1}{3} - \frac{2}{15} = \frac{6}{15} \end{aligned}$$

b- l'événement F correspond à la situation suivante : "la réalisation de A et la non réalisation de B" ou "la réalisation de B et la non réalisation de A". Cela signifie en langage d'ensembles

$$F = \left(A \cap \overline{B}\right) \cup \left(\overline{A} \cap B\right).$$

Puisque $\left(A \cap \overline{B}\right)$ et $\left(\overline{A} \cap B\right)$ sont disjoints, on a

$$P(F) = P\left(A \cap \overline{B}\right) + P\left(\overline{A} \cap B\right)$$

.

D'autre part on a

$$A = \left(A \cap \overline{B}\right) \cup (A \cap B)$$

et

$$B = \left(\overline{A} \cap B\right) \cup (A \cap B),$$

où la réunion est disjointe dans les deux cas. On en déduit les relations suivantes

$$P(A) = P\left(A \cap \overline{B}\right) + P(A \cap B),$$

$$P(B) = P\left(\overline{A} \cap B\right) + P(A \cap B).$$

D'où

$$\begin{aligned} P(F) &= P(A) + P(B) - 2P(A \cap B) \\ &= \frac{1}{5} + \frac{1}{3} - 2\frac{2}{15} \\ &= \frac{4}{15}. \end{aligned}$$

6.6 Probabilité conditionnelle

6.6.1 Définition

Soient A et B deux événements d'un même ensemble fondamental tel que $P(B) \neq 0$.

Le nombre $P(A/B)$ défini par

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \tag{6.12}$$

s'appelle probabilité de A sachant B.

6.6.2 Remarque

L'utilisation de la définition nous permet d'avoir la formule

$$P(A/B) + P\left(\overline{A}/B\right) = 1.$$

.

6.6.3 Cas particulier : ensemble équiprobable

Dans le cas d'ensemble équiprobable on a

$$P(A/B) = \frac{\text{nombre d'éléments de } (A \cap B)}{\text{nombre d'éléments de } (B)}. \tag{6.13}$$

En effet, on a

$$\begin{aligned} P(A/B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{\frac{\text{nombre d'éléments de } (A \cap B)}{N}}{\frac{\text{nombre d'éléments de } (B)}{N}}. \\ &= \frac{\text{nombre d'éléments de } (A \cap B)}{\text{nombre d'éléments de } (B)}. \end{aligned}$$

Donc, dans le cas d'ensemble équiprobable, la probabilité d'un évènement A sachant un autre évènement B, est égale à la proportion de $A \cap B$, comparativement à B.

6.6.4 Exemple

les données du groupage de sang mesurées sur 150 personnes, 100 hommes et 50 femmes sont représentées sur le tableau de contingence suivant.

sexe\groupe	O	A	B	AB	totaux
homme	50	20	20	10	100
Femme	20	10	15	5	50
totaux	70	30	35	15	150

1)- On choisit au hasard une personne parmi ces 150 personnes. Calculer les probabilités des évènements suivants.

A_1 : le groupe de la personne choisie au hasard est O.

A_2 : La personne choisie au hasard est une femme.

A_3 : La personne choisie est une femme et son groupe de sang est O.

A_4 : Le groupe de la personne choisie au hasard n'est pas A et n'est pas B.

A_5 : Le groupe de la personne choisie au hasard est AB.
2)- On choisit au hasard une femme. Calculer les probabilités suivantes.
- le groupe de cette femme est O.
- le groupe de cette femme est O ou AB.

Solution.

1)- On est dans la situation d'un espace probabilisé équiprobable. L'ensemble fondamental est l'ensemble des 150 personnes.

La probabilité de chaque évènement est égale

$$\frac{\text{nombre des cas favorables}}{\text{nombre des cas possibles}}.$$

donc,

$$P(A_1) = \frac{70}{150} = \frac{7}{15}.$$
$$P(A_2) = \frac{50}{150} = \frac{1}{3}.$$
$$P(A_3) = \frac{20}{150} = \frac{2}{15}.$$
$$P(A_4) = \frac{70+15}{150} = \frac{85}{150} = \frac{17}{30}.$$

2)-.Puisqu'on sait que la personne choisie est une femme, alors les probabilités seront des probabilités conditionnelles : sachant que la personne choisie est une femme, c'est-à-dire sachant l'évènement A_2 :

La probabilité que le groupe de la femme choisie soit O égale

$$P(A_1 \diagup A_2) = \frac{20}{50} = \frac{2}{5}.$$

La probabilité que le groupe de la femme choisie soit O ou AB égale

$$P((A_1 \cup A_5) \diagup A_2) = \frac{20+5}{50} = \frac{1}{2}.$$

6.7 Probabilités composées.

6.7.1 Cas de deux événements.

Si $P(A) \neq 0$ et $P(B) \neq 0$, on déduit de la formule (6.12), la formule suivante, appelée formule des probabilités composées.

$$P(A \cap B) = P(B)\,P(A/B) = P(A)\,P(B/A)\,. \tag{6.14}$$

6.7.2 Cas de plusieurs événements.

Soient $A_1, A_2, ..., A_n$ des événements d'un même ensemble fondamental tous de probabilité non nulle. On a

$$P(A_1 \cap A_2 \cap ... \cap A_n)\,. \tag{6.15}$$

$$= P(A_1)\,P(A_2/A_1)\,P(A_3/A_1 \cap A_2)\,...P(A_n/A_1 \cap ...A_{n-1})\,.$$

L'expression (6.15) s'écrit de la façon condensée suivante

$$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)\,.\prod_{k=2}^{n} P\left(A_k/_{A_1\cap ... \cap A_{k-1}}\right)$$

6.7.3 Exemple.

On met dans une urne 100 billes de même forme, mais de couleurs différentes : 20 rouges, 40 bleues, 15 noires et 25 jaunes.

On tire une boule, on constate sa couleur et on la met à coté. On tire une deuxième boule, on constate sa couleur et on la met à coté. On tire une troisième boule, on constate sa couleur et on la met à coté. On tire enfin une quatrième boule et on constate sa couleur.

On considère les 4 événements suivants.

A_1 : la première boule est de couleur jaune.

A_2 : la deuxième boule est de couleur jaune.

A_3 : la troisième boule est de couleur bleue.

A_4 : la quatrième boule est de couleur noire.

On peut calculer la probabilité de l'événement $A = A_1 \cap A_2 \cap A_3 \cap A_4$ par deux méthodes.

Méthode directe.

On imagine que les 100 boules sont numérotées de 1 à 100. Le tirage une à une sans remise de quatre boules donne un arrangement de 4 numéros parmi les 100 numéros, et chaque arrangement à les mêmes chances de sortir. On peut donc supposer que l'ensemble fondamental est un ensemble équiprobable où le nombre de ses éléments est égal à A_{100}^4. Dans ce cas la probabilité de chaque événement est donc égal au nombre des éléments de cet événement divisé par le nombre A_{100}^4.

Une suite de quatre tirages consécutifs est un élément de A si ses tirages donnent respectivement les couleurs suivantes, "jaune, jaune, bleue, noire".

On a 25 façons d'avoir une boule jaune dans le premier tirage.

A chaque façon du premier tirage correspond 24 façons d'avoir une boule jaune dans le deuxième tirage.

A chaque façon d'avoir une boule jaune dans le premier tirage et aussi une boule jaune dans le deuxième tirage correspond 40 façons d'avoir une boule bleue dans le troisième tirage.

A chaque façon d'avoir une boule jaune dans le premier tirage et aussi une boule jaune dans le deuxième tirage et une bleue dans le troisième tirage, correspond 15 façons d'avoir une noire dans le dernier tirage.

Le nombre d'éléments de l'événement A est donc égal à $25 \times 24 \times 40 \times 15$. Et, ainsi la probabilité de A est donnée par

$$P(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{25 \times 24 \times 40 \times 15}{100 \times 99 \times 98 \times 97}.$$

Formule des probabilités composées

Puisque le nombre des jaunes est 25 et le nombre total est 100, on a

$$P(A_1) = \frac{\text{nombres de cas favorables}}{\text{nombre de cas possibles}} = \frac{25}{100}.$$

Si le premier tirage donne jaune, il reste dans la boite 24 jaune et au total 99 boules (sans distinction entre les couleurs).

Donc la probabilité que le deuxième tirage donne jaune sachant que le premier est jaune est donnée par

$$P(A_2/A_1) = \frac{\text{nombres de cas favorables}}{\text{nombre de cas possibles}} = \frac{24}{99}$$

Si le premier et le deuxième tirage sont supposés jaunes, il reste dans la boite 40 boules bleues et au total 98 boules, donc la probabilité que le troisième tirage est bleu sachant que le premier et le deuxième est jaune est donnée par

$$P(A_3/A_1 \cap A_2) = \frac{\text{nombres de cas favorables}}{\text{nombre de cas possibles}} = \frac{40}{98}.$$

Si le premier tirage, le deuxième et le troisième sont supposés respectivement jaune, jaune et bleu, il reste dans la boite 15 boules noires et au total 97 boules, donc la probabilité que le quatrième tirage est noir sachant que le premier, le deuxième et le troisième sont jaune, jaune et bleu est donnée par

$$P(A_4/A_1 \cap A_2 \cap A_3) = \frac{\text{nombres de cas favorables}}{\text{nombre de cas possibles}} = \frac{15}{97}.$$

La formule des probabilités composées donne

$$\begin{gathered} P(A_1 \cap A_2 \cap A_3 \cap A_4) = \\ = P(A_1)\, P(A_2/A_1)\, P(A_3/A_1 \cap A_2)\, P(A_4/A_1 \cap A_2 \cap A_3) \\ = \frac{25}{100}\frac{24}{99}\frac{40}{98}\frac{15}{97}. \end{gathered}$$

On trouve évidemment le même résultat.

6.8 Système complet d'événements.

6.8.1 Définition

Un système complet d'événements est une partition de l'ensemble fondamental Ω par des événements $A_1, A_2, ...A_n$ tous de probabilité non nulle. En d'autres termes les événements $A_1, A_2, ...A_n$ forme un système complet si les conditions suivantes sont vérifiées.

a- $P(A_i) > 0$, pour tous $i = 1, 2, ...n$

b- $A_i \neq A_j$ pour tous $i \neq j$.

c- $\Omega = A_1 \cup A_2 \cup ... \cup A_n$.

6.8.2 Exemple.

On considère l'expérience du lancer de deux dés en même temps. l'ensemble fondamental est formé des couples ordonnés (p, r), où p et r sont deux entiers compris entre 1 et 6. On considère les événements A et B définies comme suit

A : la somme des résultats des deux dés est un nombre pair ($p + r$ est pair).

B :la somme des résultats des deux dés est un nombre impair ($p + r$ est impair).

Les deux événements forment un système complet.

6.8.3 Formule des probabilités totales

Si $A_1, A_2, ...A_n$ est un système complet et A un événement quelconque, alors on a

$$\begin{aligned} P(A) &= P(A \cap A_1) + P(A \cap A_2) + ... + P(A \cap A_N) \qquad (6.16) \\ &= P(A/A_1)\, P(A_1) + P(A/A_2)\, P(A_2) + ...P(A/A_N)\, P(A_N). \end{aligned}$$

a- La première identité résulte directement de la condition (6.1) et du fait que A est l'union disjointe des événements $(A \cap A_1)$,

$P(A \cap A_2)$, ... , $P(A \cap A_N)$.

b- L'expression de la deuxième ligne résulte de la formule (6.14) des probabilités composées.

6.9 Formule de Bayes.

6.9.1 Forme simple

Si A et B sont deux événements de probabilité non nulle, on a la formule suivante.

$$P(A/B) = \frac{P(A)\, P(B/A)}{P(B)}. \qquad (6.17)$$

Cette formule s'appelle formule simple de Bayes et résulte de la définition de la probabilité conditionnelle et de la formule (6.14) des probabilités composées.

6.9.2 Formule de Bayes.

Si, dans la formule (6.17), l'évènement A est remplacé par un élément A_i d'un système complet $A_1, ..., A_N$, on aura la formule de Bayes (finale) suivante

$$P(A_i/B) = \frac{P(A_i)\, P(B/A_i)}{P(A_1)\, P(B/A_1) + ... + P(A_N)\, P(B/A_N)}. \tag{6.18}$$

Pveuve. On utilise d'abord la formule simple de Bayes (6.17), et on applique ensuite la formule (6.16) des probabilités totales, pour écrire $P(B)$ sous la forme

$$P(B) = P(A_1)\, P(B/A_1) + ... + P(A_N)\, P(B/A_N).$$

■

6.9.3 Remarque.

La formule de Bayes nous permet de calculer la probabilité conditionnelle $P(A/B)$ en fonction de la probabilité conditionnelle $P(B/A)$, c'est-à-dire en inversant les rôles de A et B.

6.9.4 Exemple

Dans une classe de trente élèves, composée de 20 filles et 10 garçons, les deux langues étrangères sont le français et l'anglais.

On choisit au hasard un élève de la classe, et on définit les évènements suivants :

A : la langue étrangère de cet élève est l'anglais.

Fr : la langue étrangère de cet élève est le français.

G : cet élève est un garçon.

F : cet élève est une fille.

On suppose les données suivantes vérifiées

$$P(A/G)=\frac{2}{5},\ P(A/F)=\frac{1}{4}$$

a)- Calculer

$$P(G)\ \text{et}\ P(F).$$

b)- Calculer

$$P(A\cap G)\quad\text{et}\ P(A\cap F).$$

c)- En déduire

$$P(A)\ \text{et}\ P(F_r).$$

d)- Utiliser la formule de Bayes pour calculer

$$P(G/A).$$

e)- Déduire

$$P(F/A).$$

f)- Remplir le tableau de contingence des effectifs suivant

Sexe\Langue	anglais	français	totaux
garçon			
fille			
totaux			

Solution.

a)-

$$P(G)=\frac{10}{30}=\frac{1}{3}\ \text{et}\quad P(F)=\frac{20}{30}=\frac{2}{3}.$$

b)- on utilise la formule des probabilités composées :

$$P(A\cap G)=P(G)\,P(A/G)=\frac{1}{3}.\frac{2}{5}=\frac{2}{15},$$

$$P(A\cap F)=P(F)\,P(A/F)=\frac{2}{3}.\frac{1}{4}=\frac{1}{6}.$$

c)-

$$\begin{aligned} P(A) &= P(A\cap G)+P(A\cap F) \\ &= \frac{2}{15}+\frac{1}{6}=\frac{3}{10}, \end{aligned}$$

$$P(F_r)=1-P(A)=\frac{7}{10}.$$

d)-

$$\begin{aligned} P(G\diagup A) &= \frac{P(G\cap A)}{P(A)} \\ &= \frac{P(G)\,P(A\diagup G)}{P(A)} \\ &= \frac{\frac{1}{3}\frac{2}{5}}{\frac{3}{10}}=\frac{4.}{.9}. \end{aligned}$$

e)-

$$\begin{aligned} P(F\diagup A) &= 1-P(G\diagup A) \\ &= 1-\frac{4}{9}=\frac{5}{9}. \end{aligned}$$

f)- pour remplir le tableau de contingence on doit calculer les quatre effectifs conjoints $eff.(G\cap A)\,,\ eff.(G\cap F_r)\,,\ eff.(F\cap A)\,,\ eff.(F\cap F_r)\,.$

On a

$$\begin{aligned} eff.(G\cap A) &= P(G\cap A)\times N \\ &= \frac{2}{15}\times 30=4. \end{aligned}$$

$$\begin{aligned} eff.(G\cap F_r) &= eff(G)-eff(G\cap A) \\ &= 10-4=6. \end{aligned}$$

$$\begin{aligned} eff.(F\cap A) &= P(F\cap A)\times N \\ &= \frac{1}{6}\times 30=5. \end{aligned}$$

$$\begin{aligned} eff.\,(F \cap F_r) &= eff\,(F) - eff\,(F \cap A) \\ &= 20 - 5 = 15. \end{aligned}$$

Sexe\Langue	anglais	français	totaux
garçon	4	6	10
fille	5	15	20
totaux	9	21	30

Chapitre 7

Variables aléatoires

Les résultats possibles d'une expérience aléatoire ne sont pas toujours exprimés numériquement. Une variable aléatoire consiste en l'association à tout résultat possible une valeur numérique, et cela pour pouvoir exploiter dans la théorie des probabilités, la structure des nombres réels, comme les opérations algébriques.

7.1 Définitions

Une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{A}, \mathbf{P})$ est une application

$$X : \omega \to X(\omega)$$

qui à tout élément ω de Ω associe un nombre $X(\omega)$, telle que la condition suivante soit vérifiée.

L'image réciproque d'un intervalle de $\mathbb{R}$ ou d'une combinaison d'intervalles par les opérations sur les ensembles (union, intersection et complémentaire) est un évènement dont la probabilité est bien définie.

Si A est une partie de $\mathbb{R}$, et X est une variable aléatoire, on note par $(X \in A)$, l'ensemble des éléments de l'espace fondamentale Ω dont l'image est dans A (cet ensemble s'appelle l'image réciproque de A par X) :

$$(X \in A) = \{\omega \in \Omega : X(\omega) \in A\}, \tag{7.1}$$

et par conséquent, si l'ensemble $(X \in A)$ est un évènement,

$$P_r(X \in A) = P_r\{\omega \in \Omega : X(\omega) \in A\}. \tag{7.2}$$

En particulier, si a et b sont deux nombres réels $(a \leq b)$, alors par définition,

$$P_r(a \leq X \leq b) = P_r\{\omega \in \Omega : a \leq X(\omega) \leq b\}. \tag{7.3}$$

Si l'ensemble des images $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ est une partie discrète de $\mathbb{R}$, généralement $\mathbb{N}$ ou une partie de $\mathbb{N}$, la variable aléatoire X est dite discrète.

Une variable aléatoire X est dite finie, si l'ensemble des images $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ est une partie finie de $\mathbb{R}$, c'est-à-dire $X(\Omega)$ s'écrit sous la forme

$$X(\Omega) = \{x_1, ...x_p\},$$

où $x_1, ...x_p$ sont des nombres réels. Une variable aléatoire finie est évidemment discrète.

7.1.1 Exemple.

Considérons l'expérience du jet d'un dé. L'espace fondamental Ω lié à cette expérience est

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Pour $\omega = 1, 2, ..., 6$, On définit

$X(\omega)$ est le reste dans la division euclidienne de l'entier ω par l'entier 4.

On aura

$$X(1) = 1, X(2) = 2, X(3) = 3, X(4) = 0, X(5) = 1, X(6) = 2,$$

$$X(\Omega) = \{0, 1, 2, 3\},$$

$$(X = 2) = \{2, 6\},$$

$$\begin{aligned}(1 \leq X \leq 2) &= \{\omega \in \Omega : 1 \leq X(\omega) \leq 2\} \\ &= \{1, 2, 5, 6\},\end{aligned}$$

$$\begin{aligned}(1 < X \leq 2) &= (X = 2) \\ &= \{\omega \in \Omega : X(\omega) = 2\} \\ &= \{2, 6\},\end{aligned}$$

7.2 Variable aléatoire finie

On commence par la remarque importante suivante

Remarque 1 *Si X est une variable aléatoire finie ayant pour valeurs les nombres $x_1, x_2, ...x_n$, et $P_r(X = x_1) = p_1$, $P_r(X = x_2) = p_2, ..., P_r(X = x_n) = p_n$, alors on a*

$$p_1 + p_2 + ... + p_n = 1.$$

Pveuve. Puisque

$$\begin{aligned}\Omega &= X^{-1}(\mathbb{R}) = X^{-1}(\{x_1, x_2, ..., x_n\}) \\ &= X^{-1}(x_1) \cup X^{-1}(x_2) \cup ... \cup X^{-1}(x_n),\end{aligned}$$

alors

$$\begin{aligned}1 &= P_r(\Omega) \\ &= P_r\left(X^{-1}(x_1)\right) + P_r\left(X^{-1}(x_2)\right) + ... + P_r\left(X^{-1}(x_n)\right) \\ &= p_1 + p_2 + ... + p_n.\end{aligned}$$

■

Notation. On utilise aussi $X^{-1}(A)$, pour noter l'image réciproque par X d'une partie A de $\mathbb{R}$:

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}.$$

Représentation d'une variable aléatoire finie

On réprésente généralement une variable aléatoire X finie par un tableau de deux lignes, dans la première ligne on inscit les valeurs de X et dans la deuxième les probabilités correspondantes.

x_i	x_1	x_2	...	x_n	total
$P_r(X = x_i)$	p_1	p_2	...	p_n	1

Le tableau ci-dessus signifie que

1- X a n valeurs $x_1, ...x_n$, c'est-à-dire $X(\Omega) = \{x_1, ...x_n\}$, et

2- $P_r(X = x_1) = p_1$, $P_r(X = x_2) = p_2, ..., P_r(X = x_n) = p_n$. On dit aussi que ce tableau représente la loi de probabilité de la variable aléatoire X.

On réprésente également une variable aléatoire par un graphe " en bâtons" :

7.2.1 Exemple.

Dans l'exemple précédent (7.1.1), on a

$$X(\Omega) = \{0, 1, 2, 3\}$$

et

$$\begin{aligned}(X = 0) &= \{4\} \subset \Omega, \\ (X = 1) &= \{1, 5\} \subset \Omega, \\ (X = 2) &= \{2, 6\} \subset \Omega, \\ (X = 3) &= \{3\} \subset \Omega.\end{aligned}$$

Donc, puisque $\Omega = \{1, 2, 3, 4, 5, 6\}$ est un espace équiprobable, on a

$$\begin{aligned}P_r(X = 0) &= P_r(\{4\}) = \frac{1}{6}, \\ (X = 1) &= P_r(\{1, 5\}) = \frac{2}{6}, \\ (X = 2) &= P_r(\{2, 6\}) = \frac{2}{6}, \\ (X = 3) &= P_r(\{3\}) = \frac{1}{6}.\end{aligned}$$

Par conséquent, la variable aléatoire X est représentée par le tableau suivant

x_i	0	1	2	3	*total*
$P_r(X = x_i)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	1

7.2.2 Règle de calcul.

Dans le cas d'une variable aléatoire finie X, si A est une partie de $\mathbb{R}$, la probabilité $P_r(X \in A)$ est égale à la somme des probabilités des x_i qui sont dans A. Dans l'exemple précédent, on a

$$\begin{aligned}
P_r(1 \leq X \leq 2) &= P_r(X = 1) + P_r(X = 2) \\
&= \frac{2}{6} + \frac{2}{6} \\
&= \frac{2}{3},
\end{aligned}$$

$$\begin{aligned}
P_r(1 < X \leq 2) &= P_r(X = 2) \\
&= \frac{2}{6},
\end{aligned}$$

$$\begin{aligned}
P_r(X < 4) &= P_r(X = 0) + P_r(X = 1) + P_r(X = 2) + P_r(X = 3) \\
&= 1.
\end{aligned}$$

7.2.3 Exemple

Considérons l'expérience du jet sumultané de deux dés. L'ensemble fondamental Ω est l'ensemble des couple ordonnés (i, j) où i est le résultat du premier dé et j est le résultat du deuxième dé :

$$\Omega = \{(i, j) : 1 \leq i \leq 6,\ 1 \leq j \leq 6\},$$

et la probabilité P_r liée à cette expérience est définie, pour $A \subset \Omega$, par

$$P_r(A) = \frac{cardA}{36},$$

où $cardA$ désigne le nombre d'éléments de l'ensemble A.

Calculez la loi de probabilité de la variable aléatoire X sur Ω, qui à tout élément (i, j) de Ω associe la plus grande valeur des deux nombres i et j.

Solution. On doit calculer les valeurs possibles $X(\Omega)$ de X, et ensuite calculer la probabilité de chaque valeur.

Les valeurs possibles de X sont

$$X(\Omega) = \{1, 2, 3, 4, 5, 6\}.$$

On a, par définition, pour $1 \leq i \leq 6$

$$P_i = \frac{card\,(A_i)}{36},$$

où, A_i est l'ensemble des couples (p, r) tels que le plus grand est égale à i. Par exemple, A_1 est composé d'un seul élément $(1, 1)$, A_2 est composé de trois éléments $(1, 2)$, $(2, 1)$ et $(2, 2)$. Pour calculer facilement le nombre d'éléménts de chaque A_i, on dresse le tableau suivant, de sorte que dans la première colonne on met les valeurs de la première composante i, dans la première ligne on met les valeurs de la deuxième composante j, et dans la case correspondant à l'intersection de la ligne i et la colonne j, on met la plus grande valeur des deux nombres i et j.

i/j	1	2	3	4	5	6
1	1	2	3	3	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Le tableau ci-dessus, nous permet de construire rapidement la loi de la variable aléatoire X représentée par le tableau suivant

x_i	1	2	3	4	5	6	$\sum$
p_i	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	1

7.2.4 Exemple.

On met dans une urne 10 boules de même forme, 4 noires et 6 blanches. On tire de l'urne au hasard, et d'un seul coup, 6 boules. L'espace fondamentale Ω lié à cette expérience est l'ensemble des combinaisons de 6 boules parmi les 10 boules dans l'urne. C'est un espace équiprobable formé de $C_{10}^{6} = \frac{10!}{6!4!}$ éléments. Pour tout élément ω de Ω, on note par $X(\omega)$ le nombre de boules noires parmi les 6 boules de la combinaison ω.

1- Calculez la loi de la variable aléatoire X.

2- Calculez les probabilités

$P_r(X < 2)$, $P_r(2 \leq X \leq 4)$, $P_r(X < 0)$, $P_r(X \geq 4,2)$.

Solution 1- La loi de la variable aléatoire X est définie par le tableau

x_i	0	1	2	3	4	$\sum$
p_i	$\frac{C_6^6}{C_{10}^6} = \frac{1}{210}$	$\frac{C_6^5 C_4^1}{C_{10}^6} = \frac{24}{210}$	$\frac{C_6^4 C_4^2}{C_{10}^6} = \frac{90}{210}$	$\frac{C_6^3 C_4^3}{C_{10}^6} = \frac{80}{210}$	$\frac{C_6^2 C_4^4}{C_{10}^6} = \frac{15}{210}$	1

2-

$$\begin{aligned} P_r(X < 2) &= P_r(X = 0) + P_r(X = 1) \\ &= \frac{1}{210} + \frac{24}{210} = \frac{25}{210} = \frac{5}{42} \end{aligned}$$

$$\begin{aligned} P_r(2 \leq X \leq 4) &= P_r(X = 2) + P_r(X = 3) + P_r(X = 4) \\ &= \frac{90}{210} + \frac{80}{210} + \frac{15}{210} = \frac{185}{210} = \frac{37}{42} \end{aligned}$$

$$P_r(X < 0) = P_r(X \geq 4,2) = P_r(\emptyset) = 0$$

7.2.5 Espérance d'une variable aléatoire finie

L'espérance d'une variable aléatoire finie X, est le nombre noté $E(X)$ et défini par

$$E(X) = p_1 x_1 + p_2 x_2 + ...p_n x_n. \tag{7.4}$$

L'expression (7.4) s'écrit aussi d'une façon circoncise

$$E(X) = \sum_{i=1}^{n} p_i x_i. \tag{7.5}$$

Une façon pratique pour calculer l'espérence d'une variable aléatoire finie consiste en l'utilisation d'un tableau statistique. On applique ça à la variable alétoire finie de l'exemple **12.2.3**.

x_i	1	2	3	4	5	6	$\sum$
p_i	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	1
$p_i x_i$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{18}{36}$	$\frac{24}{36}$	$\frac{45}{36}$	$\frac{66}{36}$	$\frac{160}{36}$

On a donc

$$E(X) = \frac{160}{36} = \frac{40}{9}.$$

7.2.6 Variance d'une variable aléatoire finie

Soit X une variable aléatoire finie. La variance de X , noté $V(X)$ ou $\sigma^2(X)$, est définie par la formule

$$V(X) = \sum_{i=1}^{n} p_i \left(x_i - E(X)\right)^2. \tag{7.6}$$

Par exemple la variance de la variable aléatoire X de l'exemple **12.2.3** est égale à

$$\begin{aligned} V(X) &= \frac{1}{36}\left(1 - \frac{160}{36}\right)^2 + \frac{3}{36}\left(2 - \frac{160}{36}\right)^2 + \frac{6}{36}\left(3 - \frac{160}{36}\right)^2 \\ &\quad + \frac{6}{36}\left(4 - \frac{160}{36}\right)^2 + \frac{9}{36}\left(5 - \frac{160}{36}\right)^2 + \frac{11}{36}\left(6 - \frac{160}{36}\right)^2 \\ &= . \end{aligned}$$

7.2.7 L'écart type d'une variable aléatoire

L'écart type pour une variable aléatoire quelconque X, noté $\sigma(X)$, est la racine carrée de la variance $V(X)$:

$$\sigma(X) = \sqrt{V(X)}. \tag{7.7}$$

Remarque 1- Puisque la variance est une somme de carrés, elle est ≥ 0.

2- Puisque l'écart type est une racine carré d'un nombre ≥ 0, il est également ≥ 0.

7.2.8 Formule de Konig-Huygens.

Pour calculer la variance, au lieu d'utiliser directement la formule de la définition, on utilise généralement la formule suivante, appelée formule de Konig-Huygens.

$$V(X) = E\left(X^2\right) - E(X)^2 . \tag{7.8}$$

Cette formule donne une relation entre la variance et l'espérance. Elle signifie que la variance d'une variable aléatoire X est ègale à la différence entre l'espérence du carré X^2 de X et le carré de l'espérance de X. La formule de Konig-Huygens. est vérifiée pour toute variable aléatoire, non seulement pour les variables aléatoires finies.

Démonstration de la formule de Konig-Huygens

$$\begin{aligned} V(X) &= \sum_{i=1}^{n} p_i \left(x_i - E(X)\right)^2 \\ &= \sum_{i=1}^{n} p_i \left[x_i^2 + E(X)^2 - 2x_i E(X)\right] \\ &= \sum_{i=1}^{n} p_i x_i^2 + E(X)^2 \sum_{i=1}^{n} p_i - 2E(X) \sum_{i=1}^{n} p_i x_i \\ &= E\left(X^2\right) + E(X)^2 - 2E(X)^2 \\ &= E\left(X^2\right) - E(X)^2 . \end{aligned}$$

7.2.9 Tableau statistique

Dans le cas de variables aléatoires finies, parfois, il est plus commode de faire les calculs sur un tableau, qu'on appelle par abus de langage tableau statistique. Comme Illustration, on considére la variable X de l'exemple **12.2.3** et proposons nous de calculer les paramètres $E(X)$ et $V(X)$.

x_i	1	2	3	4	5	6	$\sum$
x_i^2	1	4	9	16	25	36	
p_i	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	1
$p_i x_i$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{18}{36}$	$\frac{24}{36}$	$\frac{45}{36}$	$\frac{66}{36}$	$\frac{160}{36}$
$p_i x_i^2$	$\frac{1}{36}$	$\frac{12}{36}$	$\frac{54}{36}$	$\frac{96}{36}$	$\frac{225}{36}$	$\frac{396}{36}$	$\frac{784}{36}$

Donc

$$E(X) = \sum_{i=1}^{6} p_i x_i = \frac{160}{36} = \frac{40}{9}$$

$$\begin{aligned} V(X) &= \sum_{i=1}^{6} p_i x_i^2 - \left(\sum_{i=1}^{6} p_i x_i \right)^2 \\ &= \frac{784}{36} - \left(\frac{40}{9} \right)^2 \\ &\cong 2,02 \end{aligned}$$

7.3 Variable aléatoire dénombrable

On a vu qu'une variable aléatoire X finie est définie par une suite finie $x_1, x_2, ..., x_n$ de nombres réels et une deuxième suite de nombres positifs $p_1, p_2, ..., p_n$ dont la somme est égale 1, les deux suites étant liées par la relation $P_r(X = x_i) = p_i$.

Une variable aléatoire dénombrable est définie de la même façon, seulement les deux suites de nombres ne sont pas finies.

7.3.1 Définition

Une variable aléatoire dénombrable X sur un espace probabilisé $(\Omega, \mathcal{A}, P_r)$ est définie par, une suite de nombres réels $x_1, x_2, ...$ et une suite $p_1, p_2, ...$ de nombres positifs dont la somme est égale à 1, les deux suites étant liées par la relation

$$P_r(X = x_i) = p_i,$$

pour tous les indices $i \in \mathbb{N}$. Contrairement au cas d'une variable aléatoire finie, on ne peut pas représenter une variable dénombrable par un tableau, car il s'agira d'une infinité de termes.

Remarque. Puisque $p_1, p_2, ...$ est une suite infinie de nombres, la somme, notée $\sum_{i=1}^{\infty} p_i$ est par définition la limite des sommes partielles, l'hypothèse que la somme est égale à 1 signifie donc

$$\lim_{n \to \infty} \sum_{i=1}^{n} p_i = 1, \tag{7.9}$$

et s'écrit

$$\sum_{i=1}^{\infty} p_i = 1. \tag{7.10}$$

7.3.2 Espérance d'une variable aléatoire dénombrable

L'espérance $E(X)$ d'une variable aléatoire dénombrable X est donnée sous la forme de limite :

$$\begin{aligned} E(X) &= \sum_{i=1}^{\infty} p_i x_i \qquad (7.11) \\ &= \lim_{n \to \infty} \sum_{i=1}^{n} p_i x_i. \end{aligned}$$

7.3.3 Variance d'une variable aléatoire dénombrable

La variance $V(X)$ d'une variable aléatoire dénombrable X est donnée sous la forme de limite :

$$\begin{aligned} V(X) &= \sum_{i=1}^{\infty} p_i (x_i - E(X))^2 \qquad (7.12) \\ &= \lim_{n \to \infty} \sum_{i=1}^{n} p_i (x_i - E(X))^2 . \end{aligned}$$

7.3.4 Formule de Konig-Huygens

La formule de Konig-Huygens pour une variable aléatoire dénombrable X s'écrit

$$\begin{aligned} V(X) &= E\left(X^2\right) - E(X)^2 \\ &= \sum_{i=1}^{\infty} p_i x_i^2 - \left(\sum_{i=1}^{\infty} p_i x_i\right)^2 . \end{aligned} \tag{7.13}$$

Exemple

On considère l'ensemble propabilisé $(\Omega, \mathcal{A}, P_r)$ où Ω est l'ensemble des entiers naturels $\mathbb{N}$, $\mathcal{A} = \mathcal{P}(\mathbb{N})$ et

$$P_r(X = n) = p_n = \frac{1}{2^n}$$

pour tout entier naturel non nulle n.

Puisque

$$\begin{aligned} \sum_{i=1}^{+\infty} p_i &= \sum_{\lim n \to +\infty} \frac{1}{2} + \frac{1}{2^2} + ... + \frac{1}{2^n} \\ &= \sum_{\lim n \to +\infty} \frac{1}{2} \frac{1 - \left(\frac{1}{2}\right)^n}{1 - \left(\frac{1}{2}\right)} \\ &= 1, \end{aligned}$$

La fonction identité $n \to X(n) = n$ est une variable aléatoire dénombrable.

7.4 Variable aléatoire définie par une densité

Fonction continue par morceaux

Dans ce cours, une fonction $f : \mathbb{R} \to \mathbb{R}$, est dite continue par morceaux si, il existe une subdivision de $\mathbb{R}$ en intervalles, tel que f est continue en chacun de ces intervalles (sauf éventuellement en ses extrémités).

Exemple 1.

La fonction

$$f(x) = \begin{cases} -1 \text{ si } x \leq 0, \\ 1 \text{ si } x > 0 \end{cases}$$

est continue par morceaux.

Exemple 2.

La fonction

$$g(x) = \begin{cases} 0 \text{ si } x < 0, \\ 1 \text{ si } 0 \leq x < 1, \\ \frac{1}{2} \text{ si } 1 \leq x < 2, \\ . \\ \frac{1}{n} \text{ si } n-1 \leq x < n, \\ . \end{cases}$$

est une fonction continue par morceaux.

7.4.1 Fonction densité

Une fonction $f : x \to f(x)$, définie de $\mathbb{R}$ dans $\mathbb{R}$, est une fonction densité si elle est continue par morceaux et vérifie les deux conditions

$$f(x) \geq 0, \tag{7.14}$$

pour tout $x \in \mathbb{R}$, et

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1. \tag{7.15}$$

Définition

Une variable aléatoire X sur un espace probabilisé est dite définie par une densité, s'il existe une fonction densité $f : x \to f(x)$, telle que

$$P_r(a \leq X \leq B) = \int_a^b f(x)\, dx, \tag{7.16}$$

pour tous nombres réels a et b où $a \leq b$ (a ou b ou les deux à la fois peuvent être infinis).

Propriété

Pour tout nombre réel x, on a

$$P_r(X = x) = \int_x^x f(t)\,dt = 0.$$

On en déduit que la probabilité d'un intervalle ne change pas si on ajoute ou on élimine une extrémité :

$$\begin{aligned} P_r(a \leq X \leq b) &= P_r(a < X \leq b) \\ &= P_r(a \leq X < b) \\ &= P_r(a < X < b). \end{aligned}$$

Remarque. Deux fonctions densité qui diffèrent uniquement en un nombre fini de points définissent la même loi de probabilité. Cela résulte du fait que l'intégrale de deux fonctions différentes seulement en un nombre fini de points est la même.

7.4.2 Exemples

Exemple 1.

On choisit deux nombres réels a et b, tels que $a < b$, et on définit une fonction $f(x)$ par

$$f(x) = \begin{cases} \frac{1}{b-a}, \text{ pour tout } a < x < b \\ \quad = 0, \text{ ailleurs.} \end{cases}$$

$f(x)$ est une fonction densité car elle vérifie les deux conditions

$$f(x) \geq 0 \text{ pour tout } x \in \mathbb{R}$$

et

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)\,dx &= \int_a^b f(x)\,dx \\ &= \frac{(b-a)}{(b-a)} \\ &= 1. \end{aligned}$$

Exemple 2.
On pose

$$g(x) = \begin{cases} \frac{x}{2}, \text{ pour tout } 0 \leq x \leq 1 \\ \frac{3(x-2)}{2} \text{ pour tout } 2 \leq x \leq 3 \\ 0 \text{ ailleurs.} \end{cases}$$

On vérifie facilement que $g(x)$ est une densité. En effet, il est évident que $g(x) \geq 0$ pour tout $x \in \mathbb{R}$. En plus

$$\begin{aligned} \int_{-\infty}^{+\infty} g(x)\, dx &= \int_0^1 \frac{x}{2} dx + \int_2^3 \frac{3(x-2)}{2} dx \\ &= \left[\frac{x^2}{4}\right]_0^1 + \frac{3}{4}\left[(x-2)^2\right]_2^3 \\ &= \left(\frac{1}{4} - 0\right) + \frac{3}{4}(1-0) \\ &= 1. \end{aligned}$$

7.4.3 Espérance, variance et écart type d'une variable aléatoire définie par une densité

Définition

Si X est une variable aléatoire définie par une densité $f(x)$,
1- L'espérance $E(X)$ de X est définie par

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, dx. \tag{7.17}$$

2- La variance $V(X)$ de X est définie par

$$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)\, dx. \tag{7.18}$$

3- Comme pour les variables discrètes, l'écart type $\sigma(X)$ d'une variable aléatoire continue X est égal à la racine carré de la variance $V(X)$:

$$\sigma(X) = \sqrt{V(X)} \tag{7.19}$$

La formule de Konig-Huygens, démontrée pour les variables aléatoires finie est aussi valable pour les variables aléatoires continues :

7.4.4 Formule de Konig-Huygens

Si X est une variable aléatoire continue de densité $f(x)$, alors

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \left(\int_{-\infty}^{+\infty} x f(x)\,dx\right)^2 \\ &= E\left(X^2\right) - E(X)^2 . \end{aligned} \tag{7.20}$$

Pveuve.

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)\,dx \\ &= \int_{-\infty}^{+\infty} \left(x^2 + E(X)^2 - 2xE(X)\right) f(x)\,dx \\ &= \int_{-\infty}^{+\infty} x^2 f(x)\,dx + E(X)^2 \int_{-\infty}^{+\infty} f(x)\,dx - 2E(X) \int_{-\infty}^{+\infty} x f(x)\,dx \\ &= \int_{-\infty}^{+\infty} x^2 f(x)\,dx + E(X)^2 - 2E(X)^2 \\ &= E\left(X^2\right) - E(X)^2 . \end{aligned}$$

■

Exemple

Une machine produit des pièces qui sont utilisées dans la fabrication des tondeuses à

gazon. La longueur de ces pièces, en mm, varie d'une pièce à l'autre selon la densité de

probabilité suivante :

$$f(x) = \begin{cases} c(4-x)(x-2) & \text{si } 2 \leq x \leq 4 \\ 0 \text{ ailleurs.} \end{cases}$$

1. Trouver la valeur de la constante c pour laquelle la fonction $f(x)$ est une densité de

probabilité. Tracer son graphe.
2. Déterminer la fonction de répartition.
3. Déterminer l'espérance et la variance.

Solution

1. On doit avoir

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

On a

$$\begin{aligned}
\int_{-\infty}^{+\infty} f(x)\,dx &= \int_{-\infty}^{2} 0dx + \int_{2}^{4} c(4-x)(x-2)\,dx + \int_{4}^{+\infty} 0dx \\
&= 0 + c\left[-\frac{x^3}{3} + 3x^2 - 8x\right]_2^4 + 0 \\
&= c\frac{4}{3},
\end{aligned}$$

d'où

$$c = \frac{3}{4}.$$

2. la fonction de répartition $F(x)$ est définie pour tout $x \in \mathbb{R}$, par

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

Les différentes expressions de $f(t)$ donnent

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0dt = 0 \text{ si } -\infty < x < 2, \\ \int_{-\infty}^{x} \frac{3}{4}(4-t)(t-2)\,dt = \frac{3}{4}\left(-\frac{x^3}{3} + 3x^2 - 8x\right) \text{ si } 2 \leq x < 4, \\ 1 \text{ si } 4 \leq x < +\infty. \end{cases}$$

3.

$$\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} xf(x)\,dx \\
&= \int_{-\infty}^{2} 0dx + \int_{2}^{4} \frac{3}{4}x(4-x)(x-2)\,dx + \int_{4}^{+\infty} 0dx \\
&= 3.
\end{aligned}$$

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} x^2 f(x)\,dx - E(X)^2 \\ &= \int_2^4 \frac{3}{4} x^2 (4-x)(x-2)\,dx - 3^2 \\ &= 9.2 - 9 \\ &= 0.2. \end{aligned}$$

7.5 Fonction de répartition d'une variable aléatoire

La loi de probabilité d'une variable aléatoire est donnée dans le cas discret par une suite de nombres non négatifs dont la somme est égale à un, et dans le cas continu par une fonction définie sur l'ensemble des réels $\mathbb{R}$, à valeurs non négatifs, dont l'intégrale sur $\mathbb{R}$ est égale un. On peut également définir la loi de probabilité d'une variable aléatoire discrète ou continue par une fonction, qui cumule les probabilités, appelée fonction de répartition. En anglais cette fonction s'appelle cumulative distribution function.

Définition

La fonction de répartition d'une variable aléatoire (discrète ou continue) X est une fonction notée F définie pour tout nombre réel x par

$$F(x) = P_r(X \leq x). \tag{7.21}$$

Propriété d'une fonction de répartition F

1. $0 \leq F(x) \leq 1$ pour tout $x \in \mathbb{R}$, car

$$\begin{aligned} F(x) &= P_r(X \leq x) \\ &\leq P_r(X < +\infty) \\ &= 1. \end{aligned}$$

On admet que $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ et $\lim_{x \to +\infty} F(x) = 1$.

2. F est une fonction croissante sur $\mathbb{R}$. En effet, on a pour tous $y \geq x$,

$$\begin{aligned} F(y) &= P_r(X \leq y) \\ &= P_r(X \leq x \text{ ou } x < X \leq y) \\ &= P_r(X \leq x) + P_r(\ x < X \leq y) \ \text{(car les deux évènements sont incompatibles)} \\ &= F(x) + P_r(\ x < X \leq y). \end{aligned}$$

D'ou

$$F(y) \geq F(x),$$

puisque $P_r(\ x < X \leq y) \geq 0$.

7.5.1 Fonction de répartition et calcul de probabilité.

Si a et b sont deux nombres et $a < b$, alors

$$P_r(a < X \leq b) = F(b) - F(a). \tag{7.22}$$

En effet,

$$\begin{aligned} F(b) &= P_r(X \leq a) + P_r(a < X \leq b) \\ &= F(a) + P_r(a < X \leq b). \end{aligned}$$

Cas d'une variable aléatoire finie

Soit X une variable aléatoire finie dont la loi est définie par le tableau

x_i	x_1	x_2	...	x_n
$P_r(X = x_i)$	p_1	p_2	...	p_n

On a

1.

La fonction de répartition de X est totalement déterminée par ses valeurs $F(x_1) = p$

$$\begin{aligned} F(x_1) &= p_1, \\ F(x_2) &= p_1 + p_2, \\ &\ldots \\ F(x_i) &= p_1 + ... + p_i, \\ &\ldots \\ F(x_n) &= p_1 + ... + p_n = 1. \end{aligned}$$

2. La fonction de répartition de la variable aléatoire finie X est décrite, en précisant sa valeur pour tous $x \in \mathbb{R}$:

$$F(x) = \begin{cases} 0 & \text{si } -\infty < x < x_1 \\ \mathrm{p}_1 & \text{si } \mathrm{x}_1 \leq x < x_2 \\ \mathrm{p}_1 + p_2 & \text{si } \mathrm{x}_2 \leq x < x_3 \\ \ldots & \ldots \\ \mathrm{p}_1 + \ldots p_n & \text{si } \mathrm{x}_{n-1} \leq x < x_n \\ 1 & \text{si } \mathrm{x}_n \leq x < +\infty \end{cases}$$

3. La fonction de répartition F de X est totalement déterminée par ses valeurs $F(x_1), F(x_2), \ldots, F(x_n)$. On peut donc la représenter par un tableau qui précise ces valeurs :

x_i	x_1	x_2	...	x_i	..	x_n
$F(x_i) = P_r(X \leq x_i)$	$F(x_1) = p_1$	$F(x_2) = p_1 + p_2$	...	$F(x_i) = p_1 + \ldots + p_i$	...	1

2.

3. On remarque que la fonction de répartition $F(x)$ de la variable aléatoire finie X est une fonction en escaliers : elle est nulle sur l'intervalle $]-\infty, x_1[$, égale la somme $p_1 + p_2 + \ldots + p_i$ sur chaque intervalle $[x_i, x_{i+1}[$ $(i = 1, \ldots n-1)$, et elle est égale à 1 sur l'intervalle $[x_n, +\infty[$.

4. Le graphe de $F(x)$ est représenté par Fig. On détecte directement sur le graphe les valeurs de X ayant une probabilité non nulle. Ce sont les points sur l'axe des abscisses où il y a un saut, c'est-à-dire une discontinuité du graphe, le probabilité en un tel point est égale à la longueur du saut.

5. Si X est une variable aléatoire discrète ayant n valeurs $x_1, x_2, \ldots x_n$, alors les probabilités poids $p_1 = P_r(X = x_1), p_2 = P_r(X = x_2), \ldots, p_n =$

$P_r(X = x_n)$ s'obtiennent des valeurs $F_1 = F(x_1), F_2 = F(x_2), ..., F_n = F(x_n)$ de la fonction de répartition par les deux relations suivantes

$$\begin{aligned} p_1 &= F_1 \\ p_i &= F_i - F_{i-1} \text{ pour } i = 2, ..., n \end{aligned} \tag{7.23}$$

7.5.2 Exemple

On prend une pièce de monnaie on inscrit sur une face le numéro zéro et sur l'autre face on inscrit un. On jette les deux pièces en même temps et on vérifie le résultat de chacune des deux pièces 1 ou 0. On définit la variable aléatoire X comme étant la somme des résultats des deux pièces. Les valeurs possibles de X sont $0, 1$ et 2.

x_i	0	1	2	$\sum$
p_i	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1
$F(x_i)$ $= P_r(X \leq x_i)$	$\frac{1}{4}$	$\frac{3}{4}$	1	

7.5.3 Exemple

Soit X une variable aléatoire discrète dont la loi de probabilité est définie par sa fonction de répartition $F(x)$ et représentée dans le tableau

x_i	-2	0	$1,5$	3	$5,6$	9	$9,5$	10	
$F_i = F(x_i)$	$0,01$	$0,2$	$0,4$	$0,5$	$0,55$	$0,7$	$0,75$	1	

1) Complétez le tableau en ajoutant les poids $p_1 = P_r(X = x_1), p_2 = P_r(X = x_2), ..., p_8 = P_r(X = x_8)$.

2) Calculez l'espérance $E(X)$ et la variance $\sigma^2(X)$ et l'écart-type $\sigma(X)$.

Réponse. 1) On complète le tableau en utilisant les relations

x_i	-2	0	$1,5$	3	$5,6$	9	$9,5$	10	$\sum$
$F_i = F(x_i)$	$0,01$	$0,2$	$0,4$	$0,5$	$0,55$	$0,7$	$0,78$	1	
$p_i = P_r(X = x_i)$	$0,01$	$0,19$	$0,2$	$0,1$	$0,05$	$0,15$	$0,08$	$0,22$	1
$p_i x_i$	$-0,02$	0	$0,3$	$0,3$	$0,28$	$1,35$	$0,76$	$2,2$	$5,17$
$p_i x_i^2$	$0,04$	0	$0,45$	$0,9$	$1,568$	$12,15$	$7,22$	22	$44,328$

2)

$$E(X) = \sum_{i=1}^{8} p_i x_i = 5.17$$

$$\begin{aligned}
\sigma^2(X) &= \sum_{i=1}^{8} p_i x_i^2 - \left(\sum_{i=1}^{8} p_i x_i \right)^2 \\
&= 44,328 - (5.17)^2 \\
&= 44,328 - 26,7289 \\
&= 17,5991
\end{aligned}$$

$$\sigma(X) = \sqrt{17,5991}$$

7.5.4 Cas d'une variable aléatoire dénombrable (non finie)

1. La fonction de répartition $F(x)$ d'une variable aléatoire dénombrable X définie par deux suites (x_i) et (p_i) est, comme pour une variable finie, une fonction en escaliers. En effet $F(x)$ est définie par

$$F(x) = \begin{cases} 0 & \text{si } -\infty < x < x_1 \\ \mathrm{p}_1 & \text{si } \mathrm{x}_1 \leq x < x_2 \\ \mathrm{p}_1 + p_2 & \text{si } \mathrm{x}_2 \leq x < x_3 \\ \ldots & \ldots \\ \mathrm{p}_1 + \ldots + p_n & \text{si } \mathrm{x}_{n-1} \leq x < x_n \end{cases} .$$

Exemple

Considérons la variable aléatoire dénombrable X, de l'exemple **12.3.4.**
La fonction de répartition $F(x)$ est définie par le tableau

$$F(x) = \begin{cases} 0 & \text{si } -\infty < x < 1 \\ \frac{1}{2} & \text{si } 1 \leq x < 2 \\ \frac{3}{4} & \text{si } x_2 \leq x < x_3 \\ ... & ... \\ 1 - \left(\frac{1}{2}\right)^{n-1} & \text{si } x_{n-1} \leq x < x_n. \\ ... & ... \end{cases}.$$

7.5.5 Cas d'une variable aléatoire définie par une densité

La fonction de répartition $F(x)$ d'une variable aléatoire X de densité $f(x)$ est une fonction définie pour tout $x \in \mathbb{R}$ par

$$F(x) = P_r(X \leq x) = \int_{-\infty}^{x} f(t)\, dt. \tag{7.24}$$

Probabilité d'un intervalle de $\mathbb{R}$. Si $a < b$ sont deux nombres réels, alors on a

$$\begin{aligned} P_r(a \leq X \leq b) &= P_r(a \leq X < b) \qquad (7.25) \\ &= P_r(a < X \leq b) \\ &= F(b) - F(a). \end{aligned}$$

Car, pour une variable aléatoire continue, la probabilité d'un point est nulle.

Exemple.

Reprenons les données de l'exemple **12.5.3** et posons la question suivante :
Les pièces de longueur inférieure à 2.2 mm et celles de longueur supérieure à 3.8 mm

sont inutilisables. Calculer la proportion de pièces inutilisables produites par cette

machine.

Solution.

4. Les pièces inutilisables sont reparties sur deux évènements incompatibles correspondants à $(X < 2.2)$ et $(X > 3.8)$.

La proportion de l'union de ces deux évènements est la probabilité de cette union :

$$
\begin{aligned}
& P_r\left((X<2.2)\cup(X>3.8)\right) \\
= {} & 1-P_r\left(\overline{(X<2.2)\cup(X>3.8)}\right) \\
= {} & 1-P_r\left(\overline{(X<2.2)}\cap\overline{(X>3.8)}\right) \quad \text{(formule de Morgan)} \\
= {} & 1-P_r\left((X\geq 2.2)\cap(X\leq 3.8)\right) \\
= {} & 1-P_r\left(2.2\leq X\leq 3.8\right) \\
= {} & 1-\left(F(3.8)-F(2.2)\right) \\
= {} & 1-(0.972-0.028) \\
= {} & 0.056 \\
= {} & 5.6\%.
\end{aligned}
$$

Caractéristiques d'une fonction de répartition d'une V.A continue

Pour vérifier qu'une fonction $F(x)$ définie sur $\mathbb{R}$ est une fonction de répartition d'une variable aléatoire continue, on doit vérifier les propriétés suivantes.

a) $F(x)$ est continue sur $\mathbb{R}$, et dérivable sur $\mathbb{R}$, (sauf peut-être en un nombre fini de points).

b) $F(x)$ est croissante.

c) $\lim\limits_{x\to-\infty} F(x)=0$.

d) $\lim\limits_{x\to+\infty} F(x)=1$.

Relation entre la fonction densité et la fonction de répartition

Si $F(x)$ et $f(x)$ sont respectivement, la fonction de répartition et la fonction densité d'une variable aléatoire continue X, alors on a (sauf peut être en un nombre fini de points),

$$f(x) = \frac{dF}{dx}(x). \tag{7.26}$$

Exemple

On définit sur $\mathbb{R}$ la fonction $F(x)$ par

$$F(x) = \begin{cases} 0 \text{ si } x \leq Ln2 \\ 1 + \frac{1}{1-e^x} \text{ si } x > Ln2. \end{cases}$$

1) Montrer que $F(x)$ est une fonction de répartition d'une variable aléatoire continue X.

2) Trouvez la fonction densité correspondante.

3) Calculez les probabilités $P_r(X \leq Ln2)$; $P_r(X \geq Ln10)$; $P_r(Ln1 \leq X \leq Ln10)$.

Solution. 1)

a) Pour la continuité de $F(x)$, il suffit de montrer la continuité au point $x = Ln2$, car pour le reste des points c'est évident.

On a

$$\lim_{\substack{x \to Ln2 \\ x > Ln2}} F(x) = \left[1 + \frac{1}{1 - e^{Ln2}}\right] = 0.$$

Il est également évident F est dérivable en tout point x différent de $Ln2$.

b) Pour la croissance de F, on utilise la continuité de $F(x)$ sur tout $\mathbb{R}$, et la dérivée

$$\frac{dF}{dx}(x) = \begin{cases} 0 \text{ pour } x < 2 \\ \frac{e^x}{(1-e^x)^2} \text{ pour } x > 2. \end{cases}$$

$\frac{dF}{dx}(x)$ est positive sur les intervalles $]-\infty, Ln2[$ et $]Ln2, +\infty[$ et $F(x)$ est continue sur tout $\mathbb{R}$, par conséquent elle est croissante sur $\mathbb{R}$.

Pour les deux dernières propriétés, c'est évident.

2)

$$f(x) = \frac{dF}{dx}(x) = \begin{cases} 0 \text{ pour } x < Ln2 \\ \frac{e^x}{(1-e^x)^2} \text{ pour } x > Ln2. \end{cases}$$

On n'a pas défini $f(x)$ au point $x = Ln2$, car on peut lui donner n'importe qu'elle valeur.

3)

$$\begin{aligned} P_r(X \leq Ln2) &= F(Ln2) - F(-\infty) \\ &= 0 - 0 \\ &= 0 \end{aligned}$$

$$\begin{aligned} P_r(X \geq Ln10) &= F(Ln(+\infty)) - F(Ln(10)) \\ &= 1 - \left[1 + \frac{1}{1 - e^{Ln(10)}}\right] \\ &= 1 - \left[1 + \frac{1}{1-10}\right] \\ &= \frac{1}{9} \end{aligned}$$

$$\begin{aligned} P_r(Ln1 \leq X \leq Ln10) &= F(Ln10) - F(Ln1) \\ &= F(Ln10) - 0 \\ &= \left[1 + \frac{1}{1 - e^{Ln10}}\right] \\ &= 1 + \frac{1}{1-10} \\ &= \frac{9}{10}. \end{aligned}$$

Quantile d'ordre α d'une variable aléatoire X

Définition.

Un nombre x est un quantile d'ordre $0 \leq \alpha \leq 1$, d'une variable aléatoire X , de fonction de répartition F, si

$$F(x) = \alpha.$$

Donc la notion de quantile correspond à l'inverse de la fonction de répartition.

Exemple.

Soit la variable aléatoire discrète X définie par le tableau

x_i	-2	1	4	8
p_i	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{3}{10}$

Pour calculer les quantiles on ajoute au tableau une ligne pour la fonction de répartition :

x_i	-2	1	4	
p_i	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	
F_i	$\frac{2}{5}$	$\frac{3}{5}$	1	

Dans cet exemple, tout nombre strictement inférieur à -2 est un quantile de zéro, par exemple, -2.5 est un quantile de zéro :

$$x_0 < -2.$$

Pour les autres quantiles, on a

$$-2 \leq x_{\frac{2}{5}} < 1, \;\; 1 \leq x_{\frac{3}{5}} < 4, \;\; 4 \leq x_1 < +\infty.$$

Unicité

On a remarqué que pour α donné, le quantile x_α n'est pas toujours unique. On fait, le quantile x_α de α est unique, seulement si la fonction de répartition F est injective. Dans ce cas, le quantile x_α de α est l'image réciproque de α par la fonction de répartition F.

Exemple

On définit une fonction f sur $\mathbb{R}$ de la façon suivante

$$f(x) = \begin{cases} \frac{1}{5x^2} \text{ pour } x < -1 \text{ ou } x > 1 \\ \frac{1}{10} \text{ pour } -1 \leq x \leq 1. \end{cases}$$

1. Montrer que $f(x)$ est une densité.
2. Calculer la fonction de répartition $F(x)$.
3. Montrer que $F(x)$ est strictement croissante.
4. Calculer les quantiles $x_{0.5}$ et $x_{0.3}$.

Solution.

1.

$$\begin{aligned}
\int_{-\infty}^{+\infty} f(x)\,dx &= \frac{1}{5}\int_{-\infty}^{-1}\frac{1}{x^2}dx + \int_{-1}^{1}\frac{1}{10}dx + \frac{1}{5}\int_{1}^{+\infty}\frac{1}{x^2}dx \\
&= \frac{1}{5}\left[-\frac{2}{x}\right]_{-\infty}^{-1} + \left[\frac{x}{10}\right]_{-1}^{1} + \frac{1}{5}\left[-\frac{2}{x}\right]_{1}^{+\infty} \\
&= \frac{2}{5} + \frac{1}{5} + \frac{2}{5} = \mathbf{1}.
\end{aligned}$$

2. Par définition, pour tout $x \in \mathbb{R}$,

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

Donc,

pour $x < -1$,

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x}\frac{1}{t^2}dt \\
&= \left[-\frac{2}{t}\right]_{-\infty}^{x} \\
&= \frac{-2}{x} - \frac{2}{-\infty} \\
&= \frac{-2}{x},
\end{aligned}$$

pour $-1 \leq x \leq 1$,

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(t)\,dt \\
&= \frac{1}{5}\int_{-\infty}^{-1} \frac{1}{t^2}dt + \int_{-1}^{-x} \frac{1}{10}dt \\
&= \frac{1}{5}\left[-\frac{2}{t}\right]_{-\infty}^{-1} + \left[\frac{t}{10}\right]_{-1}^{x} \\
&= \frac{2}{5} + \frac{x}{10} + \frac{1}{10} \\
&= \frac{x}{10} + \frac{1}{2},
\end{aligned}$$

et, pour $1 < x < +\infty$,

$$\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(t)\,dt \\
&= \frac{1}{5}\int_{-\infty}^{-1} \frac{1}{t^2}dt + \int_{-1}^{-1} \frac{1}{10}dt + \frac{1}{5}\int_{1}^{+\infty} \frac{1}{t^2}dt \\
&= \frac{1}{5}\left[-\frac{2}{t}\right]_{-\infty}^{-1} + \left[\frac{t}{10}\right]_{-1}^{1} + \frac{1}{5}\int_{1}^{+\infty} \frac{1}{t^2}dt \\
&= \frac{2}{5} + \frac{2}{10} + \frac{1}{5}\left[-\frac{2}{t}\right]_{1}^{x} \\
&= \frac{3}{5} + \frac{1}{5}\left(\frac{-2}{x} + \frac{2}{1}\right) \\
&= 1 - \frac{2}{5x}.
\end{aligned}$$

7.6 Exemples de lois discrètes

7.6.1 Loi binomiale

Loi de Bernoulli

Considérons une expérience ayant deux résultats possibles l'un codé par 0 et l'autre par 1, généralement l'évènement codé par 1 s'appelle succès et l'évènement codé par 0 s'appelle échec. La variable aléatoire X associe à cette expérience s'appelle variable aléatoire ou loi, de Bernoulli, si on note

$$P_r(X=1)=p, \tag{7.27}$$

alors,

$$P_r(X=0)=1-p, \tag{7.28}$$

car les deux évènements $(X=1)$ et $(X=0)$ sont deux évènements complémentaires.

L'espérance $E(X)$ d'une variable X de Bernoulli est

$$E(X)=p, \tag{7.29}$$

Sa variance $V(X)$ est

$$\begin{aligned} V(X) &= (1-p)^2+(0-p)^2 \\ &= p(1-p) \end{aligned} \tag{7.30}$$

et son écart type $\rho(X)$ est

$$\sigma(X)=\sqrt{p(1-p)}. \tag{7.31}$$

Définition d'une Loi binomiale

On considère une expérience de Bernoulli, ayant donc deux résultats possibles 1 ou 0, et on répète l'expérience un nombre fini n fois, dans les mêmes conditions. Le résultat possible aprés la réalisation de ces n expériences de Bernoulli est une suite de nombres $(a_1, a_2, ..., a_n)$, chacun de ces nombres étant égal à 0 ou 1.

On définit la variable aléatoire X comme étant le nombre de fois où l'expérience donne le résultat 1. Les valeurs possibles de la variable aléatoire X sont les entiers de 0 à n.

Une loi binomiale, noté $B(n,p)$, est déterminée par deux paramètres l'entier n égale au nombre de répétition de l'expérience de Bernoulli et le nombre $0 \leq p \leq 1$ désignant la probabilité du résultat positif. L'écriture

$$X \sim B(n,p)$$

signifie que la loi de probabilité du variable aléatoire X est la loi binomiale $B(n,p)$.

Caractéristiques d'une loi binomiale

On pense à une loi binomiale $B(n,p)$ lorsque les conditions suivantes sont vérifiées.

1) Il s'agit d'une expérience qui a deux résultats possibles, succès avec une probabilité p et échec avec la probabilité $1-p$ (donc expérience de Bernoulli).

2) Cette expérience se répète un nombre n fini de fois.

3) Cette expérience se répète dans les mêmes conditions (on dit qu'il y a indépendance).

Formule de la loi binomiale

Les valeurs possibles d'une variable aléatoire X gouvernée par une loi binomiale $B(n,p)$ sont les entiers de 0 à n.

et on a pour tout entier $0 \leq k \leq n$,

$$\begin{aligned} P_r(X=k) &= C_n^k p^k (1-p)^{n-k} \qquad (7.32) \\ &= \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}. \end{aligned}$$

On peut donc représenter ce variable aléatoire finie par le tableau

x_k	0	1	...	k	...	n	$\sum$
$p_k = P_r(X=k)$	$(1-p)^n$	$p(1-p)^{n-1}$	...	$C_n^k p^k (1-p)^{n-k}$	...	p^n	1

Espérance $E(X)$, variance $V(X)$ et écart type $\sigma(X)$. Si $X \sim B(n,p)$ est une loi binomiale, on prouve que

$$E(X) = np, \qquad (7.33)$$

$$V(X) = np(1-p) \qquad (7.34)$$

et

$$\sigma(X) = \sqrt{np(1-p)}. \qquad (7.35)$$

Démonstration de la formule de la loi binomiale On démontre la formule (7.32) de la loi binomiale par étapes.

Le cas $n = 1$. Le cas $n = 1$ signifie qu'on a réalisé l'expérience de Bernoulli une seule fois, donc la loi $B(1, p)$ coïncide avec la loi de Bernoulli.

Le cas $n = 2$. Dans ce cas les résultats possibles de l'expérience sont $(1,0)$, $(1,1)$, $(0,0)$, $(0,1)$. Les probabilités de ces 4 évènements se calculent comme suit

$$\begin{aligned} P_r(1,0) &= P_r(a_1 = 1 \text{ et } a_2 = 0) \\ &= P_r((a_1 = 1) \cap (a_2 = 0)) \\ &= P_r(a_1 = 1) . P_r(a_1 = 0) \text{ (puisque les deux évènements sont indépendants).} \\ &= p(1-p) . \end{aligned}$$

De la même façon on trouve

$$\begin{aligned} P_r(1,1) &= p^2 \\ P_r(0,0) &= (1-p)^2 \\ P_r(0,1) &= (1-p)p. \end{aligned}$$

On a donc

$$\begin{aligned} P_r(X = 0) &= P_r(0,0) = 0 \\ P_r(X = 1) &= P_r((1,0) \cup (0,1)) \\ &= p(1-p) + (1-p)p \text{ (car les deux évènements sont incompatibles).} \\ &= 2p(1-p) . \\ P_r(X = 2) &= P_r(1,1) = p^2. \end{aligned}$$

Ainsi la loi de X est définie dans ce cas par le tableau

x_i	0	1	2	$\sum$
$p_i = P_r(X = x_i)$	$(1-p)^2$	$2p(1-p)$	p^2	1

Le cas général $n \geq 1$ On considère une partie de l'ensemble des numéros de 1 à n, contenant exactement r numéros ($r \leq n$), $j_1, j_2, ...j_r$.

On prouve que La probabilité de l'évènement où, parmi les n expériences de Bernoulli, les r expériences $j_1, j_2, ...j_r$ donnent chacune 1 (chacune des autres $n - r$ expériences donne forcément 0) est égale à

$$p^r(1-p)^{n-r} .$$

L'évènement $(X = r)$ est l'union des évènements qu'on vient de décrire. Le nombres de ces évènements est égal à C_n^r et ils sont deux à deux incompatibles, d'où finalement

$$P(X = r) = C_n^r p^r (1-p)^{n-r}.$$

Remarque Pour prouver que la somme des probabilités est égale à 1, on utilise la formule du binôme :

$$\begin{aligned} 1 &= (p + (1-p))^n \qquad (7.36) \\ &= \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k}. \end{aligned}$$

Exemple

on met dans une urne 10 boules de même dimension, 3 de couleur blanche et 7 de couleur noire. On tire au hasard une boule, on constate sa couleur et on la remet dans l'urne, on répète l'opération 6 fois pour avoir au total fait 6 tirages. Dans chacun des six tirages, la probabilité que la boule tirée soit de couleur noire est $p = \frac{7}{10}$, et la variable aléatoire X, désignant le nombre de boules noires parmi les 6 boules tirées, suit la loi binomiale $B\left(6; \frac{7}{10}\right)$:

$$P_r(X = k) = \frac{6!}{k!(6-k)!}(0.7)^k (0.3)^{6-k}, \text{ pour } k = 0, ..., 6.$$

On se propose de calculer les probabilités $P_r(X = 0)$, $P_r(X < 0)$, $P_r(X = 5)$, $P_r(X = \text{ au moins } 5)$, $P_r(X = \text{ au plus } 2)$ et $P_r(X \neq 1)$.

Solution

$$
\begin{aligned}
P_r(X=0) &= (0.3)^6 . \\
P_r(X<0) &= P_r(\emptyset) = 0. \\
P_r(X=5) &= \frac{6!}{5!1!}(0.7)^5(0.3)^1 = . \\
P_r(X = \text{ au moins } 5) &= P_r(X \geq 5) \\
&= P_r(X = 5 \text{ ou } 6) \\
&= P_r(X=5) + P_r(X = \ 6) \\
&= \frac{6!}{5!1!}(0.7)^5(0.3) + \frac{6!}{5!1!}(0.7)^5(0.3) \\
&= . \\
P_r(X = \text{ au plus } 2) &= P_r(X \leq 2) \\
&= P_r(X = 0 \text{ ou } 1 \text{ ou } 2\ ) \\
&= P_r(X=0\ ) + P_r(X=1) + P_r(X=2) \\
&= . \\
P_r(X \neq 1) &= 1 - P_r(X=1), \text{ car les évènements } (X \neq 1) \text{ et } (X=1) \\
&\quad \text{sont complémentaires.} \\
&= 1 - \frac{6!}{1!5!}(0.7)(0.3)^5 \\
&= .
\end{aligned}
$$

Remarque Dans l'expérience de l'exemple précédent, après chaque tirage, avant de faire le tirage suivant, on remet la boule tirée dans l'urne : On dit qu'il s'agit de tirage avec remise. Si après chaque tirage, on ne remet pas la boule tirée dans l'urne, on dit qu'il s'agit de tirage sans remise.

Dans le cas de tirage avec remise, les résultats des six tirages sont indépendants, c'est-à-dire le résultat d'un tirage ne dépend pas du résultat du tirage précédent, car la configuration de l'urne est la même pour chacun des six tirages. C'est pour cette raison que la loi définie dans l'exemple précédent est une loi binomiale.

Dans le cas de tirages sans remise la loi binomiale ne s'applique pas car les résultats des tirages successifs ne sont pas indépendants.

Exemple

Dans une ville 20 pour cent de la population sont des universitaires, on prend au hasard 10 personnes de cette ville et on désigne par X le nombre d'universitaires parmi ces 10 personnes. Calculez la loi de la variable aléatoire X ainsi définie.

Solution. L'expérience qu'on considère ici est le choix au hasard d'une personne de cette ville. La probabilité que la personne tirée soit universitaire est $P = 0,20$ et la probabilité que la personne choisie ne soit pas universitaire est $1 - p = 0,80$. Pour choisir 10 personnes, on imagine qu'on a répété l'expérience 10 fois ($n = 10$). En plus, on peut supposer que le résultat d'une expérience ne dépend pas du résultat de l'expérience précédente, car le pourcentage des universitaires reste presque le même (il y a donc indépendance). En conclusion on peut dire que la variable aléatoire X suit la loi normale $B(10; 0,20)$.

7.6.2 Loi de Poisson

Définition

Une loi de Poisson de moyenne $\lambda > 0$ est une loi discrète dénombrable définie par

$$P_r(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \text{ pour } k = 0, 1, 2, ... \tag{7.37}$$

Les mathématiciens avaient montré que la loi de Poisson convient pour une variable aléatoire décrivant le nombre d'évènements indépendants qui surviennent en une durée limitée de temps et en petites quantités, par exemple, le nombre de tremblements de terre, le nombres d'interventions d'un médecin de garde ou le nombres de catastrophes naturelles.

Remarque Une loi de Poisson est définie par un seul paramètre qui est l'espérance (ou en langage statistique la moyenne) λ.

Espérance $E(X)$, variance $V(X)$ et écart type $\sigma(X)$.

Si $X \sim P(\lambda)$ est une variable aléatoire de Poisson on prouve facilement que

$$\begin{aligned} E(X) &= \lambda \\ V(X) &= \lambda \\ \sigma(X) &= \sqrt{\lambda}. \end{aligned} \tag{7.38}$$

En effet, on a

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} kP_r(X=k) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} \\ &= \frac{0\lambda^0}{0!} + \sum_{k=1}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \frac{0\lambda^0}{0!} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}e^{-\lambda} \\ &= 0 + \lambda\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} = \lambda\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} \\ & \lambda \end{aligned}$$

et

$$\begin{aligned} V(X) &= \sum_{k=0}^{\infty} k^2 P_r(X=k) - (E(X))^2 \\ &= \sum_{k=0}^{\infty} k^2 P_r(X=k) - \lambda^2 = \sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} - \lambda^2 \\ &= 0 + \frac{\lambda}{1}e^{-\lambda} + \sum_{k=2}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} - \lambda^2 \\ &= 0 + \lambda e^{-\lambda} + \lambda\sum_{k=2}^{\infty} k\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda} - \lambda^2 \\ &= = \lambda e^{-\lambda} + \lambda\sum_{k=1}^{\infty} (k+1)\frac{\lambda^k}{k!}e^{-\lambda} - \lambda^2 \\ &= \lambda e^{-\lambda} + \lambda\sum_{k=1}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} + \lambda\sum_{k=1}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} - \lambda^2 \\ &= \lambda e^{-\lambda} + \lambda\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} + \lambda\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}e^{-\lambda} - \lambda e^{-\lambda} - \lambda^2 \\ &= \lambda e^{-\lambda} + \lambda^2 + \lambda - \lambda e^{-\lambda} - \lambda^2 \\ &= \lambda \end{aligned}$$

Exemple

Soit X la variable aléatoire désignant le nombre de tremblements de terre durant une année, et on suppose que la variable aléatoire X suit la loi de Poisson de moyenne $\lambda = 4$.

Calculez les probabilités des évènements suivants.

A : il y aura exactement 2 tremblements de terre cette année.

B : il y aura au moins un tremblement de terre l'année prochaine.

C : il y aura au moins un tremblement de terre durant les deux prochaines années, en supposant que le nombre de tremblements de terre durant une année ne dépend pas du même nombre durant l'année précédente.

Solution On a

$$P_r(X = k) = \frac{4^k}{k!}e^{-4}, \text{ pour } k = 0, 1, ...$$

$$\begin{aligned} P_r(A) &= P_r(X = 2) \\ &= \frac{4^2}{2!}e^{-4} \\ &= . \end{aligned}$$

$$\begin{aligned} P_r(B) &= P_r(X \geq 1) \\ &= 1 - P_r(X < 1) \text{ (évènements complémentaires)} \\ &= 1 - P_r(X = 0) \\ &= 1 - e^{-4}. \end{aligned}$$

On considère les deux évènements

C_1 : Il y aura au moins un tremblement de terre durant la première année.

C_2 : Il y aura au moins un tremblement de terre durant la deuxième année.

$$\begin{aligned} P_r(C) &= P_r(C_1 \cup C_2) \\ &= P_r(C_1) + P_r(C_2) - P_r(C_1 \cap C_2) \\ &= P_r(C_1) + P_r(C_2) - P_r(C_1).P_r(C_2) \text{ (car } C_1 \text{ et } C_2 \text{ sont indépendants)} \\ &= \left[1 - e^{-4}\right] + \left[1 - e^{-4}\right] - \left[1 - e^{-4}\right].\left[1 - e^{-4}\right] \\ &= . \end{aligned}$$

7.7 Loi normale $\mathcal{N}(\mu, \sigma)$

7.7.1 Définitions

La loi de probabilité la plus célèbre, définie par une densité, est la loi normale ou la loi de Gauss, notée $\mathcal{N}(\mu, \sigma^2)$ et dont La fonction densité $f(x)$ est définie pour tout $x \in \mathbb{R}$ par

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}. \tag{7.39}$$

La notation $\mathcal{N}(\mu, \sigma^2)$ est justifiée par le fait que la moyenne de cette loi est égale à μ et sa variance est égale à σ^2.

Loi normale réduite centrée.

Une loi de probabilité est dite centrée si sa moyenne est égale à zéro, elle est dite réduite si sa variance est égale à un.

La loi normale $Z \rightsquigarrow \mathcal{N}(0, 1)$, ayant pour moyenne 0 et pour variance 1, est appelée loi normale réduite centrée, et sa densité est définie pour tout $t \in \mathbb{R}$ par

$$f(t) = \frac{1}{\sqrt{2\pi}} \exp -\frac{t^2}{2}. \tag{7.40}$$

Relation entre loi normale et loi normale réduite centrée.

Soient

$$Z \sim \mathcal{N}(0, 1)$$

et

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$$

(c'est-à-dire Z est une variable aléatoire normale réduite centrée et X est une variable aléatoire normale de moyenne μ et d'écart type σ.)

Si $\Pi(z)$ et $F(x)$ sont les fonctions de répartition de Z et X, respectivement, alors on la relation

$$F(x) = \Pi\left(\frac{x-\mu}{\sigma}\right) \tag{7.41}$$

pour tout $x \in \mathbb{R}$.

Idée de la démonstration.

Pour démontrer cette relation on utilise le changement de variable

$$s = \frac{t - \mu}{\sigma}.$$

Quantile d'ordre α

Un quantile d'ordre $0 < \alpha < 1$ d'une variable aléatoire X est un nombre réel x_α vérifiant

$$P_r(X \leq x_\alpha) = \alpha. \tag{7.42}$$

Notation

Dans ce document on notera le quantile d'ordre α de la loi normale réduite centrée Z par z_α.

La notion de quantile sera utilisée dans ce document, dans l'étude de la théorie de l'estimation et les tests statistiques.

7.7.2 Espérance $E(X)$, Variance $V(X)$ et Ecart type $\sigma(X)$.

Si X est une variable aléatoire et $X \sim \mathcal{N}(\mu, \sigma)$, alors on peut monter que

$$\begin{aligned} E(X) &= \mu, \\ V(X) &= \sigma^2, \end{aligned} \tag{7.43}$$

7.7.3 Calcul de probabilité pour une loi normale.

La fonction de répartition d'une variable normale n'a pas une expression analytique. Néanmoins, une table (table1) donnant les valeurs de la fonction de répartition $\Pi(x)$ de la loi normale réduite centrée $Z \sim \mathcal{N}(0,1)$ est réalisée à l'aide d'un logiciel. La relation (7.41) nous permet d'utiliser cette table pour une loi normale quelconque $\mathcal{N}(\mu, \sigma^2)$.

Remarque.

La table statistique des valeurs de la fonction de répartition $\Pi(z)$ ne fournit pas (presque dans tous les livres) les valeurs de $\Pi(z)$ pour $z < 0$.

Pour calculer $\Pi(z)$ pour un z inférieur à 0, on passe par la propriété suivante de Π, vérifiée par toute densité paire,

$$\Pi(z) = 1 - \Pi(-z), \text{ pour tout } z \in \mathbb{R}. \tag{7.44}$$

Utilisation de la table statistique des valeurs de $\Pi(z)$ (Table1 en annexe).

Exemple
Soit $X \sim \mathcal{N}(10, 4)$.
A. Calculer les probabilités
$P_r(X \leq 10)$,
$P_r(X \geq 15)$,
$P_r(5 \leq X \leq 10)$.
B. Calculez la valeur du nombre x dans les situations suivantes :
$P_r(X \leq x) = 0.95$,
$P_r(X \geq x) = 0.70$,
$P_r(5 \leq X \leq 10)$.

Solution. On note par $F(x) = P_r(X \leq x)$ la fonction de répartition de X.
A.

$$\begin{aligned} P_r(X \leq 10) &= F(10) \\ &= \Pi\left(\frac{10-10}{4}\right) \\ &= \Pi(0) \\ &= 0,5000 \text{ (par la table statistique)} \end{aligned}$$

$$\begin{aligned}
P_r(X \geq 15) &= 1 - P_r(X \leq 15) \\
&= 1 - F(15) \\
&= 1 - \Pi\left(\frac{15-10}{4}\right) \\
&= 1 - \Pi(1,25) \\
&= 1 - 0,8944 \\
&= 0,1056
\end{aligned}$$

$$\begin{aligned}
P_r(5 \leq X \leq 10) &= F(10) - F(5) \\
&= \Pi\left(\frac{10-10}{4}\right) - \Pi\left(\frac{5-10}{4}\right) \\
&= \Pi(0) - \Pi(-1,25) \\
& \quad 0.5000 - [1 - \Pi(1,25)] \\
&= 0,5000 - 0,1056 \\
&= 0,3944
\end{aligned}$$

B)

$$\begin{aligned}
P_r(X \leq x) &= 0,95 \\
F(x) &= 0,95 \\
\Pi\left(\frac{x-10}{4}\right) &= 0,95
\end{aligned}$$

Puisque $0,95 \geq 0.5000$ le quantile $z = \frac{x-10}{4}$ est à chercher directement sur la table statistique donnant approximativement la valeur des quantiles dont l'image par la fonction caractéristique Π est supérieure ou égale à 0.5000.

Sur la table des valeurs de Π, il n'y a pas un quantile z tel que $\Pi(z) = 0.95$. Pour avoir une valeur approximative de $z = \frac{x-10}{4}$, on procède de la façon suivante.

a) On prend les deux nombres $z_1 = 0.9495$ et $z_2 = 0.9505$ dont les images $\Pi(1,64) = 0,9495$ et $\Pi(1,65) = 0,9505$ encadrent $0,95$.

b) On prend comme valeur approximative de $z = \frac{x-10}{4}$, soit la moyenne $1,645 = \frac{1,64+1,65}{2}$, soit la valeur dont l'image $\Pi(.)$ est la proche de $0,95$.

c) on pose donc

$$\frac{x-10}{4} \simeq 1,645$$

d'où

$$\begin{aligned} x &\simeq 4\,(1,645) + 10 \\ x &\simeq 16,58 \end{aligned}$$

Donc le quantile x qui vérifie $P_r\,(Y \leq y) = 0.95$ est

$$\mathbf{x \simeq 16,58}.$$

$$\begin{aligned} P_r\,(X \geq x) &= 0.70 \\ 1 - P_r\,(X \leq x) &= 0,70 \\ 1 - P_r\,(X \leq x) &= 0,70 \\ P_r\,(X \leq x) &= 0,30 \\ F\,(x) &= 0,30 \\ \Pi\left(\frac{x-10}{4}\right) &= 0,30. \end{aligned}$$

Puisque $0,30 < 0.5000$, Le quantile $z = \frac{x-10}{4}$ ne figure pas directement sur la table. On cherche d'abord son opposé $-z$.

a) on a

$$\begin{aligned} \Pi\,(-z) &= 1 - \Pi\,(z) \\ &= 1 - 0,30 \\ &= 0,70 \end{aligned}$$

b) On cherche sur la table, le quantile dont l'image par Π est égale à 0.70 (ou les deux quantiles dont les images sont les plus proches de 0.70). On trouve

$$\Pi\,(0,52) = 0,6985 \text{ et } \Pi\,(0,53) = 0,7019.$$

On prend comme valeur de $-z$, la moyenne des deux quantiles $0,52$ et $0,53$:

$$-z = -\left(\frac{x-10}{4}\right) \simeq \frac{0,52\ +0,53}{2} = 0,525$$

d'où

$$\left(\frac{x-10}{4}\right) = -0,525.$$

et

$$\mathbf{x = 7,9}$$

Utilisation de la table statistique de l'écart réduit (table2. en annexe).

Soit Z une variable normale réduite centrée : $Z \sim N(0,1)$ et soit α un nombre compris entre 0 et 1. on note par $\overline{z}_\alpha$ le nombre positif qui vérifie

$$P_r(|Z| \geq \overline{z}_\alpha) = \alpha.$$

La table2. en annexe nous donne la valeur de $\overline{z}_\alpha$ pour chaque $0 < \alpha < 1$, avec un pas de 0.01.

Exemple.

Soit $Z \sim N(0,1)$. Calculer $\overline{z}$ dans les situations suivantes :

a.

$$P_r(|Z| \geq \overline{z}) = 0.23$$

b.

$$P_r(|Z| \geq \overline{z}) = 0.254$$

c.

$$P_r(|Z| \leq \overline{z}) = 0.90$$

Solution.

a. $\alpha = 0.23$ figure sur la table2., la valeur de $\overline{z}$ se trouve dans la cellule intersection de la ligne contenant la partie décimale 0.2 et la colonne contenant la partie centième 0.03 : donc $\overline{z} = \overline{z}_{0.03} = 1.200$.

b. $\alpha = 0.254$ ne figure pas sur la table. Dans ce cas,

- soit on prend la valeur la plus proche de $\alpha = 0.254$, c'est-à dire on pose $\alpha \simeq 0.25$ et par conséquent on aura

$$\overline{z} = \overline{z}_{0.254} \simeq \overline{z}_{0.25} = \mathbf{1.150}.$$

- Soit en considère les deux valeurs 0.25 et 0.26 qui encadrent $\alpha = 0.254$ et on prend la moyenne, et par conséquent on aura

$$\begin{aligned} \overline{z} &= \overline{z}_{0.254} \simeq \frac{\overline{z}_{0.25} + \overline{z}_{0.26}}{2} \\ &= \frac{1.150 + 1.126}{2} \\ &= \mathbf{1.138}. \end{aligned}$$

c.

$$\begin{aligned} 1 - P_r(|Z| \leq \overline{z}) &= 1 - 0.90 \\ P_r(|Z| \geq \overline{z}) &= 0.10. \end{aligned}$$

la table 2. de l'écart réduit donne

$$\overline{z} = \overline{z}_{0.10} = \mathbf{1.645}.$$

Calcul de $\overline{z}_\alpha$ par la table1.

Soit $0 < \alpha < 1$. et soit $\overline{z}_\alpha$ le nombre qui vérifie

$$P_r(|Z| \geq \overline{z}_\alpha) = \alpha. \tag{7.45}$$

On a

$$\begin{aligned}
P_r(|Z| \geq \overline{z}_\alpha) &= P_r(Z \leq -\overline{z}_\alpha) + P_r(Z \geq \overline{z}_\alpha) \\
\alpha &= P_r(Z \leq -t) + (1 - P_r(Z \leq \overline{z}_\alpha)) \\
\alpha &= \Pi(-\overline{z}_\alpha) + (1 - \Pi(\overline{z}_\alpha)) \\
\alpha &= 1 - \Pi(\overline{z}_\alpha) + (1 - \Pi(\overline{z}_\alpha)) \\
\alpha &= 2(1 - \Pi(\overline{z}_\alpha)).
\end{aligned}$$

D'où,

$$\Pi(\overline{z}_\alpha) = 1 - \frac{\alpha}{2}.$$

Donc $\overline{z}_\alpha$ vérifiant (7.45) est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale réduite centrée :

$$\overline{z}_\alpha = z_{1-\frac{\alpha}{2}}. \tag{7.46}$$

Ainsi, on peut calculer $\overline{z}_\alpha$, en utilisant la table1. de la fonction Π de répartition de la loi normale réduite centrée.

Exemple.

Utiliser la table1. de la fonction de répartition Π de la loi normale réduite centrée, pour calculer $\overline{z}$ qui vérifie

a.

$$P_r(|Z| \geq \overline{z}) = 0.418.$$

b.

$$P_r(|Z| \geq \overline{z}) = 0.1224.$$

Solution.

a. On a

$$\begin{aligned} 1-\frac{\alpha}{2} &= 1-\frac{0.418}{2} \\ &= 0.7910. \end{aligned}$$

La table1. donne

$$\Pi(0.81) = 0.7910, \text{ c'est-dire } z_{0.7910} = 0.81.$$

Donc, dans ce cas,

$$\begin{aligned} \overline{z} &= \overline{z}_{0.418} \\ &= z_{1-\frac{0.418}{2}} \text{ (formule 7.46)} \\ &= z_{0.7910} \\ &= \mathbf{0.81}. \end{aligned}$$

b. On a

$$\begin{aligned} \overline{z} &= \overline{z}_{0.1224} \\ &= z_{1-\frac{0.1224}{2}} \\ &= z_{0.9388}. \end{aligned}$$

Dans ce cas, il n'existe pas sur la table1. un nombre z qui vérifie

$$\Pi(z) = 0.9388.$$

Dans ce cas,

1. soit on prend le quantile du nombre 0.9382 (ou 0.9394) le plus proche du nombre 0.9388 comme valeur approchée de $\overline{z} = z_{0.9388}$. On pose alors

$$\begin{aligned} \overline{z} &= \overline{z}_{0.9388} \\ &\simeq z_{0.9382} \\ &= \mathbf{1.54} \end{aligned}$$

2. soit on prend la moyenne des quantiles 1.55 et 1.56 des deux nombres 0.9382 et 0.9394 qui encadrent le nombre 0.9388. On pose alors

$$\begin{aligned} \overline{z} &= \overline{z}_{0.9388} \\ &\simeq \frac{z_{0.9382}+z_{0.9386}}{2} \\ &= \frac{1.54+1.55}{2} \\ &= \mathbf{1.545}. \end{aligned}$$

7.7.4 Approximation de lois discrètes par la loi normale

Il est parfois trés pénible de calculer des probabilités pour certaines lois discrètes. Heureusement, les mathématiciens ont montré qu'on peut sous certaines conditions approximer ces lois par une loi normale. On résume dans cette section les résultats essentiels de ces approximations.

Approximation d'une loi binomiale.

On admet le résultat suivant.

Une loi normale $B(n;p)$ peut être remplacée par une loi normale $N\left(np;\sqrt{np(1-p)}\right)$ de moyenne np et d'écart-type $\sqrt{np(1-p)}$, si

a) $n \geq 30$,
b) $np > 5$
c) $n(1-p) > 5$.

Approximation d'une loi de Poisson.

On admet le résultat suivant.

Une loi de Poisson $P(\lambda)$ peut être remplacée par une loi normale $B\left(\lambda;\sqrt{\lambda}\right)$ de moyenne λ et d'écart-type $\sqrt{\lambda}$, si $\lambda > 5$.

Correction de continuité.

La probabilité d'un point n'est pas nulle pour une loi binomiale ou une loi de Poisson, mais cette même probabilité est toujours nulle pour une loi normale. Pour remédier à cette contradiction, quand on approxime une loi binomiale ou une loi de Poisson par une loi normale, on utilise ce qu'on appelle une correction de continuité qui se résume au remplacement de la probabilité d'un point par un intervalle centré en ce même point et de longueur 1. On résume les situations essentielles, qui peuvent se présenter, dans le tableau suivant.

Si X suit une loi normale ou de Poisson et Y la loi normale qui approxime X, n et m deux entiers naturels, alors

$$P_r(X = n) \quad \text{est approximé par } P_r\left(n - \frac{1}{2} \leq Y \leq n + \frac{1}{2}\right)$$
$$P_r(m \leq X \leq n) \quad \text{est approximé par } P_r\left(m - \frac{1}{2} \leq Y \leq n + \frac{1}{2}\right)$$
$$P_r(m < X \leq n) \quad \text{est approximé par } P_r\left(m + \frac{1}{2} \leq Y \leq n + \frac{1}{2}\right)$$
$$P_r(m < X < n) \quad \text{est approximé par } P_r\left(m + \frac{1}{2} \leq Y \leq n - \frac{1}{2}\right)$$
$$P_r(X \leq n) \quad \text{est approximé par } P_r\left(Y \leq n + \frac{1}{2}\right)$$
$$P_r(m < X) \quad \text{est approximé par } P_r\left(m + \frac{1}{2} \leq Y\right)$$

Quatrième partie