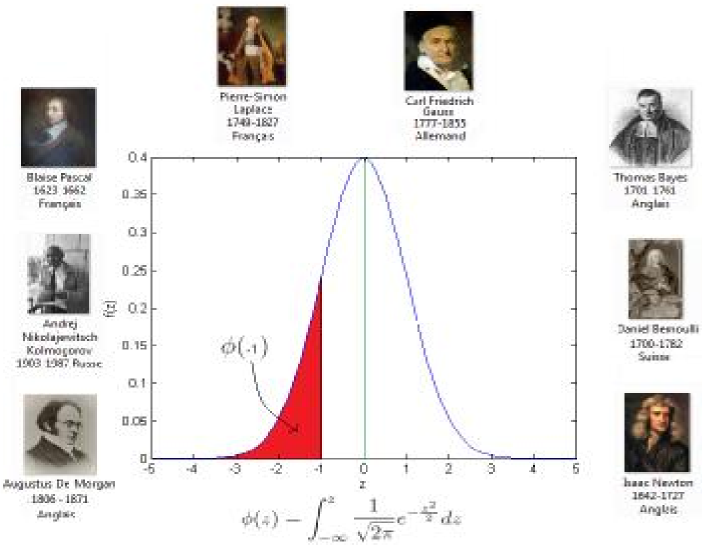
République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Faculté de Technologie, Université de Laghouat

Département des Sciences et Techniques



Cours de Statistique et Probabilités pour les Sciences et Techniques

Dr. Mohand BENTOBACHE

Maître de Conférences, Université de Laghouat

Chargé de recherche, Laboratoire de Modélisation et d'Optimisation des Systèmes LAMOS, Université de Béjaia.

e-mail : m.bentobache@mail.univ-lagh.dz, mbentobache@yahoo.com

Année Universitaire : 2014-2015.

Programme (Durée : Cours 22h30mn en 15 semaines, TD 22h30mn)

- ▶ **Chapitre 1 :** Introduction, définitions et terminologie, caractère qualitatif (Durée=3h).
- ▶ **Chapitre 2 :** Variable statistique discrète (Durée=4.5h).
- ▶ **Chapitre 3 :** Variable statistique continue (Durée=3h).
- ▶ **Chapitre 4 :** Variable statistique double et ajustement linéaire (Durée=1h30).
- ▶ **Chapitre 5 :** Analyse combinatoire et théorie fondamentale des probabilités (Durée=4.5h).
- ▶ **Chapitre 6 :** Variables aléatoires et lois usuelles de probabilités (Durée=6h).

Description

Ce module appelé MATHS4 (Statistique et Probabilités) est assuré aux étudiants de deuxième année licence LMD spécialité Sciences et Techniques. Il fait partie de l'unité méthodologie, coefficient=4, nombre de crédits=4. L'objectif du module est l'acquisition des compétences nécessaires (méthodes et techniques statistiques) pour la récolte, le traitement et l'analyse des données statistiques issues de différents domaines des sciences de l'ingénieur.

Logiciels utilisés

Microsoft Excel (Statistique) et Matlab (Probabilités).

Chapitre 1: Introduction, Définitions et Terminologie, Caractère Qualitatif

I. Plan du chapitre

Cours 1: Introduction, définitions et terminologie, caractère qualitatif (Durée=3h)

- ► Exemples Introductifs.
- ► Introduction.
- ► Vocabulaire de la statistique: population, caractère, échantillon, variables, etc.
- ► Caractère statistique qualitatif et représentations graphiques.

II. Introduction

La statistique est l'ensemble des méthodes scientifiques à partir desquelles on recueille, organise, résume et analyse des données, et qui permettent d'en tirer des conclusions et prendre de bonnes décisions.

Pour effectuer une étude statistique, on suit généralement les étapes suivantes:

Étape 1:

- Récolter les données;
- Trier et regrouper les observations;
- Établir la distribution des fréquences et le tableau statistique.

Étape 2:

- Visualiser les données grace à des représentations graphiques (diagramme en secteur, diagrammes en tuyaux d'orgues, diagramme en bâtons, histogramme,...).

Étape 3:

- Synthétiser les données en quelques grandeurs représentatives (moyenne, mode, médiane,...);
- Mesurer la dispersion des différentes observations par rapport à ces grandeurs (variance, écart type,...).

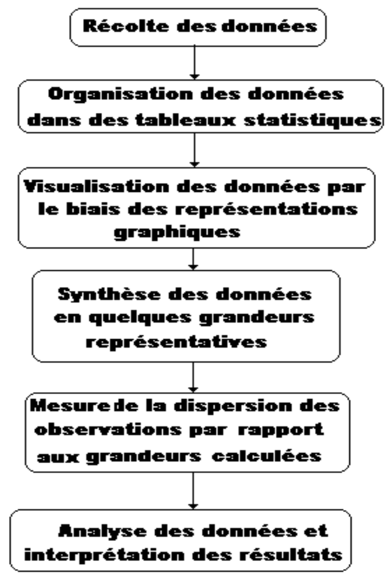
Ces différentes étapes sont résumées dans le schéma de la figure 1.

Avant de définir les termes les plus importants dans la statistique, nous présentons d'abord quelques exemples introductifs.

III. Exemples introductifs

Exemple 1: Le tableau suivant donne la répartition du personnel d'une entreprise selon la catégorie socioprofessionnelle:

Figure 1. Étapes à suivre pour mener une étude statistique



Catégorie socioprofessionnelle	Nombre d'employés
Simples employés	45
Ingénieurs	10
Techniciens	25
Techniciens supérieurs	20
Total	100

Exemple 2: Soit la distribution statistique donnant la répartition de 116 familles selon leurs nombre d'enfants:

Nombre d'enfants	Nombre de familles
0	6
1	18
2	25
3	33
4	21
5	13
Total	116

Exemple 3: Considérons la distribution du poids de 133 étudiants d'une certaine université:

Poids (en kg)	Nombre d'étudiants
[56,58[	5
[58,60[	12
[60,62[	18
[62,64[	39
[64,66[	36
[66,69[	15
[69,72[	08
Total	133

IV. DÉFINITIONS

A. *Population*

► L'ensemble sur lequel porte l'étude statistique s'appelle la *population*.
► Les éléments de cet ensemble s'appellent *individus* ou *unités statistiques*.
► La population peut être un groupe de personnes, d'objets, de choses concrètes ou même abstraites.
► Le nombre total d'individus qui constituent la population est appelé *effectif total*. On le note par N.
► Un sous ensemble d'individus choisi d'une manière adéquate est appelé *échantillon*.

B. *Caractère statistique*

Le caractère est la propriété choisie pour l'étude statistique.
Exemple Une population constituée d'un ensemble d'étudiants peut faire l'objet d'une étude concernant la "taille", le "poids" ou bien leurs "notes en mathématiques".

C. *Modalités*

Les différentes positions que peut prendre un caractère s'appellent les modalités.

Dans l'exemple 2, le caractère étudié est le nombre d'enfants par famille et ses modalités sont: 0, 1, 2, 3, 4 et 5.

Remarque 1: Un individu ne doit pas avoir plus d'une modalité.

D. *Effectif*

Lorsque la population est répartie sur les différentes modalités, nous obtenons pour chacune d'elle un nombre: c'est le nombre d'individus ayant cette modalité, on l'appelle *effectif* ou aussi *fréquence absolue*. On note généralement l'effectif correspondant à la modalité i, par n_i.

E. *Série statistique*

Les couples (m_i, n_i), $i = \overline{1,k}$, constitués de la modalité m_i et de l'effectif correspondant forment une suite qu'on appelle série statistique.
Dans l'exemple 2, la série statistique étudiée est (0,6), (1,18), (2,25), (3,33), (4,21), (5,13). Dans l'exemple 3: ([56,58[,5), ([58,60[,12),...

F. *Distribution statistique*

Soit une population concernant N individus. Le classement des individus de cette population selon les différentes modalités du caractère donne naissance à une *distribution statistique* généralement présentée dans un *tableau statistique*.

G. *Fréquences relatives*

La fréquence relative qui correspond à la modalité m_i et qu'on note f_i représente la proportion d'individus de la population ayant la modalité m_i. Elle est donnée par:

$$f_i = \frac{n_i}{N}, \; i = 1, 2, \cdots, k. \quad (1)$$

Remarque 2:

$$\begin{aligned} \textstyle\sum_{i=1}^{k} f_i &= f_1 + f_2 + \cdots + f_k \\ &= \tfrac{n_1}{N} + \tfrac{n_2}{N} + \cdots + \tfrac{n_k}{N} \\ &= \tfrac{n_1+n_2+\cdots+n_k}{N} = 1. \end{aligned}$$

H. *Tableau statistique*

Un tableau statistique est un tableau à deux colonnes. La première colonne de gauche est toujours réservée aux modalités du caractère à étudier tandis que celle de droite est utilisée pour indiquer les effectifs correspondant aux différentes modalités.

Modalités	Effectifs	Fréquences relatives
m_1	n_1	f_1
m_2	n_2	f_2
$\vdots$	$\vdots$	$\vdots$
m_k	n_k	f_k
Total	$N = \sum_{i=1}^{k} n_i$	$\sum_{i=1}^{k} f_i = 1$

I. *Nature du caractère statistique*

Le caractère statistique peut être soit *qualitatif* ou *quantitatif*.
► Un caractère est dit quantitatif quand ses modalités ne sont pas mesurables.
► Un caractère est dit quantitatif quand ses modalités sont mesurables.

J. *Variable statistique*

Dans le cas où le caractère statistique est quantitatif, on l'appelle *variable statistique*, elle peut être soit *discrète* si elle prend des valeurs isolées ou bien *continue* si elle prend ses valeurs dans un intervalle.

Dans l'exemple 1, le caractère étudié est la catégorie socioprofessionnelle des employés. C'est un caractère qualitatif car ses modalités (Simples employés, Ingénieurs,...) ne sont pas mesurables. Dans l'exemple 2, le caractère étudié est le nombre d'enfants par famille et il est quantitatif discret car ses modalités sont mesurables. De plus, elle sont isolées. Enfin, dans le dernier exemple (exemple 3), le caractère étudié est le poids des étudiants. C'est un caractère quantitatif continu car il prend ses valeurs dans des intervalles.

Figure 2. Classification du caractère statistique

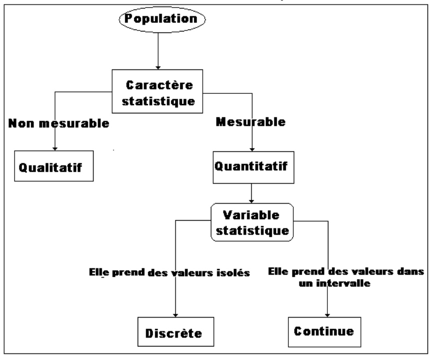


Figure 3. Diagramme en tuyaux d'orgue de la répartition des employés selon la catégorie socioprofessionnelle

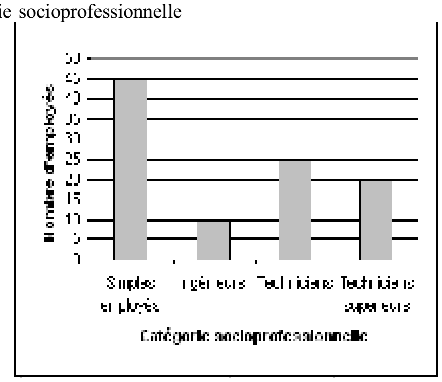
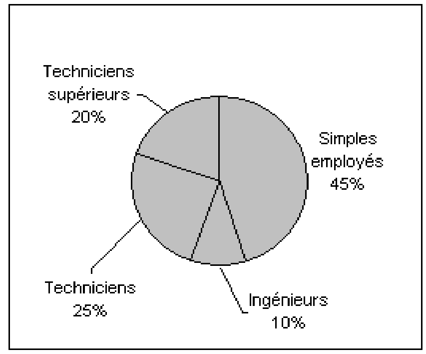


Figure 4. Diagramme en secteurs de la répartition des employés selon la catégorie socioprofessionnelle



V. Caractère qualitatif

Pour bien comprendre comment mener une étude statistique sur un caractère qualitatif, étudions la distribution de l'exemple 1.

Dans l'exemple 1, il s'agit d'étudier la catégorie socio-professionnelle d'une population constituée de 100 employés d'une entreprise. Pour cela, dressons un tableau là où on va calculer les proportions d'employés de chaque catégorie, f_i, le pourcentage des employés de chaque catégorie, $p_i = f_i \times 100$, ainsi que les angles correspondant à chaque catégorie, $\theta_i = f_i \times 360$.

Catégorie socioprofessionnelle	Nombre d'employés	f_i	p_i (%)	θ_i (°)
Simples employés	45	0.45	45	162
Ingénieurs	10	0.1	10	36
Techniciens	25	0.25	25	90
Techniciens supérieurs	20	0.2	20	72
Total	100	1	100	360

Il existe deux représentations graphiques adéquates pour mieux visualiser une distribution à caractère qualitatif: le diagramme en tuyaux d'orgue et le diagramme en secteur. Le diagramme en tuyaux d'orgue ainsi que le diagramme en secteurs représentant les données de l'exemple introductif 1 sont montrés dans les figures 3 et 4.

VI. Exercices de TD

EXERCICE 1:

On s'intéresse au type d'activité principale de 40 entreprises. Il y a 5 types possibles (Catégories) : I (Industrie), T (Transport), C (Communications), S (Services), A (Autres). Les données sont résumées dans le tableau suivant:

Activité principale	Nombre d'entreprises
Industrie	15
Communications	10
Transport	8
Services	3
Autres	4
Total	40

1) Identifiez la population, le caractère étudié et sa nature.
2) Représenter graphiquement les données de cette distribution par un diagramme en tuyaux d'orgue; puis par un diagramme en secteurs angulaires.

VII. Exercices supplémentaires de TD

EXERCICE 2:

Le tableau suivant représente l'évolution de la population mondiale pendant les années: 1950, 1960 et 1970 répartie sur les différents régions géographiques.

Région/Année	1950	1960	1970
Afrique	217	278	344
Asie	1388	1659	2056
Europe	392	425	462
Amérique du nord	166	199	228
Amérique du sud	162	213	283
Australie	13	16	19
Russie	180	214	243

1) Représenter les données de ce tableau dans un seul diagramme en tuyaux d'orgue.
2) Représenter les données de chaque année par un diagramme en secteurs.
3) Représenter par des courbes les données des séries chronologiques correspondant à l'évolution de la population pour chaque région.

Chapitre 2: Variable Statistique Discrète

I. Plan du chapitre

Cours 1: Variable statistique discrète (VSD) et représentations graphiques (Durée=3h)

- Exemples Introductifs.
- Définition et tableau statistique d'une VSD.
- Distribution statistique des effectifs (fréquences) et leurs représentations graphiques.
- Effectifs (fréquences) cumulés croissants (décroissants) et leurs représentations graphiques.

Cours 2: Paramètres de position centrale et de dispersion d'une VSD (Durée=1h30mn)

- Mode, médiane et quartiles.
- Moyennes (arithmétique, quadratique, géométrique et harmonique).
- Variance, écart-type, coefficient de variation, étendue, écart interquartile.

II. Exemple introductif

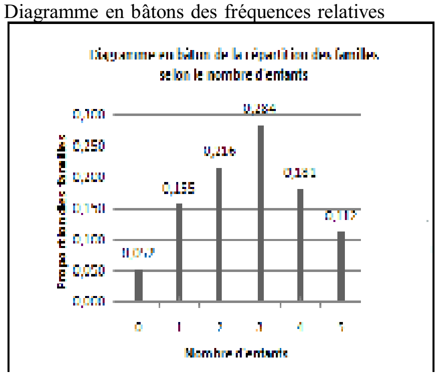
***Exemple** 1:* Soit la distribution statistique donnant la répartition de 116 familles selon leurs nombre d'enfants:

Nombre d'enfants	Nombre de familles
0	6
1	18
2	25
3	33
4	21
5	13
Total	116

III. Définition et tableau statistique d'une VSD

***Définition** 1:*
Une variable statistique est dite *discrète* si elle prend des valeurs isolées.
Soit X une variable statistique discrète qui prend les valeurs (modalités) x_1, x_2,…,x_k, avec $x_1 < x_2 < \ldots < x_k$ et supposons qu'ils existent n_i individus ayant la valeur x_i. Le tableau statistique correspondant à une variable statistique discrète est généralement présenté sous la forme suivante:

Figure 1. Diagramme en bâtons des fréquences relatives



x_i	Effectifs n_i	Fréquences relatives f_i
x_1	n_1	f_1
x_2	n_2	f_2
$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k
Total	$N = \sum_{i=1}^{k} n_i$	$\sum_{i=1}^{k} f_i = 1$

où $f_i = \frac{n_i}{N}$ représente la proportion d'individus ayant la valeur x_i, $i = \overline{1,k}$.

IV. Représentation graphique des effectifs et fréquences relatives

Les effectifs n_i (resp. les fréquences relatives f_i) pour une variables statistique discrète sont représentés par un *diagramme en bâtons*:

- sur l'axe des abscisses, on positionne les différentes valeurs de la variable statistique X;
- sur l'axe des ordonnées on représente les effectifs n_i (resp. les fréquences relatives f_i);
- pour chaque valeur x_i, on trace un trait (bâton) perpendiculaire à l'axe des abscisses dont la hauteur est proportionnelle à l'effectif n_i (resp. la fréquence relative f_i).

***Exemple** 2:* Représentons les fréquences relatives pour l'exemple 1 (X:nombre d'enfants par famille): voir Figure 1.

V. Effectifs cumulés et fréquences cumulées

*A. **Effectifs cumulés croissants et fréquences cumulées croissantes***

***Définition** 2:* L'effectif cumulé croissant (resp. fréquence cumulée croissante) correspondant à la valeur x_i est le nombre d'individus (resp. proportion d'individus) ayant une valeur de la variable statistique X inférieure ou égale à x_i ou bien strictement inférieure à x_{i+1}. On le note par N_i^+ (resp. F_i^+).

On a

$$\begin{aligned} N_1^+ &= n_1 \\ N_2^+ &= n_1 + n_2 = N_1^+ + n_2 \\ &\vdots \\ N_i^+ &= n_1 + n_2 + \dots n_i = N_{i-1}^+ + n_i \\ &\vdots \\ N_k^+ &= n_1 + n_2 + \dots n_k = N_{k-1}^+ + n_k = N. \end{aligned}$$

D'une façon générale, on a

$$\begin{cases} N_1^+ = n_1, \\ N_i^+ = N_{i-1}^+ + n_i, i = 2, 3, \dots, k. \end{cases} \tag{1}$$

De même, on aura

$$\begin{cases} F_1^+ = f_1, \\ F_i^+ = F_{i-1}^+ + f_i, i = 2, 3, \dots, k. \end{cases} \tag{2}$$

B. *Effectifs cumulés décroissants et fréquences cumulées décroissantes*

***Définition** 3:* L'effectif cumulé décroissant (resp. fréquence cumulée décroissante) correspondant à la valeur x_i est le nombre d'individus (resp. proportion d'individus) ayant une valeur de la variable statistique X supérieure ou égale à x_i ou bien strictement supérieure à x_{i-1}. On le note par N_i^- (resp. F_i^-).
On a

$$\begin{aligned} N_1^- &= N \\ N_2^- &= n_2 + n_3 + \dots + n_k = N_1^- - n_1 \\ &\vdots \\ N_i^- &= n_i + n_{i+1} + \dots n_k = N_{i-1}^- - n_{i-1} \\ &\vdots \\ N_k^- &= n_k = N_{k-1}^- - n_{k-1} = N. \end{aligned}$$

D'une façon générale, on a

$$\begin{cases} N_1^- = N, \\ N_i^- = N_{i-1}^- - n_{i-1}, i = 2, 3, \dots, k. \end{cases} \tag{3}$$

De même, on aura

$$\begin{cases} F_1^- = 1, \\ F_i^- = F_{i-1}^- - f_{i-1}, i = 2, 3, \dots, k. \end{cases} \tag{4}$$

***Remarque** 1:* Pour vérifier l'exactitude des calculs, on peut utiliser les égalités suivantes:

$$N_k^+ = N, \ N_k^- = n_1, \ F_k^+ = 1, \ F_k^- = f_1.$$

***Remarque** 2:* Remarquons que

$$F_i^+ = \frac{N_i^+}{N} \quad \text{et} \quad F_i^- = \frac{N_i^-}{N}, \ i = 1, 2, \dots, k. \tag{5}$$

***Remarque** 3:*

$$\begin{aligned} N_i^- &= n_i + n_{i+1} + \dots + n_k \\ &= N - (n_1 + n_2 + \dots + n_{i-1}) \\ &= N - N_{i-1}^+. \end{aligned}$$

D'où

$$N_{i-1}^+ + N_i^- = N. \tag{6}$$

De même, on aura

$$F_{i-1}^+ + F_i^- = 1. \tag{7}$$

***Exemple** 3:* Le tableau statistique suivant donne la répartition du poids de 15 caisses.

Poids (kg)	Nombre de caisses
15.5	3
20	5
25	6
50	1
Total	15

Ici la variable statistique désigne le poids des caisses. C'est une variable statistique discrète car elles prend des valeurs isolés. Calculons les fréquences relatives (f_i), les effectifs cumulés croissants (N_i^+), les effectifs cumulés décroissants (N_i^-), les fréquences cumulées croissantes (F_i^+) et les fréquences cumulées décroissantes (F_i^-) pour cette série statistique:

x_i	n_i	f_i	N_i^+
15.5	3	0.2	$N_1^+ = 3$
20	5	0.333	$N_2^+ = 3 + 5 = 8$
25	6	0.4	$N_3^+ = 8 + 6 = 14$
50	1	0.067	$N_4^+ = 14 + 1 = 15$
Total	15	1	

N_i^-	F_i^+	F_i^-
$N_1^- = 15$	$F_1^+ = 0.2$	$F_1^- = 1$
$N_2^- = 15 - 3 = 12$	$F_2^+ = 0.533$	$F_2^- = 0.8$
$N_3^- = 12 - 5 = 7$	$F_3^+ = 0.933$	$F_3^- = 0.467$
$N_4^- = 7 - 6 = 1$	$F_4^+ = 1$	$F_4^- = 0.067$

- $N_3^+ = n_1 + n_2 + n_3 = N_2^+ + n_3 = 14$ représente le nombre de caisses ayant un poids inférieur ou égal à $x_3 = 25kg$ ou strictement inférieur à $x_4 = 50kg$.
- $F_3^+ = f_1 + f_2 + f_3 = F_2^+ + f_3 = 0.933$ représente la proportion de caisses ayant un poids inférieur ou égal à $x_3 = 25kg$ ou strictement inférieur à $x_4 = 50kg$. En d'autre termes, 93.3% des caisses ont un poids inférieur ou égal à 25 kg.
- $N_3^- = n_3 + n_4 = N_2^- - n_2 = 7$ représente le nombre de caisses ayant un poids supérieur ou égal à $x_3 = 25kg$ ou strictement supérieur à $x_2 = 20kg$.
- $F_3^- = f_3 + f_4 = F_2^- - f_2 = 0.467$ représente la proportion de caisses ayant un poids supérieur ou égal à $x_3 = 25kg$ ou strictement supérieur à $x_2 = 20kg$. En d'autre termes, 46.7% des caisses ont un poids supérieur ou égal à 25 kg.
- Remarquons aussi que $N - N_1^+ = 15 - 3 = 12 = N_2^- \Rightarrow N_1^+ + N_2^- = N$. On a aussi: $N_2^+ + N_3^- = 8 + 7 = N$ et $N_3^+ + N_4^- = 14 + 1 = N$.

Dans l'exemple précédent, les quantités N_i^+ et F_i^+ sont calculées pour répondre à la question suivante: quel est le nombre ou la proportion de caisses ayant un poids inférieur ou égal à une valeur x_i de la variable statistique X. De plus, N_i^- et F_i^- sont calculées pour répondre à la question: quel est le nombre ou la proportion de caisses ayant un poids supérieur ou égal à une valeur x_i? Mais comment peut-on calculer:

- le nombre ou la proportion de caisses ayant un poids inférieur ou égal à une valeur α quelconque de $\mathbb{R}$?
- le nombre ou la proportion de caisses ayant un poids supérieur ou égal à une valeur $\beta \in \mathbb{R}$?
- le nombre de caisses ayant un poids compris entre α et β, avec $\alpha < \beta$?

Afin de répondre à ces questions importantes en pratique, nous introduisons ce qu'on appelle les fonctions de répartition de la variable statistique X.

Définissons alors les fonctions suivantes:

$\mathbf{N}^+ : \mathbb{R} \to [0, N]$

$x \mapsto \mathbf{N}^+(x)$, où $\mathbf{N}^+(x)$ représente le nombre d'individus ayant une valeur de la variable statistique X inférieure ou égale à x.

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

$x \mapsto \mathbf{F}^+(x)$, où $\mathbf{F}^+(x)$ représente la proportion d'individus ayant une valeur de la variable statistique X inférieure ou égale à x.

$\mathbf{N}^- : \mathbb{R} \to [0, N]$

$x \mapsto \mathbf{N}^-(x)$, où $\mathbf{N}^-(x)$ représente le nombre d'individus ayant une valeur de la variable statistique X supérieure ou égale à x.

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

$x \mapsto \mathbf{F}^-(x)$, où $\mathbf{F}^-(x)$ représente la proportion d'individus ayant une valeur de la variable statistique X supérieure ou égale à x.

Dressons les tableaux des valeurs des différentes fonctions:

x	$\mathbf{N}^+(x)$	$\mathbf{F}^+(x)$
$x < x_1$	0	0
$x_1 \le x < x_2$	N_1^+	F_1^+
$x_2 \le x < x_3$	N_2^+	F_2^+
$\vdots$	$\vdots$	$\vdots$
$x_{k-1} \le x < x_k$	N_{k-1}^+	F_{k-1}^+
$x \ge x_k$	$N_k^+ = N$	$F_k^+ = 1$

x	$\mathbf{N}^-(x)$	$\mathbf{F}^-(x)$
$x \le x_1$	$N_1^- = N$	1
$x_1 < x \le x_2$	N_2^-	F_2^-
$x_2 < x \le x_3$	N_3^-	F_3^-
$\vdots$	$\vdots$	$\vdots$
$x_{k-1} < x \le x_k$	F_k^+	F_k^+
$x > x_k$	0	0

On peut aussi écrire les expressions des différentes fonctions de la façon suivante:

$$\mathbf{N}^+(x) = \begin{cases} 0, & \text{si } x < x_1; \\ N_i^+, & \text{si } x \in [x_i, x_{i+1}[, i = 1, \ldots, k-1; \\ N, & \text{si } x \ge x_k. \end{cases}$$

$$\mathbf{F}^+(x) = \begin{cases} 0, & \text{si } x < x_1; \\ F_i^+, & \text{si } x \in [x_i, x_{i+1}[, i = 1, \ldots, k-1; \\ 1, & \text{si } x \ge x_k. \end{cases}$$

$$\mathbf{N}^-(x) = \begin{cases} N, & \text{si } x \le x_1; \\ N_i^-, & \text{si } x \in]x_{i-1}, x_i], i = 2, \ldots, k; \\ 0, & \text{si } x > x_k. \end{cases}$$

$$\mathbf{F}^-(x) = \begin{cases} 1, & \text{si } x \le x_1; \\ F_i^-, & \text{si } x \in]x_{i-1}, x_i], i = 2, \ldots, k; \\ 0, & \text{si } x > x_k. \end{cases}$$

Notons par $\mathbf{N}_{\alpha\beta}$ (resp. $\mathbf{F}_{\alpha\beta}$) le nombre (resp. la proportion) des individus ayant une valeur de la variable statistique X comprise entre α et β. Il est facile de remarquer que

$$N_{\alpha\beta} = \mathbf{N}^+(\beta) - \mathbf{N}^+(\alpha) = \mathbf{N}^-(\alpha) - \mathbf{N}^-(\beta).$$

***Remarque** 4:* On peut remarquer que:

$$\mathbf{F}^+(x) = \frac{\mathbf{N}^+(x)}{N} \text{ et } \mathbf{F}^-(x) = \frac{\mathbf{N}^-(x)}{N}.$$

***Exemple** 4:* Dressons les tableaux des valeurs des différentes fonctions de répartition de la variable X (poids) pour l'exemple 3:

x	$\mathbf{N}^+(x)$	$\mathbf{F}^+(x)$
$x < 15.5$	0	0
$15.5 \le x < 20$	3	0.2
$20 \le x < 25$	8	0.533
$25 \le x < 50$	14	0.933
$x \ge 50$	15	1

x	$\mathbf{N}^-(x)$	$\mathbf{F}^-(x)$
$x \le 15.5$	15	1
$15.5 < x \le 20$	12	0.8
$20 < x \le 25$	7	0.467
$25 < x \le 50$	1	0.067
$x > 50$	0	0

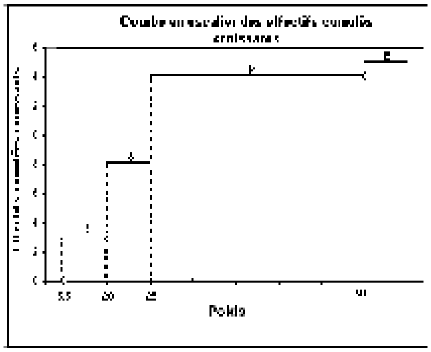
***Exercice** 1:* Déduire pour l'exemple 3:

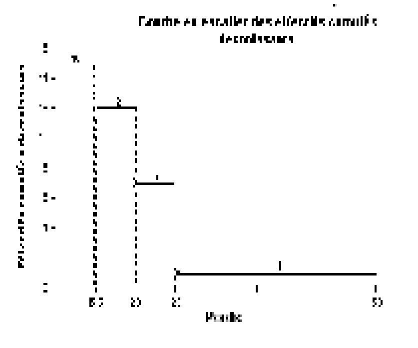
- $\mathbf{F}^+(22)$ qui représente la proportion de caisses ayant un poids inférieur ou égal à $22kg$;
- $\mathbf{F}^+(27)$ qui représente la proportion de caisses ayant un poids inférieur ou égal à $27kg$;
- $\mathbf{F}^-(22)$ qui représente la proportion de caisses ayant un poids supérieur ou égal à $22kg$;
- $\mathbf{F}^-(27)$ qui représente la proportion de caisses ayant un poids supérieur ou égal à $27kg$;
- $\mathbf{F}_{\alpha\beta}, \alpha = 22, \beta = 27$ qui représente la proportion de caisses ayant un poids compris entre 22 et 27 kg.

C. *Représentation graphique des effectifs cumulés et fréquences cumulées*

Les N_i^+ et F_i^+ sont représentés par la courbe en escalier cumulative croissante tandis que les N_i^- et F_i^- sont représentés par la courbe en escalier cumulative décroissante.

Afin de mieux comprendre comment représenter les effectifs cumulés et les fréquences cumulées, nous allons représenter les N_i^+ et N_i^- pour l'exemple 3.





VI. Caractéristiques de position centrale d'une VSD

A. *Mode*

Le mode est la valeur de la variable statistique qui correspond au plus grand effectif. On le note par Mo.

Dans l'exemple de la répartition des familles selon le nombre d'enfants (Exemple 1), le mode est $Mo = 3$ enfants. Car l'effectif maximum $n_{max} = 33$ familles.

Dans l'exemple de la répartition des caisses selon leurs poids, le mode est $Mo = 25$ kg.

***Remarque** 5:* Graphiquement (Diagramme en bâtons) le mode correspond à l'abscisse du plus haut bâton.

B. *Médiane*

La médiane représente la valeur de la variable statistique qui partage la population en deux sous-ensembles d'individus d'effectifs égaux. En d'autres termes, les individus appartenant à la première moitié ont des valeurs de la variable statistique au dessous de la médiane et ceux appartenant à la deuxième moitié ont des valeurs au dessus de la médiane. On note la médiane par Me.

1) Calcul de la médiane dans le cas d'une série statistique sans effectifs*:*

Une série statistique sans effectifs est présentée sous forme d'une suite de valeurs $x_1, x_2, \ldots, x_n$, où chaque valeur x_i de la VSD X est associée à l'individu i de la population étudiée.

***Remarque** 6:* Le nombre n, l'indice de la dernière valeur, représente la taille de la population considérée.

Pour calculer la médiane d'une série statistique sans effectifs, on suit les étapes suivantes:

(1) On ordonne par ordre croissant les valeurs de la série $(x_i)_{i=\overline{1,n}}$; on obtient une nouvelle série ordonnée qu'on note $(y_i)_{i=\overline{1,n}}$, avec $y_1 \le y_2 \le \ldots \le y_n$;

(2) on distingue alors deux cas:

(2.1) n est impair ($n = 2p+1$). Alors

$$Me = y_{p+1} = y_{\frac{n+1}{2}}.$$

(2.2) n est pair ($n = 2p$). Alors

$$Me = \frac{y_p + y_{p+1}}{2} = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2}.$$

***Exemple** 5:* Considérons la série statistique des notes de 9 étudiants dans le module de statistique:
$(x_i)_{i=\overline{1,9}}$: 13, 5, 7, 8, 6, 12, 13, 10, 15.
D'abord, on commence par ordonner cette série. On obtient alors la série suivante:
$(y_i)_{i=\overline{1,n}}$: 5, 6, 7, 8, 10, 12, 13, 13, 15.
On a $n = 9 = 2 \times 4 + 1 \Rightarrow Me = y_{\frac{n+1}{2}} = y_5 = 10$.

***Exemple** 6:* Considérons maintenant la série suivante:
$(x_i)_{i=\overline{1,8}}$: 13, 5, 7, 8, 6, 12, 13, 10.
La série ordonnée est comme suit:
$(y_i)_{i=\overline{1,8}}$: 5, 6, 7, 8, 10, 12, 13, 13.
On a dans ce cas $n = 8 = 2 \times 4 \Rightarrow Me = \frac{y_{\frac{n}{2}} + y_{\frac{n}{2}+1}}{2} = \frac{y_4 + y_5}{2} = 9$.

***Remarque** 7:* En réalité, dans l'exemple 6, la médiane n'existe pas. La note médiane $Me = 9$ ne représente qu'une valeur approximative de la médiane. Dans ce cas, on parle aussi de l'intervalle médian:

$$I_M =]y_p, y_{p+1}[=]y_{\frac{n}{2}}, y_{\frac{n}{2}+1}[.$$

2) Calcul de la médiane dans le cas d'une série statistique avec effectifs*:*

Rappelons qu'une série statistique avec effectifs est présentée sous la forme d'une suite de couples (x_i, n_i), $i = \overline{1,k}$, avec $x_1 < x_2 < \cdots < x_k$, et $\sum_{i=1}^{k} n_i = N$ représente l'effectif total de la population.

Dans ce cas, la determination de la médiane se fait en trois étapes:

(1) on calcule les N_i^+;

(2) on identifie le plus petit N_i^+ supérieur ou égal à $\frac{N}{2}$ (notons cet N_i^+ par N^{+*}) ou la plus petite F_i^+ supérieure ou égale à 0.5 (notons cette fréquence par F^{+*});

(3) Me est alors la valeur de la variable statistique correspondant à N^{+*} ou F^{+*}.

***Exemple** 7:* Calculons la médiane pour l'exemple 3:

x_i	n_i	f_i	N_i^+	F_i^+
15.5	3	0.2	3	0.2
20	5	0.333	**8**	**0.533**
25	6	0.4	14	0.933
50	1	0.067	15	1
Total	15	1		

Le plus petit N_i^+ supérieur ou égal à $\frac{N}{2} = 7.5$ est $N^{+*} = N_2^+ = 8$ caisses. Donc la médiane est $Me = x_2 = 20kg$.
La plus petite F_i^+ supérieure ou égale à 0.5 est $F^{+*} = F_2^+ = 0.533$. Donc la médiane est $Me = x_2 = 20kg$.

C. *Les quartiles*

Il existe trois quartiles:

- Le premier quartile, noté Q_1, représente la valeur de la variable statistique qui divise la population en deux sous-ensembles d'individus d'effectifs égaux à $\frac{N}{4}$ et $\frac{3N}{4}$ respectivement.
- Le deuxième quartile, noté Q_2, est égale à la médiane.
- Le troisième quartile, noté Q_3, représente la valeur de la variable statistique qui divise la population en deux sous-ensembles d'individus d'effectifs égaux à $\frac{3N}{4}$ et $\frac{N}{4}$ respectivement.

Pour l'exemple 3, on a

- Le premier quartile $Q_1 = x_2 = 20kg$, car le plus petit N_i^+ supérieur ou égal à $\frac{N}{4} = 3.75$ est $N^{+*} = N_2^+ = 8$ et la plus petite F_i^+ supérieure ou égale à 0.25 est $F^{+*} = F_2^+ = 0.533$.
- Le deuxième quartile $Q_2 = Me = 20kg$.
- Le troisième quartile $Q_3 = x_3 = 25kg$, car le plus petit N_i^+ supérieur ou égal à $\frac{3N}{4} = 11.25$ est $N^{+*} = N_3^+ = 14$ et la plus petite F_i^+ supérieure ou égale à 0.75 est $F^{+*} = F_3^+ = 0.933$.

D. *Les moyennes*

Considérons la série statistique (x_i, n_i), $i = \overline{1,k}$, avec $\sum_{i=1}^{k} n_i = N$.

*1) **Moyenne arithmétique**:* La moyenne arithmétique, notée par $\bar{X}$, est donnée par la formule:

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i. \tag{8}$$

*2) **Moyenne quadratique**:* La moyenne quadratique, notée par Q, est donnée par la formule:

$$Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i x_i^2}, \; x_i > 0, i = \overline{1,k}. \tag{9}$$

*3) **Moyenne géométrique**:* La moyenne géométrique, notée par G, est donnée par la formule:

$$G = (x_1^{n_1} \times x_2^{n_2} \times \cdots x_k^{n_k})^{\frac{1}{N}}, \; x_i > 0, i = \overline{1,k}. \tag{10}$$

Dans la pratique, on utilise la formule suivante:

$$G = e^{\frac{1}{N}\sum_{i=1}^{k} n_i ln(x_i)}. \tag{11}$$

Les formules (10) et (11) sont équivalentes. En effet, on a

$$\begin{aligned} ln(G) &= \tfrac{1}{N} ln(x_1^{n_1} \times x_2^{n_2} \times \cdots x_k^{n_k}) \\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^{k} n_i ln(x_i). \end{aligned}$$

D'où

$$G = e^{\frac{1}{N}\sum_{i=1}^{k} n_i ln(x_i)}.$$

*4) **Moyenne harmonique**:* La moyenne harmonique, notée par H, est donnée par la formule:

$$H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}, \; x_i \neq 0, i = \overline{1,k}. \tag{12}$$

***Remarque** 8:* Les différentes moyennes vérifient l'inégalité suivante:

$$H \leq G \leq \bar{X} \leq Q.$$

VII. CARACTÉRISTIQUES DE DISPERSION D'UNE VSD

Les caractéristiques de dispersion sont calculées dans le but de mesurer la dispersion des différentes valeurs de la variable statistique autour d'une valeur centrale comme la moyenne. Dans ce qui suit nous présentons les caractéristiques de dispersion les plus utilisées:

A. *Variance*

La variance est notée $V(X)$ et elle est donnée par la formule suivante:

$$V(X) = \frac{1}{N}\sum_{i=1}^{k} n_i (x_i - \bar{X})^2. \tag{13}$$

Pour des calculs pratiques, on utilise généralement la formule suivante:

$$V(X) = \frac{1}{N}\sum_{i=1}^{k} n_i x_i^2 - \bar{X}^2. \tag{14}$$

***Exercice** 2:* Démontrer la formule (14).

B. *Écart-type*

L'écart type est notée σ_X et il est donnée par la formule suivante:

$$\sigma_X = \sqrt{V(X)}. \tag{15}$$

Si σ_X est grand, alors les différentes valeurs de la variable statistique X sont fortement dispersées autour de la moyenne. Dans le cas contraire, les valeurs sont faiblement dispersées autour de la moyenne. On dit alors dans ce dernier cas que la moyenne est un bon paramètre explicatif.

***Remarque** 9:* Dans certains ouvrages, la variance est aussi notée par $Var(X)$ ou σ_X^2.

C. *Coefficient de variation*

Lorsque on veut comparer la dispersion de deux séries dont les variables statistiques sont mesurées avec deux unités de mesure différentes, on utilise le coefficient de variation noté par CV. Ce coefficient est calculé par la formule suivante:

$$CV_X = \frac{\sigma_X}{\bar{X}}. \tag{16}$$

Soient X et Y deux variables statistiques discrètes. Si $CV_X < CV_Y$, alors on dit que les valeurs de la VSD X sont moins dispersées autour de la moyenne que celles de la variable statistique Y.

D. *Étendue*

L'étendue est notée par "E". Elle représente la différence entre la plus grande valeur (x_{max}) et la plus petite valeur (x_{min}) observées:

$$E = x_{max} - x_{min}. \tag{17}$$

E. *Écart interquartile*

L'écart interquartile est noté par "EIQ". Il représente la différence entre le troisième et le premier quartile. Cet écart est donné par la formule suivante:

$$EIQ = Q_3 - Q_1. \tag{18}$$

Notons que l'intervalle interquartile:

$$IQ = [Q_1, Q_3]$$

contient 50% des observations.

Considérons une série statistique sans effectifs: $\{x_1, x_2, \dots, x_n\}$, où x_i représente la valeur de la VSD X correspondant à l'individu i. Les différentes caractéristiques de la VSD X sont données par les formules suivantes:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i,\ Q = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2},$$

$$G = e^{\frac{1}{n}\sum_{i=1}^{n} ln(x_i)},\ H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}},$$

$$V(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{X}^2,\ \sigma_X = \sqrt{V(X)}.$$

VIII. EXERCICES DE TD

EXERCICE 1: Une substance radioactive a été observée pendant 2608 intervalles de temps de $7,5$ secondes chacun; on a observé le nombre de particules x atteignant un compteur durant chaque intervalle de temps (tableau ci-dessous).

x_i	0	1	2	3	4	5	6	7
n_i	57	203	383	525	532	408	273	139

x_i	8	9	10	11	12
n_i	45	27	10	4	2

où x_i représente le nombre de particules observé et n_i le nombre de périodes pendant lesquelles on a observé x_i particules.
1) Identifiez la population, le caractère étudié et sa nature.
2) Calculer les fréquences relatives (f_i), les effectifs cumulés croissants et décroissants (N_i^+ et N_i^-); les fréquences cumulées croissantes et décroissantes (F_i^+ et F_i^-).
3) Représenter graphiquement les effectifs n_i.
4) Soit $x \in \mathbb{R}$ et considérons les fonctions suivantes N^+ et N^- définies par: $N^+ : \mathbb{R} \longrightarrow [1, 2608]$, telle que $N^+(x)$ représente le nombre de périodes pendant lesquelles le compteur a compté un nombre de particules inférieur ou égal à x particules et $N^- : \mathbb{R} \longrightarrow [1, 2608]$, telle que $N^-(x)$ représente le nombre de périodes pendant lesquelles le compteur a compté un nombre de particules supérieur ou égal à x particules. Dresser le tableau des valeurs des fonctions N^+ et N^-; puis tracer leurs graphes dans un repère orthonormé.

EXERCICE 2: Considérons les séries statistiques suivantes:
a) 1, 2, 4, 4, 4, 5, 6, 7, 8, 8, 9, 9, 10, 11, 12;
b) 1, 7, 8, 10, 12, 15, 17, 23, 24, 29.
Calculer le mode, la médiane et les différentes moyennes (arithmétique, géométrique, quadratique et harmonique) pour les différentes séries statistiques.

EXERCICE 3: Soit la distribution statistique donnant la répartition de 110 familles selon leurs nombre d'enfants:

Nombre d'enfants	Nombre de familles
1	18
2	25
3	33
4	21
5	13
Total	110

1) Identifiez la population, le caractère étudié et sa nature.
2) Calculer les différentes caractéristiques de position centrale (mode, médiane, 1**er** quartile, 2**eme** quartile, 3**eme** quartile et les différentes moyennes), les caractéristiques de dispersion (étendue, variance, écart type, écart interquartile et le coefficient de variation).

IX. EXERCICES SUPPLÉMENTAIRES DE TD

EXERCICE 4: Dans un examen hématologique, un prélèvement sanguin a été divisé en 170 régions égales et on a observé le nombre de globules rouges dans chaque région (voir le tableau suivant).

x_i	4	5	6	7	8	9	10	11	12	13	14
n_i	1	3	5	8	13	14	15	15	21	18	17

x_i	14	15	16	17	18	19	20	21
n_i	17	16	9	6	3	3	2	1

où x_i représente le nombre de globules rouges et n_i le nombre de régions dans lesquelles on a observé x_i globules rouges.
1) Identifiez la population, le caractère étudié et sa nature.
2) Calculer les fréquences relatives (f_i), les effectifs cumulés croissant et décroissants (N_i^+ et N_i^-), les fréquences cumulées croissantes et décroissantes (F_i^+ et F_i^-).
3) Représenter graphiquement les fréquences relatives f_i.
4) Soit $x \in \mathbb{R}$ et considérons les fonctions suivantes F^+ et F^- définies par: $F^+ : \mathbb{R} \longrightarrow [0, 1]$, telle que $F^+(x)$ représente la proportion des régions pendant lesquelles on a observé un nombre de globules rouges inférieur ou égal à x globules et $F^- : \mathbb{R} \longrightarrow [0, 1]$, telle que $F^-(x)$ représente la proportion des périodes pendant lesquelles on a observé un nombre de globules rouges supérieur ou égal à x globules. Dresser le tableau des valeurs des fonctions F^+ et F^-; puis tracer leurs graphes dans un repère orthonormé.

EXERCICE 5: Considérons les séries statistiques suivantes:
a) 8, 1, 10, 7, 17, 23, 12, 15, 24;
b) 4, 5, 8, 8, 9, 11, 12, 14, 17, 19.
Calculer le mode, la médiane et les différentes moyennes (arithmétique, géométrique, quadratique et harmonique) pour les différentes séries.

EXERCICE 6: Calculer pour les exercices 1 et 4 les différentes caractéristiques de position centrale (mode, médiane, 1**er** quartile, 2**eme** quartile, 3**eme** quartile et les différentes moyennes), les caractéristiques de dispersion (étendue, variance, écart type, écart interquartile et le coefficient de variation).

Chapitre 3: Variable Statistique Continue

I. Plan du chapitre

Cours 1: Variable statistique continue (VSC) et représentations graphiques (Durée=1h30mn)

- ► Exemples Introductifs.
- ► Définition et tableau statistique d'une VSC.
- ► Distribution statistique des effectifs (fréquences) et sa représentation graphique.
- ► Effectifs (fréquences) cumulés croissants (décroissants) et leurs représentations graphiques.

Cours 2: Paramètre de position centrale et de dispersion d'une VSC (Durée=1h30mn)

- ► Mode, médiane et quartiles (Calcul graphique et analytique).
- ► Moyennes (arithmétique, quadratique, géométrique et harmonique).
- ► Variance, écart-type, coefficient de variation, étendue, écart interquartile.

II. Exemples introductifs

Exemple 1: Considérons la distribution du poids de 100 étudiants:

Poids (en kg)	Nombre d'étudiants
[56,58[	5
[58,60[	10
[60,62[	18
[62,64[	29
[64,66[	16
[66,68[	14
[68,70[	08
Total	100

Exemple 2: Considérons la distribution du poids de 100 étudiants:

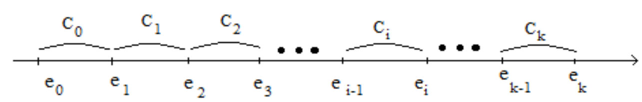
Poids (en kg)	Nombre d'étudiants
[56,58[	5
[58,60[	10
[60,62[	18
[62,64[	29
[64,66[	16
[66,70[	22
Total	100

III. Définitions et tableau statistique

Définition 1:
► Une variable statistique est dite *continue* si elle prend ses valeurs dans un intervalle.
► Considérons une certaine population ayant N individus. On s'interesse à récolter des données concernant une variable statistique continue (VSC) X. Lorsque le nombre d'observations est grand, on procède au classement des données. Le classement des données consiste à regrouper les différentes observations sous forme d'intervalles disjoints. Ces intervalles sont les modalités da la VSC X. On les appelle *classes*. La classe N° i sera notée par

$$C_i = [e_{i-1}, e_i[,\ i = 1, \ldots, k.$$

► Les e_{i-1} et e_i sont appelés respectivement borne inférieure et borne supérieure de la classe i.



-Figure 1-

► Chaque classe C_i est caractérisée par son *amplitude* notée par a_i et son *centre* noté par x_i, où

$$a_i = e_i - e_{i-1} \text{ et } x_i = \frac{e_{i-1} + e_i}{2}.$$

► Le tableau statistique d'une variable statistique continue est présenté en général sous la forme suivante:

C_i	Effectifs n_i	Fréquences relatives f_i
$[e_0, e_1[$	n_1	f_1
$[e_1, e_2[$	n_2	f_2
$\vdots$	$\vdots$	$\vdots$
$[e_{k-1}, e_k[$	n_k	f_k
Total	$N = \sum_{i=1}^{k} n_i$	$\sum_{i=1}^{k} f_i = 1$

où $f_i = \frac{n_i}{N}$ représente la proportion d'individus ayant une valeur de la variable statistique appartenant à la classe C_i, $i = \overline{1, k}$.

IV. Fréquences relatives, effectifs cumulés et fréquences cumulées

De la même manière qu'une variable statistique discrète, on définit pour chaque classe C_i, les quantités suivantes:
► n_i: le nombre d'individus ayant une valeur de la variable statistique appartenant à la classe C_i. On a $\sum_{i=1}^{k} n_i = N$.
► La fréquence relative $f_i = \frac{n_i}{N}$: la proportion d'individus ayant une valeur de la variable statistique appartenant à la classe C_i. On a alors $\sum_{i=1}^{k} f_i = 1$.
► L'effectif cumulé croissant N_i^+ (resp. la fréquence cumulée

croissante F_i^+): le nombre (resp. la proportion) d'individus ayant une valeur de la VSC X inférieure strictement à e_i. On a donc

$$\begin{cases} N_1^+ = n_1, \\ N_i^+ = N_{i-1}^+ + n_i, \quad \text{pour } i = 2, \dots, k, \end{cases}$$

et

$$\begin{cases} F_1^+ = f_1, \\ F_i^+ = F_{i-1}^+ + f_i, \quad \text{pour } i = 2, \dots, k. \end{cases}$$

► L'effectif cumulé décroissant N_i^- (resp. la fréquence cumulée décroissante F_i^-): le nombre (resp. la proportion) d'individus ayant une valeur de la VSC X supérieure ou égale à e_{i-1}. On a

$$\begin{cases} N_1^- = N, \\ N_i^- = N_{i-1}^- - n_{i-1}, \quad \text{pour } i = 2, \dots, k, \end{cases}$$

et

$$\begin{cases} F_1^- = 1, \\ F_i^- = F_{i-1}^- - f_{i-1}, \quad \text{pour } i = 2, \dots, k. \end{cases}$$

► La fonction à variable réelle $N^+(x)$ (resp. $F^+(x)$): le nombre (resp. la proportion) d'individus ayant une valeur de la VSC X inférieure ou égale à x.
► La fonction à variable réelle $N^-(x)$ (resp. $F^-(x)$): le nombre (resp. la proportion) d'individus ayant une valeur de la VSC X supérieure ou égale à x.
► Si l'on suppose que le nombre d'individus est uniformément distribué par rapport à la variable statistique X. Alors, on pourra écrire:

$$N^+(x) = \begin{cases} 0, & \text{si } x < e_{i-1}; \\ N_{i-1}^+ + \frac{n_i(x-e_{i-1})}{a_i}, & \text{si } x \in [e_{i-1}, e_i[; \\ N, & \text{si } x \geq e_k. \end{cases} \quad (1)$$

$$F^+(x) = \begin{cases} 0, & \text{si } x < e_{i-1}; \\ F_{i-1}^+ + \frac{f_i(x-e_{i-1})}{a_i}, & \text{si } x \in [e_{i-1}, e_i[; \\ 1, & \text{si } x \geq e_k. \end{cases} \quad (2)$$

$$N^-(x) = \begin{cases} N, & \text{si } x < e_{i-1}; \\ N_i^- - \frac{n_i(x-e_i)}{a_i}, & \text{si } x \in [e_{i-1}, e_i[; \\ 0, & \text{si } x \geq e_k. \end{cases} \quad (3)$$

$$F^-(x) = \begin{cases} 1, & \text{si } x < e_{i-1}; \\ F_i^- - \frac{f_i(x-e_i)}{a_i}, & \text{si } x \in [e_{i-1}, e_i[; \\ 0, & \text{si } x \geq e_k. \end{cases} \quad (4)$$

► Supposons que l'on cherche la valeur de x pour laquelle $N^+(x)$ ou $F^+(x)$ sont connues. Pour ce faire, on exprime x en fonction de $N^+(x)$ ou $F^+(x)$. De l'équation (1) et (2), on trouve pour $x \in [e_{i-1}, e_i[$:

$$x = e_{i-1} + a_i \frac{[N^+(x) - N_{i-1}^+]}{n_i} = e_{i-1} + a_i \frac{[F^+(x) - F_{i-1}^+]}{f_i}. \quad (5)$$

En d'autre termes, cela revient à travailler avec les fonctions inverses des fonctions $N^+(x)$ et $F^+(x)$.
► Calculons les f_i, N_i^+, N_i^-, F_i^+ et F_i^- pour les deux exemples 1 et 2:

C_i (en kg)	n_i	f_i	N_i^+	N_i^-	F_i^+	F_i^-
[56,58[	5	0.05	5	100	0.05	1
[58,60[	10	0.1	15	95	0.15	0.95
[60,62[	18	0.18	33	85	0.33	0.85
[62,64[	29	0.29	62	67	0.62	0.67
[64,66[	16	0.16	78	38	0.78	0.38
[66,68[	14	0.14	92	22	0.92	0.22
[68,70[	08	0.08	100	8	1	0.08
Total	100	1				

C_i (en kg)	n_i	f_i	N_i^+	N_i^-	F_i^+	F_i^-
[56, 58[	5	0,050	5	100	0,050	1,000
[58, 60[	10	0,100	15	95	0,150	0,950
[60, 62[	18	0,180	33	85	0,330	0,850
[62, 64[	29	0,290	62	67	0,620	0,670
[64, 66[	16	0,160	78	38	0,780	0,380
[66, 70[	22	0,220	100	22	1,000	0,220
	100	1				

V. Représentation graphique des effectifs et fréquences relatives

Les effectifs (resp. les fréquences relatives) sont représentés par un *histogramme*. On distingue deux cas:

Cas 1. Les amplitudes des classes sont égales. Dans ce cas, on positionne les bornes des différentes classes sur l'axe horizontal de l'histogramme, et on positionne les effectifs n_i (resp. les fréquences relatives f_i) sur son axe vertical. Puis on trace pour chaque classe C_i un rectangle dont la largeur est proportionnelle à l'amplitude de la classe et la longueur est proportionnelle à l'effectif n_i (resp. la fréquence relative f_i).

Cas 2. Les amplitudes des classes sont différentes. Dans ce cas, on positionne les bornes des différentes classes sur l'axe horizontal de l'histogramme, et on positionne les effectifs corrigés $n'_i = \frac{n_i}{a_i}$ (resp. les fréquences relatives corrigées $f'_i = \frac{f_i}{a_i}$) sur son axe vertical. Puis on trace pour chaque classe C_i un rectangle dont la largeur est proportionnelle à l'amplitude de la classe et la longueur est proportionnelle à l'effectif corrigé n'_i (resp. la fréquence relative corrigée f'_i).

Calculons les effectifs corrigés et les fréquences relatives corrigées pour l'exemple 2, puis représentons les effectifs et les fréquences relatives pour les deux exemples 1 et 2:

C_i	a_i	n_i	f_i	$n'_i = \frac{n_i}{a_i}$	$f'_i = \frac{f_i}{a_i}$
[56, 58[	2	5	0,050	2,50	0,025
[58, 60[	2	10	0,100	5,00	0,050
[60, 62[	2	18	0,180	9,00	0,090
[62, 64[	2	29	0,290	14,50	0,145
[64, 66[	2	16	0,160	8,00	0,080
[66, 70[	4	22	0,220	5,50	0,055
		100	1		

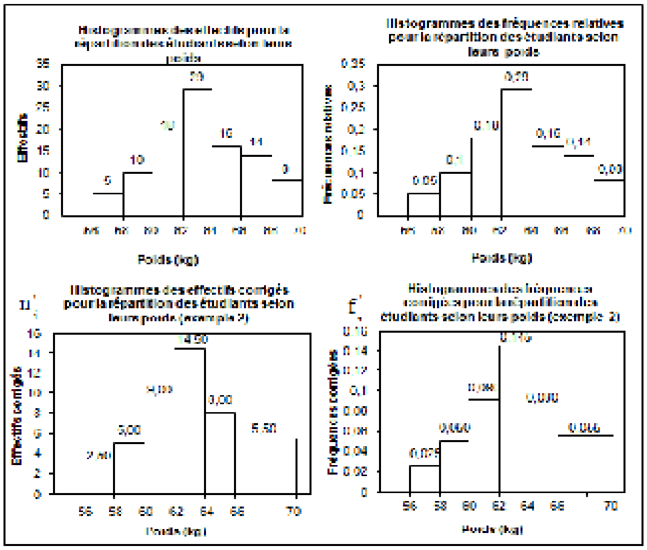
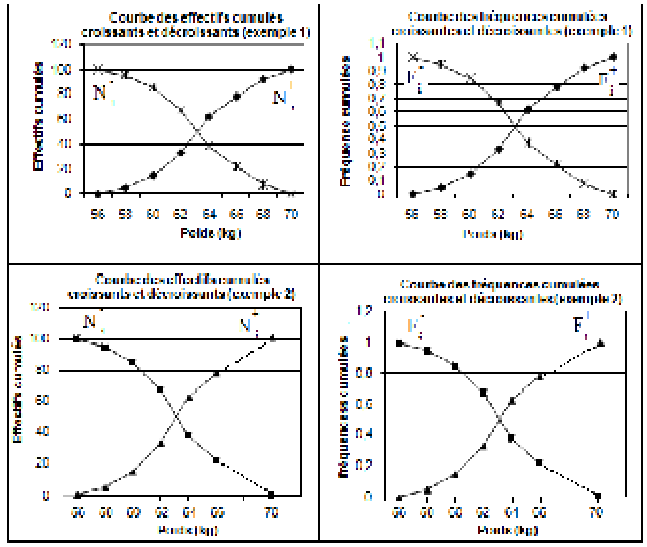


Figure 2



-Figure 3-

VI. Représentation graphique des effectifs cumulés et les fréquences cumulées

► Les effectifs cumulés croissants N_i^+ (resp. les fréquences cumulées croissantes F_i^+) sont représentés par un graphique appelé *courbe cumulative croissante*: sur l'axe horizontal, on positionne les bornes des différentes classes; sur l'axe vertical, on positionne les N_i^+ (resp. les F_i^+). Puis on relie par des segments continus, les points

$$(e_0, 0),\ (e_1, N_1^+), \dots, (e_i, N_i^+), \dots, (e_k, N).$$

$$\text{(resp. } (e_0, 0),\ (e_1, F_1^+), \dots, (e_i, F_i^+), \dots, (e_k, 1)\text{)}.$$

En d'autre termes, on relie les points dont l'abscisse est la borne supérieure de la classe C_i et l'ordonnée est la valeur de F_i^+.

► Les effectifs cumulés décroissants N_i^- (resp. les fréquences cumulées décroissantes F_i^-) sont représentés par un graphique appelé *courbe cumulative décroissante*: sur l'axe horizontal, on positionne les bornes des différentes classes; sur l'axe vertical, on positionne les N_i^- (resp. les F_i^-). Puis on relie par des segments continus, les points

$$(e_0, 1),\ (e_1, N_2^-), \dots, (e_{i-1}, N_i^-), \dots, (e_k, 0).$$

$$\text{(resp. } (e_0, 1),\ (e_1, F_2^-), \dots, (e_{i-1}, F_i^-), \dots, (e_k, 0)\text{)}.$$

En d'autre termes, on relie les points dont l'abscisse est la borne inférieure de la classe C_i et l'ordonnée est la valeur de F_i^-. Représentons les N_i^+, N_i^-, puis les F_i^+ et F_i^- pour l'exemple 1:

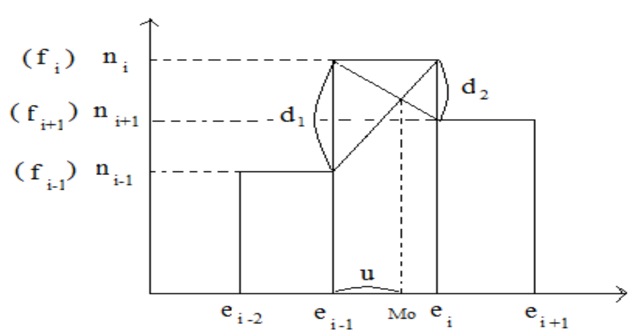
Remarque 1: Le point d'intersection de la courbe des N_i^+ et celle des N_i^- (resp. la courbe des F_i^+ et celle des F_i^-) est le point $(Me, \frac{N}{2})$ (resp. $(Me, 0.5)$).

VII. Caractéristiques de position centrale pour une variable statistique continue

A. *La classe modale et le mode*

La classe correspondant au plus grand effectif (ou à la plus grande fréquence relative) est appelée la *classe modale*. On la note par CMo. Dans les exemples 1 et 2, la classe modale est la classe de poids $CMo = [62, 64[$.

On peu aussi donner une approximation pour le mode. Supposons que la classe modale est la classe $CMo = [e_{i-1}, e_i[$ et que les amplitudes des classes sont égales. Considérons le schéma suivant:



-Figure 4-

Graphiquement on a: $Mo = e_{i-1} + u$. Analytiquement, le mode sera égal à:

$$Mo = e_{i-1} + a_i\left(\frac{d_1}{d_1 + d_2}\right), \tag{6}$$

où a_i et e_{i-1} représentent respectivement l'amplitude et la borne inférieure de la classe modale, et

$$\begin{cases} d_1 = n_i - n_{i-1}, \\ d_2 = n_i - n_{i+1}, \end{cases} \text{ ou } \begin{cases} d_1 = f_i - f_{i-1}, \\ d_2 = f_i - f_{i+1}. \end{cases}$$

Dans le cas où les classes ont des amplitudes différentes, le mode sera toujours calculé avec la formule (6) mais cette fois, on utilise les effectifs corrigés ou les fréquences relatives corrigées pour calculer d_1 et d_2:

$$\begin{cases} d_1 = n'_i - n'_{i-1}, \\ d_2 = n'_i - n'_{i+1}, \end{cases} \text{ ou } \begin{cases} d_1 = f'_i - f'_{i-1}, \\ d_2 = f'_i - f'_{i+1}. \end{cases}$$

Pour l'exemple 1, on a $\max n_i = n_4$. Donc $CMo = [e_{i-1}, e_i[= [e_3, e_4[= [62, 64[$, $a_i = 2$, $d_1 = n_4 - n_3 = 29 - 18 = 11$ et $d_2 = n_4 - n_5 = 29 - 16 = 13$. D'où

$$Mo = e_{i-1} + a_i(\frac{d_1}{d_1 + d_2}) = 62 + 2(\frac{11}{11 + 13}) = 62.92 \ kg.$$

Pour l'exemple 2, les amplitudes des classes ne sont pas égales donc le mode sera calculé comme suit: $\max n'_i = n'_4 = 14.5 \Rightarrow CMo = [e_{i-1}, e_i[= [e_3, e_4[= [62, 64[$, $a_i = 2$, $d_1 = n'_4 - n'_3 = 14.5 - 9 = 5.5$ et $d_2 = n'_4 - n'_5 = 14.5 - 8 = 6.5$. D'où

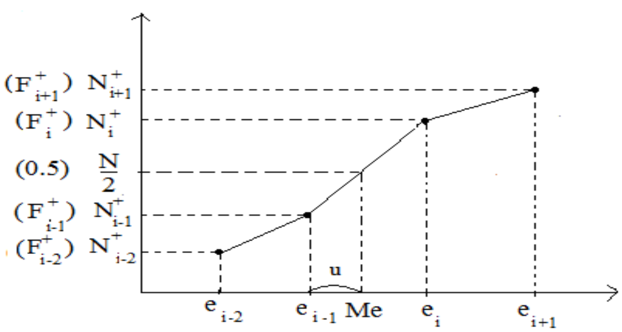
$$Mo = e_{i-1} + a_i(\frac{d_1}{d_1 + d_2}) = 62 + 2(\frac{5.5}{5.5 + 6.5}) = 62.92 \ kg.$$

B. *La classe médiane et la médiane*

► La classe correspondant à la l'effectif cumulé croissant N_i^+ vérifiant

$$N_{i-1}^+ \leq \frac{N}{2} \leq N_i^+ \text{ ou } F_{i-1}^+ \leq 0.5 \leq F_i^+$$

est appelée la *classe médiane*. Elle est notée par CMe. Supposons que $CMe = [e_{i-1}, e_i[$ et considérons le schéma suivant:



-Figure 5-

Graphiquement, la médiane représente l'abscisse du point d'ordonnée $\frac{N}{2}$ dans la courbe des N_i^+ (ou 0.5 dans la courbe des F_i^+): $Me = e_{i-1} + u$.

Analytiquement, la médiane est donnée par la formule suivante:

$$Me = e_{i-1} + a_i(\frac{N/2 - N_{i-1}^+}{n_i}) \text{ ou } Me = e_{i-1} + a_i(\frac{0.5 - F_{i-1}^+}{f_i}).$$

Pour l'exemple 1, la classe médiane est $CMe = [62, 64[$ et la médiane est égale à:

$$Me = 62 + 2(\frac{50 - 33}{29}) = 63.17 \ kg.$$

C. *Les quartiles*

► La classe correspondant à l'effectif cumulé croissant N_i^+ vérifiant

$$N_{i-1}^+ \leq \frac{N}{4} \leq N_i^+ \text{ ou } F_{i-1}^+ \leq 0.25 \leq F_i^+$$

est appelée la *classe du premier quartile*. Elle est notée par CQ_1. On a alors $CQ_1 = [e_{i-1}, e_i[$ et le premier quartile, noté Q_1, est donné par la formule suivante:

$$Q_1 = e_{i-1} + a_i(\frac{\frac{N}{4} - N_{i-1}^+}{n_i}) \text{ ou } Q_1 = e_{i-1} + a_i(\frac{0.25 - F_{i-1}^+}{f_i}).$$

► La classe correspondant à la l'effectif cumulé croissant N_i^+ vérifiant

$$N_{i-1}^+ \leq \frac{N}{2} \leq N_i^+ \text{ ou } F_{i-1}^+ \leq 0.5 \leq F_i^+$$

est appelée la *classe du deuxième quartile*. Elle est notée par CQ_2. On a $CQ_2 = CMe$ et le deuxième quartile $Q_2 = Me$.

► La classe correspondant à la l'effectif cumulé croissant N_i^+ vérifiant

$$N_{i-1}^+ \leq \frac{3N}{4} \leq N_i^+ \text{ ou } F_{i-1}^+ \leq 0.75 \leq F_i^+$$

est appelée la *classe du troisième quartile*. Elle est notée par CQ_3. On a alors $CQ_3 = [e_{i-1}, e_i[$ et le troisième quartile, noté Q_3, est donné par la formule suivante:

$$Q_3 = e_{i-1} + a_i(\frac{\frac{3N}{4} - N_{i-1}^+}{n_i}) \text{ ou } Q_3 = e_{i-1} + a_i(\frac{0.75 - F_{i-1}^+}{f_i}).$$

Remarque 2: On peut facilement déduire la médiane, le premier quartile et le troisième quartile à partir de la formule (5). En effet, il suffit de poser $N^+(x) = \frac{N}{2}$ ou $F^+(x) = 0.5$ pour trouver la formule de Me, $N^+(x) = \frac{N}{4}$ ou $F^+(x) = 0.25$, pour trouver la formule de Q_1 et $N^+(x) = \frac{3N}{4}$ ou $F^+(x) = 0.75$, pour trouver la formule de Q_3.

D. *Les moyennes*

Les différentes moyennes sont données par les formules suivantes:

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{k} n_i x_i, \ Q = \sqrt{\frac{1}{N}\sum_{i=1}^{k} n_i x_i^2},$$

$$G = e^{\frac{1}{N}\sum_{i=1}^{k} n_i ln(x_i)}, \ H = \frac{N}{\sum_{i=1}^{k} \frac{n_i}{x_i}}.$$

On a

$$H \leq G \leq \bar{X} \leq Q.$$

VIII. CARACTÉRISTIQUES DE DISPERSION

A. *La variance et l'écart-type*

► La variance, notée $V(X)$ est donnée par la formule suivante:

$$V(X) = \frac{1}{N}\sum_{i=1}^{k} n_i(x_i - \bar{X})^2 = (\frac{1}{N}\sum_{i=1}^{k} n_i x_i^2) - \bar{X}^2.$$

► L'écart-type est alors:

$$\sigma_X = \sqrt{V(X)}.$$

B. *L'étendue et l'écart interquartile*

L'étendue et l'écart interquartile sont donnés par:

$$E = e_k - e_0 \text{ et } EIQ = Q_3 - Q_1.$$

IX. Exercices de TD

EXERCICE 1:

Le tableau suivant représente les données brutes recueillies sur les âges de 100 employés d'une entreprise:

60	39	23	30	29	26	29	41	40	32
63	22	32	52	46	35	25	28	33	33
20	25	42	34	29	43	41	31	30	36
58	21	24	55	51	28	18	40	44	38
32	21	30	31	25	49	31	26	33	36
43	34	35	22	33	38	34	34	33	34
23	26	57	23	26	36	39	31	35	34
34	51	40	50	35	45	28	36	32	39
26	48	17	45	45	25	25	30	36	30
43	25	27	21	53	25	38	33	37	33

1) Trier les données par ordre croissant;
2) Regrouper les données en 10 classes d'amplitude 5 ([17,22[,...), puis dresser un tableau statistique qui contiendra: les effectifs, les fréquences relatives ainsi que les fréquences cumulées.
3) Représenter graphiquement les effectifs par un histogramme.
4) Représenter dans le même repère les fréquences cumulées croissantes et décroissantes.
5) Calculer l'âge moyen des employés.
6) Calculer l'écart type de cette distribution statistique.
7) Calculer le mode et la médiane graphiquement et analytiquement.
8) Calculer le premier et le troisième quartile, puis déterminer l'écart interquartile.

EXERCICE 2: Un tour automatique produit des axes cylindriques. Les diamètres en $(1/10$ de mm), mesurés sur un lot de 1000 pièces ont donné les résultats suivants :

C_i	[244, 246[	[246, 248[	[248, 249[	[249, 250[
n_i	11	132	152	200
C_i	[250, 251[	[251, 252[	[252, 254[	[254, 258[
n_i	194	158	139	14

1) Tracer un histogramme du caractère X (diamètre mesurée).
2) Calculer le mode et la médiane.
3) Calculer la moyenne, la variance et l'écart type.

X. Exercices supplémentaires de TD

EXERCICE 3: Une entreprise désire importer un certain type de lampes électriques. Les données concernant chaque type de lampes sont reportées dans les tableaux suivants

Type 1:

C_i	[0,3[	[3,6[	[6,9[	[9,12[	[12,15[	[15,18[	[18,21[
n_i	2	4	8	12	17	6	1

Type 2:

C_i	[0,3[	[3,5[	[5,10[	[10,15[	[15,18[	[18,21[
n_i	1	2	7	20	15	5

où C_i représentent les classes des durées de vie (en mois) des lampes et n_i le nombre de lampes ayant une durée de vie dans la classe C_i.
1) Quel est le nombre de lampes ayant une durée de vie inférieure à 16 mois?
2) Quelle est la durée de vie correspondant au plus grand nombre de lampes pour les deux types? (Calculer le mode pour les deux séries).
3) Quelle est la durée de vie médiane pour chaque type?
4) Quel est la série qui a la plus forte dispersion? (comparer les coefficients de variation des deux séries).

EXERCICE 4: Considérons les observations x_i et y_i vérifiant:

$$\sum_{i=1}^{50} x_i = 300, \ \sum_{i=1}^{50} x_i^2 = 1950;$$

$$\sum_{i=1}^{50} y_i = 280, \ \sum_{i=1}^{50} y_i^2 = 2100.$$

1) Déduire la moyenne arithmétique et l'écart type de chaque série.
2) Quelle est la série qui a la plus forte dispersion autour de la moyenne.
3) Si l'on regroupe les deux séries en une seule série, quel serait la moyenne arithmétique de la série obtenue?

EXERCICE 5: Considérons la variable statistique X telle que: $\bar{X} = 50, n = 30, \sigma = 8$. Trouver $\sum_{i=1}^{n} x_i^2$.

Chapitre 4: Variable Statistique Double et Ajustement Linéaire

I. Plan du chapitre

Cours 1: Variable statistique double et ajustement linéaire (Durée=1h30mn)

- ► Exemples introductifs.
- ► Moyenne et variance.
- ► Covariance de deux variables statistiques.
- ► Ajustement linéaire avec la méthode des moindres carrés.
- ► Qualité de l'ajustement: coefficient de corrélation.

II. Exemples introductifs

Exemple 1: On donne le tableau d'observations suivant :

x_i (en mm)	2	4	6	8
y_i $(m^2.°C)$	0.83	1.34	1.63	2.29
x_i (en mm)	10	12	15	20
y_i $(m^2.°C)$	2.44	2.93	4.06	4.48

où x représente l'épaisseur d'un mur et y sa résistance thermique.
1) Représenter ce tableau de données par un nuage de points.
2) Ajuster linéairement y en x et déterminer le coefficient de corrélation linéaire.
3) Représenter la droite de regression de y en x.
4) Si l'on construit un mur d'épaisseur $x = 22mm$, alors combien sera sa résistance thermique?

Exemple 2: On donne le tableau d'observations suivant :

x_i	1	2	3
y_i	1	2	2

1) Représenter ce tableau de données par un nuage de points.
2) Ajuster linéairement y en x et déterminer le coefficient de corrélation linéaire.
3) Représenter graphiquement la droite de regression de y en x.

Problématiques:
► Existe t-il une relation linéaire entre la variable x et la variable y?
► Trouver l'equation de la droite dite de *regression* ou *d'ajustement* liant x et y.
► Mesurer la qualité de l'ajustement: est ce que la droite de regression est la meilleure relation mathématique qui lie x et y ou bien existe t-il une autre relation non linéaire liant x et y (*regression non linéaire*)?
► Une fois la relation mathématique trouvée, on peut prévoir la valeur de y pour une valeur donnée de x.

III. Tableau de données et nuage de points

Dans de nombreuse situations pratiques, on constate qu'il existe un lien entre deux caractères d'une population, par exemple, entre le poids et la taille, entre la note d'un étudiant et le volume horaire de son travail, etc.
On définit alors une série statistique à deux variables x et y, prenant les valeurs:
► $x \rightsquigarrow x_1, x_2, \dots, x_n$;
► $y \rightsquigarrow y_1, y_2, \dots, y_n$.

A. *Tableau de données*

Un tableau de donnée pour une série statistique à deux variables peut être présenté sous deux formes différentes:
Forme 1:

x_i/y_i	y_1	y_2	$\dots$	y_p
x_1	n_{11}	n_{12}	$\dots$	n_{1p}
x_2	n_{21}	n_{22}	$\dots$	n_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\dots$	n_{kp}

où n_{ij} représente le nombre d'individus ayant la valeur x_i du caractère x et la valeur y_i du caractère y. On a alors:
► L'effectif total de la population:

$$N = \sum_{i=1}^{k} \sum_{j=1}^{p} n_{ij}.$$

Forme 2:

x_i	x_1	x_2	$\dots$	x_n
y_i	y_1	y_2	$\dots$	y_n

où n est l'effectif total de la population et x_j, y_j représentent les valeurs des caractères x et y respectivement, observées pour chaque individu i.
Dans ce qui suit on ne s'intéressera qu'à la forme 2.

B. *Nuage de points*

Pour chaque couple d'observation $(x_i, y_i), i = \overline{1, n}$, on lui associe un point M_i de coordonnées (x_i, y_i) dans un repère orthonormé.

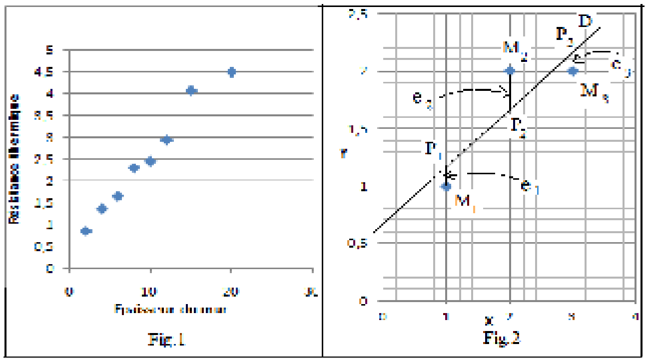


Fig.1 Fig.2

► Le nuage de points correspondant à l'exemple 1 est montré dans la Figure 1.
► Le nuage de points correspondant à l'exemple 2 est montré dans la Figure 2.

IV. Moyenne, variance et covariance

► Moyenne de x:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

► Variance de x:

$$V(x) = \sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2.$$

► Moyenne de y:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

► Variance de y:

$$V(y) = \sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2.$$

► Covariance de x et y:

$$\text{Cov}(x,y) = \sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}.$$

V. Ajustement linéaire par la méthode des moindres carrés (Gauss)

Pour mieux comprendre la méthode des moindres carrés, travaillons sur l'exemple 2: soit D une droite d'équation $y = ax + b$, où a et b sont des coefficients à déterminer. La droite D qui ***minimise*** la quantité suivante:

$$\sum_{i=1}^{n}(M_iP_i)^2 = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}[y_i - (ax_i + b_i)]^2$$

est dite *droite d'ajustement de y en x* ou *droite de regression de y en x*.
Pour résoudre ce problème de minimisation, on définit la fonction Q de la manière suivante:

$$Q(a,b) = \sum_{i=1}^{n}(y_i - ax_i - b_i)^2.$$

Les valeurs de a et b qui minimisent la fonction $Q(a,b)$ sont solution du système linéaire à deux inconnues suivant:

$$\begin{cases} \frac{\partial Q(a,b)}{\partial a} = \sum_{i=1}^{n} x_i(y_i - ax_i - b_i) = 0, \\ \frac{\partial Q(a,b)}{\partial b} = \sum_{i=1}^{n}(y_i - ax_i - b_i) = 0. \end{cases}$$

En développant les expressions, on obtient

$$\sum_{i=1}^{n} y_i = a\sum_{i=1}^{n} x_i + nb, \quad (1)$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i. \quad (2)$$

Remarquons que l'équation (1) peut être écrite sous la forme:

$$\bar{y} = a\bar{x} + b.$$

D'autre part, si on multiplie la première équation par $\bar{x}$ et on la soustrait de la deuxième, on obtient:

$$a = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}.$$

En divisant le numérateur et le dénominateur par n, on obtient:

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2} = \frac{\text{Cov}(x,y)}{V(x)} = \frac{\sigma_{xy}}{\sigma_x^2}.$$

D'où

$$\begin{cases} a = \frac{\sigma_{xy}}{\sigma_x^2} \\ \text{et} \\ b = \bar{y} - a\bar{x}. \end{cases}$$

VI. Coefficient de corrélation linéaire

Afin de mesurer la qualité de l'ajustement effectué en utilisant la méthode des moindres carrés, on calcule le *coefficient de correlation linéaire*, ρ, de la manière suivante:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Remarque 1: on a: $-1 \leq \rho \leq 1$.
► Si $\rho = 1$ ou $\rho = -1$, alors tous les points se trouvent sur la droite d'ajustement. On dit alors que l'ajustement est parfait.
► Si $\rho = 0$, alors les points sont les plus loins possibles de la droite d'ajustement. Donc la liaison n'est pas linéaire.
► Plus ρ s'approche de -1 ou 1, plus l'ajustement par une droite est bon.
► Plus ρ s'approche de 0, plus l'ajustement par une droite est mauvais.
► Dans la pratique, on dit que l'ajustement linéaire est bon si

$$\rho \geq 0.7 \text{ ou } \rho \leq -0.7.$$

Revenons maintenant à l'exemple 2. Afin de trouver la droite qui est la mieux ajustée aux points du nuage, on dresse le tableau suivant:

x_i	y_i	x_i^2	y_i^2	x_iy_i
1	1	1	1	1
2	2	4	4	4
3	2	9	4	6
6	5	14	9	11

On a alors:
► $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{6}{3} = 2.$
► $\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{14}{3} - 2^2 = \frac{2}{3} \Rightarrow \sigma_x = 0.82.$
► $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{5}{3} = 1.67.$
► $\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2 = \frac{9}{3} - \frac{25}{9} = \frac{2}{9} \Rightarrow \sigma_y = 0.47.$
► $\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y} = \frac{11}{3} - 2 \times \frac{5}{3} = \frac{1}{3} = 0.33.$
D'où

$$\begin{cases} a = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{1/3}{2/3} = \frac{1}{2} = 0.5 \\ \text{et} \\ b = \bar{y} - a\bar{x} = \frac{5}{3} - \frac{1}{2} \times 2 = \frac{2}{3} = 0.67. \end{cases}$$

Donc l'équation de la droite de regression de y en x est:

$$y = \frac{1}{2}x + \frac{2}{3}.$$

Le coefficient de corrélation est donné par:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{0.33}{0.82 \times 0.47} = 0.86.$$

On déduit alors que l'ajustement linéaire effectué est bon.

VII. EXERCICES DE TD

EXERCICE 1: On donne le tableau d'observations suivant :

h_i (heures)	10	14	17	20
t_i (%)	9.15	23.25	37.04	51.7

où t_i représentent les heures de communication, par le ministère de l'intérieur, des résultats concernant les taux de participation aux élections présidentielles du 17 avril 2014 en Algérie (D'après le journal El Watan du 25 avril 2014) et t_i les taux de participation aux heures h_i.
1) Représenter ce tableau de données par un nuage de points.
2) Trouver la relation mathématique liant t et h et déterminer le coefficient de corrélation linéaire.
3) Représenter graphiquement la droite de regression de t en h.
4) Quelle est le taux de participation attendu à 21h.

VIII. EXERCICES SUPPLÉMENTAIRES DE TD

EXERCICE 2: On donne le tableau d'observations suivant :

x_i (en mm)	2	4	6	8
y_i $(m^2.°C)$	0.83	1.34	1.63	2.29

x_i (en mm)	10	12	15	20
y_i $(m^2.°C)$	2.44	2.93	4.06	4.48

où x représente l'épaisseur d'un mur et y sa résistance thermique.
1) Représenter ce tableau de données par un nuage de points.
2) Ajuster linéairement y en x et déterminer le coefficient de corrélation linéaire.
3) Représenter graphiquement la droite de regression de y en x.
4) Si l'on construit un mur d'épaisseur $x = 22mm$, alors combien sera sa résistance thermique?

EXERCICE 3: On donne le tableau d'observations suivant:

t_i	11.9	14.5	15.5	17.3	17.4	17.7	19	19.2
x_i	11.1	14.2	15.1	17.9	17.1	17.1	18.3	19.2

1) Ajuster linéairement x en t et déterminer le coefficient de corrélation linéaire.

Chapitre 5: Analyse Combinatoire et Théorie Fondamentale des Probabilités

I. Plan du chapitre

Cours 1: Analyse combinatoire (Durée=1h30mn)

- ► Exemples Introductifs.
- ► Les p-listes.
- ► Les arrangements.
- ► Les combinaisons et leurs propriétés.
- ► Triangle de Pascal et Binôme de Newton.

Cours 2: Théorie fondamentale des probabilités (Durée=3h)

- ► Notion d'expérience aléatoire.
- ► Vocabulaire des événements.
- ► Axiomes du calcul des probabilités.
- ► Probabilités conditionnelles et théorème de Bayes.

II. Exemples introductifs

Exemple 1: Combien de mots de passe de deux caractères peut-on former avec les caractères '1', '2' et 'A'?

Exemple 2: Les étudiants de notre université organisent un tournoi de football. Notre université dispose de cinq cités universitaires A, B, C, D et E. Chaque cité universitaire est représentée par une équipe. A la fin du tournoi, la première équipe gagnera un prix de 100000,00 DA, la deuxième un prix de 50000,00 DA et la troisième un prix de 30000,00 DA. Pour qu'à la fin du tournoi les équipes gagnantes trouvent leur listes déjà prêtes, il nous faut préparer au préalable toutes les listes possibles. Combien y en a-t-il?

Exemple 3: Si à la fin du tournoi du football, les trois premières équipes gagnent un même prix égal à 100000,00 DA. Combien de listes y en a-t-il?

Réponses:

Exemple 1: (voir Figure 1).

Les mots de passe que l'on peut former sont comme suit:

$$11, 12, 1A, 21, 22, 2A, A1, A2, AA.$$

► La répétition existe: un chiffre peut figurer deux fois dans un mot de passe (11 a un sens).

► L'ordre est important: le mot de passe 1A est différent du mot de passe A1.

Le nombre de mots de passe que l'on peut former est le nombre de listes de 2 éléments choisis parmi 3: $3^2 = 9$.

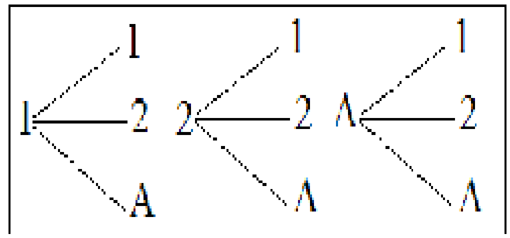


Figure 1. L'arbre des possibilités pour l'exemple 1

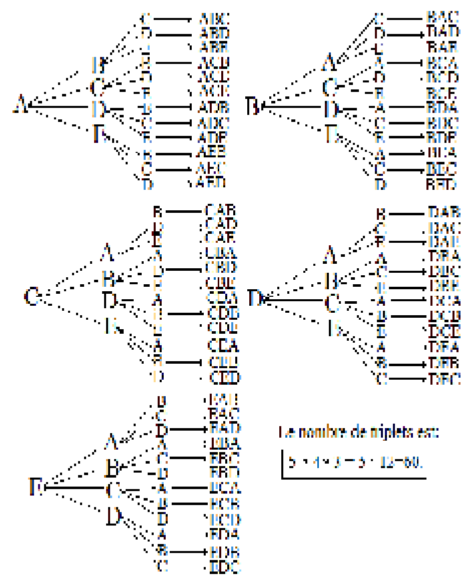


Figure 2. L'arbre des possibilités pour l'exemple 2: $p = 3$, $n = 5$

► D'une façon générale, le nombre de p-listes d'un ensemble E à n éléments est: $\mathbf{n^p}$.

Exemple 2: (voir Figure 2).

► La répétition n'existe pas: une équipe ne peut pas figurer deux fois dans une liste.

► L'ordre est important: la liste ABC est différente de la liste ACB.

Le nombre de listes de gagnants que l'on peut former est le nombre d'arrangements de 3 éléments choisis parmi 5:

$$A_5^3 = \frac{5!}{(5-3)!} = \frac{5 \times 4 \times 3 \times 2!}{2!} = 5 \times 4 \times 3 = 60.$$

Supposons maintenant que notre université possède unique-

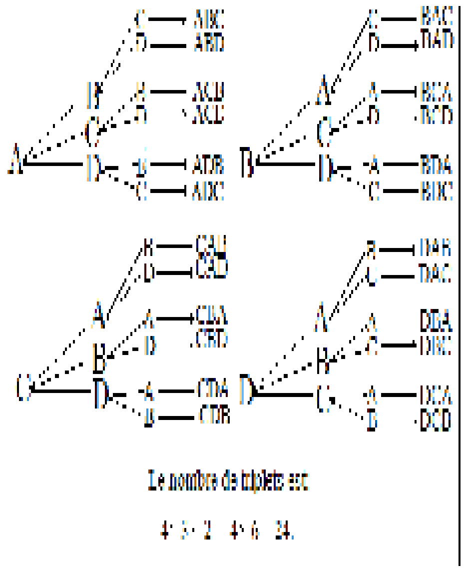


Figure 3. L'arbre des possibilités pour l'exemple 2: $p = 3$, $n = 4$

ment 4 cités universitaires A, B, C et D, alors quel serait le nombre de listes des 3 équipes gagnantes?

Le nombre de listes des équipes gagnantes que l'on peut former est le nombre d'arrangements de 3 éléments choisis parmi 4:

$$A_4^3 = \frac{4!}{(4-3)!} = \frac{4 \times 3 \times 2 \times 1}{1!} = 24.$$

L'arbre des possibilités pour ce cas est représenté dans la Figure 3.

D'une façon générale, le nombre d'arrangements de p éléments choisis parmi n éléments est:

$$A_n^p = n(n-1)\cdots(n-p+1) = \frac{n!}{(n-p)!}.$$

$A_n^p = 0$ pour tout $p > n$.

Cas particulier: Permutation

Si $p = n$, alors

$$A_n^p = P_n = n!.$$

► Le nombre de façons d'arranger 3 livres dans une étagère est $P_3 = 3! = 3 \times 2 = 6$.

► Le nombre de manières de s'asseoir de 5 personnes autour d'une table est: $P_5 = 5! = 5 \times 4 \times 3 \times 2 = 120$.

Exemple 3:

► La répétition n'existe pas: une équipe ne peut pas figurer deux fois dans une liste.

► L'ordre n'est pas important: la liste ABC est ACB sont identiques car les trois équipes auront le même prix.

Le nombre de listes des équipes gagnantes que l'on peut former est le nombre de combinaisons de 3 éléments choisis parmi 5:

$$C_5^3 = \frac{5!}{3!(5-3)!} = 10.$$

D'une façon générale, le nombre de combinaisons de p éléments choisis parmi n éléments est:

$$C_n^p = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!}.$$

III. Propriétés des combinaisons

Propriété 1:

$$C_n^0 = \frac{n!}{0!(n-0)!} = \frac{n!}{n!} = 1.$$

Propriété 2:

$$C_n^n = \frac{n!}{n!(n-n)!} = \frac{n!}{n!} = 1.$$

Propriété 3:

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

Propriété 4:

$$C_{n-1}^{p-1} + C_{n-1}^p = C_n^p.$$

En effet,

$$\begin{aligned}
C_{n-1}^{p-1} + C_{n-1}^p &= \frac{(n-1)!}{(p-1)!((n-1)-(p-1))!} + \frac{(n-1)!}{p!((n-1)-p)!} \\
&= \frac{(n-1)!}{(p-1)!(n-p)!} + \frac{(n-1)!}{p!(n-p-1)!} \\
&= \frac{p(n-1)!}{p(p-1)!(n-p)!} + \frac{(n-1)!(n-p)}{p!(n-p-1)!(n-p)} \\
&= \frac{p(n-1)!}{p!(n-p)!} + \frac{(n-1)!(n-p)}{p!(n-p)!} \\
&= \frac{(p+n-p)(n-1)!}{p!(n-p)!} \\
&= \frac{n(n-1)!}{p!(n-p)!} \\
&= \frac{n!}{p!(n-p)!} = C_n^p.
\end{aligned}$$

A. *Triangle de Pascal et binôme de Newton*

n/p	0	1	2	3	4	5
0	1					
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

► On construit d'abord la première colonne du triangle de Pascal en utilisant la première propriété des combinaisons ($C_n^0 = 1$).

► On construit la diagonale du triangle de Pascal en utilisant la deuxième propriété des combinaisons ($C_n^n = 1$).

► On construit les autres éléments du triangle de Pascal en utilisant la troisième propriété des combinaisons ($C_{n-1}^{p-1} + C_{n-1}^p = C_n^p$).

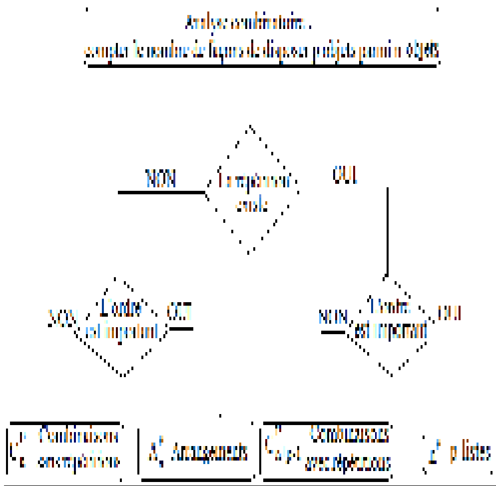


Figure 4. Organigramme récapitulatif

Binôme de Newton

$(a+b)^n = \sum_{p=0}^{n} C_n^p a^p b^{n-p}$. Pour $n = 5$:

$$\begin{aligned}(a+b)^5 &= \sum_{p=0}^{5} C_n^p a^p b^{n-p} \\ &= C_5^0 a^0 b^5 + C_5^1 a^1 b^4 + C_5^2 a^2 b^3 \\ &\quad + C_5^3 a^3 b^2 + C_5^4 a^4 b^1 + C_5^5 a^5 b^0 \\ &= b^5 + 5ab^4 + 10a^2b^3 \\ &\quad + 10a^3b^2 + 5a^4b + a^5.\end{aligned}$$

Résumé: (voir Figure 4)

IV. THÉORIE FONDAMENTALE DES PROBABILITÉS

A. *Notion d'expérience aléatoire*

Définition 1:

► Une expérience ayant un nombre fini d'issues possibles est appelée expérience aléatoire s'il est impossible de savoir à l'avance quelle en sera l'issue.

► L'ensemble de toutes les issues possibles est appelé l'univers des possibilités associé à cette expérience; il est généralement noté Ω.

► Chaque sous ensemble de Ω contenant un seul élément, c'est à dire chaque issue possible est appelée événement élémentaire.

Remarque 1: Si on note chaque issue possible $w_1, \dots, w_n$ alors $\Omega = \{w_1, \dots, w_n\}$ et chaque $\{w_i\}$ est un événement élémentaire.

B. *Vocabulaire des événements*

Définition 2: Soit E une expérience aléatoire et Ω l'univers des possibilités associé à cette expérience.

► L'ensemble de toutes les parties de Ω, $\mathcal{P}(\Omega)$, est l'ensemble des événements liés à Ω.

► Ω est l'événement certain.

► $\emptyset$ est l'événement impossible.

Exemple 4:

► L'expérience qui consiste à jeter un dé est une expérience aléatoire car on ne pas pas prévoir les résultats à l'avance.

► $\Omega = \{1, 2, 3, 4, 5, 6\}$.

► Les événements élémentaires sont: $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ et $\{6\}$.

► A: "avoir le chiffre 7" est un événement impossible ($A = \emptyset$).

► B: "avoir un chiffre compris entre 1 et 6" est un événement certain ($B = \Omega$).

Exemple 5: L'expérience qui consiste à jeter une pièce de monnaie est une expérience aléatoire car on ne peut pas prévoir les résultats au préalable.

► $\Omega = \{P, F\}$ (P:Pile, F:Face).

► A: "avoir pile" est un événement élémentaire ($A = \{P\}$).

► B: "avoir pile et face" est un événement impossible ($B = \emptyset$).

► C: "avoir pile ou face" est un événement certain ($C = \Omega$).

Événement contraire:

Soit A un événement lié à une expérience aléatoire E d'univers associé Ω. L'événement contraire est appelé complémentaire de A ou « non A ». Il est noté $\bar{A}$. A est donc une partie de Ω.

► $w \in \bar{A} \Rightarrow w \in C_A = \{w_i \in \Omega : w_i \notin A\}$.

► $\bar{A}$ se réalise si A ne se réalise pas.

Exemple 6: E:"jeter un dé", A:"avoir un chiffre inférieur à 2", B:"avoir un chiffre supérieur ou égal à 2". On a:
$A = \{1\}$ et $B = \{2, 3, 4, 5, 6\}$. Donc $B = \bar{A}$.

C. *Composition d'événements*

a) **Événement** $A \cup B$

► La loi $\cup$ dans $\mathcal{P}(\Omega)$ correspond à l'emploi du « ou » entre deux événements.

► L'événement $A \cup B$ se réalise si l'un des deux événements au moins se réalise.

► $w \in A \cup B \Rightarrow w \in A$ ou $w \in B$.

Exemple 7: E:"jeter un dé"; A:"avoir un chiffre pair", B:"avoir un chiffre impair". Trouver $A \cup B$?
$A \cup B$: " avoir un chiffre pair ou impair".
$\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 4, 6\}$, $B = \{1, 3, 5\}$. D'où $A \cup B = \{1, 2, 3, 4, 5, 6\} = \Omega$.

b) **Événement** $A \cap B$

► La loi $\cap$ dans $\mathcal{P}(\Omega)$ correspond à l'emploi du « et » entre deux événements.

► L'événement se réalise si A et B se réalisent en même temps.

► $w \in A \cap B \Rightarrow w \in A$ et $w \in B$.

Exemple 8: E:"jeter un dé"; A:"avoir un chiffre pair", B:"avoir un chiffre supérieur ou égal 3". Trouver $A \cap B$?
$A \cap B$: " avoir un chiffre pair supérieur ou égal à 3".
$\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 4, 6\}$, $B = \{3, 4, 5, 6\}$. D'où $A \cap B = \{4, 6\}$.

Remarque 2: On dit que deux événements A et B sont incompatibles s'il ne peuvent pas se réaliser en même temps: $A \cap B = \emptyset$. Par exemple, E:"jeter une pièce de monnaie"; A:"avoir pile"; B:"avoir face". $A = \{\text{Pile}\}$, $B = \{\text{Face}\}$; $A \cap B = \emptyset$, donc A et B sont incompatibles.

c) **Différence de deux événements** $A \setminus B$

► $w \in A \setminus B \Rightarrow w \in A$ et $w \notin B$.

d) **Différence symétrique de deux événements** $A\Delta B$
► $w \in A\Delta B \Rightarrow w \in A \cap \bar{B}$ ou $w \in \bar{A} \cap B$.
Lois de De Morgan
► $\overline{A \cap B} = \bar{A} \cup \bar{B}$;
► $\overline{A \cup B} = \bar{A} \cap \bar{B}$.

D. *Axiomes du calcul des probabilités*

Soit E une expérience aléatoire et Ω son univers associé. On appelle probabilité, notée P, toute application de l'ensemble des événements $\mathcal{P}(\Omega)$ dans $\mathbb{R}$ vérifiant les trois axiomes suivants (Axiomes de Kolmogorov):
A1: $\forall A \in \mathcal{P}(\Omega),\ 0 \leq P(A) \leq 1$.
A2: $P(\Omega) = 1$.
A3: si $A \cap B = \emptyset$, alors $P(A \cup B) = P(A) + P(B)$.
Conséquences
► $P(\bar{A}) = 1 - P(A)$;
► $P(\emptyset) = 0$;
► $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
Cas particulier important: l'équiprobabilité

$$P(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{\text{Nombre de cas favarables}}{\text{Nombre de cas possibles}}.$$

V. Probabilité conditionnelle et théorème de Bayes

Probabilité conditionnelle

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

Indépendance de deux événements
► On dit que A et B sont deux événements indépendants si

$$P(A \cap B) = P(A)P(B).$$

►**Théorème de Bayes pour deux événements**
On a $P(A \cap B) = P(A/B)P(B) = P(B/A)P(A)$. D'où

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)}.$$

Cas général: Si $B_1, B_2, \ldots, B_n$ est un système complet d'événements (les événements B_i sont incompatibles deux à deux et leurs union est égale à Ω: $B_i \cap B_j \neq \emptyset$, $i \neq j$ et $\bigcup_{i=1}^n B_i = \Omega$) et si A est un événement quelconque, alors
► **Formule des probabilités totales**

$$P(A) = \sum_{k=1}^{n} P(A/B_k)P(B_k).$$

► **Formule de Bayes**

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{\sum_{k=1}^n P(A/B_k)P(B_k)}.$$

VI. Exercices de TD

EXERCICE 1: Les mots de passe de la carte magnétique de de la poste d'Algérie sont composés de 4 chiffres. Trouver le nombre de mots de passe possibles? Supposons maintenant que les quatres chiffres des mots de passe sont tous différents. Trouver le nombre de mots de passe possibles?
EXERCICE 2: Nous disposons de 11 composants électroniques: 4 sont de bonne qualité, 5 de qualité moyenne et 2 de mauvaise qualité.
1) Trouver le nombre de lots possibles de 3 composants choisi parmi 11.
2) Trouver le nombre de lots possibles de 3 composants parmi 11, tels que chaque lot contient au moins un composant de mauvaise qualité.
3) On choisit au hasard un lot de 3 composants. Quelle est la probabilité que l'on trouve dans le lot choisi au moins un composant de mauvaise qualité?
EXERCICE 3: Calculer $(a + b)^7$ puis déduire 2^7.
EXERCICE 4: Les étudiants d'une section composée de deux groupes (groupe A et groupe B) ont passé un examen. Certains d'entre eux ont réussi, d'autres ont échoué. Les résultats faisant état de la situation ont été reportés dans le tableau suivant:

	Groupe A	Groupe B
Admis	20	21
Ajournés	5	9

Si un étudiant est choisi au hasard dans cette section, quelle est la probabilité que:
a) il appartient au groupe A.
b) il est admis.
c) il appartient au groupe B et il est admis.
d) il appartient au groupe A et il est ajourné.
e) il est ajourné sachant qu'il appartient au groupe A.
EXERCICE 5: Une machine électronique s'arrête de fonctionner quand s'arrête l'un des composants A ou B. La probabilité que la machine s'arrête est 0,6. La probabilité que le composant B s'arrête est 0,4. Après l'observation de l'arrêt de B, la probabilité que A s'arrête aussi est 0,2. Calculer la probabilité que A s'arrête.
EXERCICE 6: Les pourcentages de production de composants électroniques de trois machines M1, M2 et M3 sont respectivement 50%, 30% et 20% et les pourcentages de déchet sont respectivement 3%, 4% et 5%. Un composant électronique est choisi au hasard, on constate qu'il est défectueux. Alors quelle est la probabilité qu'il soit défectueux et quelle est la probabilité qu'il provienne de la machine M2?

VII. Exercices supplémentaires de TD

EXERCICE 7: Un groupe de 30 étudiants participe à un jeu de questions-réponses. Une question a été posée et 14 d'entre eux connaissent la réponse.
1) Trouver le nombre de groupes possibles de 7 étudiants choisis parmi 30.
2) Trouver le nombre de groupes possibles de 7 étudiants où au moins l'un d'entre eux connaisse la réponse.
3) Sept étudiants seront choisis au hasard pour aller répondre à cette question. Quelle est la probabilité qu'au moins l'un des étudiants choisis connaisse la réponse?
EXERCICE 8: Un bloc opératoire est alimenté en électricité quand l'un ou l'autre des générateurs A ou B fonctionne. La probabilité que A fonctionne est $P(A) = 0.7$. La probabilité que B fonctionne est $P(B) = 0.8$. La probabilité que le bloc soit alimenté en électricité est égale à 0.9.

1) Quelle est la probabilité que les deux générateurs fonctionnent en même temps?
2) Si on sait que le générateur B fonctionne, quelle est la probabilité que le générateur A fonctionne aussi?
3) Si on sait que le générateur A fonctionne, quelle est la probabilité que le générateur B fonctionne aussi?
4) Quelle est la probabilité que le générateur B soit le seul à fonctionner?
5) Si on sait que le générateur B fonctionne, alors quelle serait la probabilité que le générateur A soit en panne?

Chapitre 6: Variables Aléatoires et Lois de Probabilités

I. Plan du chapitre

Cours 1: Variable aléatoire discrète (Durée=3h)

- ► Exemples Introductifs.
- ► Notion de variable aléatoire discrète.
- ► Loi de probabilité et sa représentation graphique.
- ► Fonction de répartition et sa représentation graphique.
- ► Espérance mathématique, variance et écart-type.
- ► Lois discrètes usuelles: loi de Bernoulli, loi binômiale, loi géométrique, loi de Poisson.

Cours 2: Variable aléatoire continue (Durée=3h)

- ► Notion de variable aléatoire continue.
- ► Loi de probabilité et sa représentation graphique.
- ► Fonction de répartition et sa représentation graphique.
- ► Espérance mathématique, variance et écart-type.
- ► Lois continues usuelles: loi uniforme, loi exponentielle, loi normale.

II. Exemples introductifs

Exemple 1: Deux étudiants A et B décident de jouer au jeu pile ou face: on lance une seule fois une pièce de monnaie non truquée de 20 dinars (pile représente 20 et face le lion). Si on obtient pile, alors c'est l'étudiant A qui gagne, dans le cas contraire c'est l'étudiant B qui gagne. Soit X la variable aléatoire défini comme suit: $X = 1$, si l'étudiant A gagne et $X = 0$, si ce dernier perd. Trouver la loi de probabilité de X.

Exemple 2: Une urne contient $m_1 = 2$ boules blanches et $m_2 = 3$ boules noires. On tire une boule au hasard. On définit la variable aléatoire X de la façon suivante: $X = 1$ si la boule tirée est blanche et $X = 0$, sinon (si la boule tirée est noire). Trouver la loi de probabilité de X.

Exemple 3: Essayons maintenant de changer les règles du jeu décrit dans l'exemple 1: on lance la pièce de monnaie non truquée 2 fois successivement. Soit X, le nombre de face obtenu (le nombre de fois où l'étudiant A gagne). Trouver la loi de probabilité de X.

Exemple 4: On jette une pièce de monnaie. On définit la variable aléatoire X par le nombre de jets successifs nécessaire pour obtenir le côté pile pour la première fois. Trouver la loi de probabilité de X.

III. Définitions

Définition 1: Considérons une certaine expérience aléatoire. Soit Ω l'ensemble fondamental de cette expérience. Si à chacun des événements élémentaires w de Ω, on fait correspondre un nombre, on définit une variable aléatoire (VA) notée X. Si à chacune des valeurs possibles de la VA X, on fait associer la probabilité de l'événement correspondant, on obtient une loi de probabilité (ou distribution de probabilité) de la VA X. On note l'ensemble des valeurs de X, Val(X).

Définition 2: On dit qu'une variable aléatoire X est discrète si ses différentes valeurs possibles sont en nombre fini ou infini dénombrable.

► Par exemple, la variable aléatoire étudiée dans l'exemple 1 est défini comme suit:

$$X = \begin{cases} 1, & \text{si l'on obtient pile (succès de l'étudiant);} \\ 0, & \text{si l'on obtient face (échec de l'étudiant).} \end{cases}$$

L'ensemble des valeurs de la VA X est Val$(X) = \{0, 1\}$. Par conséquent, X prend un nombre de valeurs fini et dénombrable. De ce fait, elle est dite discrète.

► Dans l'exemple 4, la variable étudiée X représente "le nombre de jets pour obtenir pile pour la première fois". Alors Val$(X) = \{1, 2, \ldots\}$. Dans ce cas, on ne sait pas quand est ce que on va s'arrêter. Dans cette situation, l'ensemble des valeurs est infini dénombrable. Par conséquent, la variable X est discrète.

IV. Variable aléatoire discrète (VAD)

A. *Loi de probabilité d'une variable aléatoire discrète*

Soit X une variable aléatoire discrète, avec Val$(X) = \{x_1, x_2, \ldots, x_n\}$. Si on associe à chaque valeur x_i de X, une probabilité $p_i = P(x_i) = P(X = x_i)$, alors on dit que l'on a établi (calculé) la loi de probabilité de la VAD X. On lit $P(X = x_i)$ la probabilité que la VAD X prend la valeur x_i.

Remarque 1: Généralement, la loi de probabilité d'une VAD est décrite dans le tableau suivant:

X	x_1	x_2	$\ldots$	x_n
$P(X = x_i)$	p_1	p_2	$\ldots$	p_n

Propriété: Toute loi de probabilité d'une VAD doit vérifier la propriété suivante:

$$\sum_{i=1}^{n} p_i = 1. \tag{1}$$

Essayons de trouver la loi de probabilité pour chaque exemple:
► Pour l'exemple 1, on a Val$(X) = \{0, 1\}$. Ici $p_1 = P(X = 0) = 1/2$ représente la probabilité d'échec ou la probabilité d'avoir face et $p_2 = P(X = 1) = 1/2$ la probabilité de succès ou la probabilité d'avoir pile. La loi de probabilité de X est résumée dans le tableau suivant:

X	0	1
$P(X = x_i)$	1/2	1/2

► Pour l'exemple 2, on a Val$(X) = \{0, 1\}$. On a

X	0	1
$P(X = x_i)$	2/5	3/5

En effet, $P(X = 0)$ représente la probabilité que la boule tirée soit blanche. Comme on a en tout 5 boules, dont 2 sont blanches et 3 sont noires, alors $P(X = 0) = 2/5$ et $P(X = 1) = 3/5$. On peut aussi remarquer que $P(X = 1) = 1 - P(X = 0)$ car il n'existent que deux couleurs différentes.
► Pour l'exemple 3, lorsqu'on jette la pièce de monnaie deux fois, on ne peut avoir que 0, 1 ou 2 piles. Comme X représente le nombre de piles obtenu lors des deux jets, alors Val$(X) = \{0, 1, 2\}$.

Notons que l'espace fondamental de cette expérience aléatoire est $\Omega = \{FF, PF, FP, PP\}$, où P représente pile et F représente face. Trouver $P(X = 0)$ revient à trouver la probabilité de l'événement "obtenir face lors du premier jet et face lors du deuxième jet". Donc $P(X = 0) = 1/4$. De même, on trouve $P(X = 2) = 1/4$. La probabilité $P(X = 1)$ est égale à la probabilité de l'événement composé suivant: "obtenir pile lors du premier jet et face lors du deuxième jet **ou** obtenir face lors du premier jet et pile lors du deuxième jet". Donc $P(X = 1) = 2/4 = 1/2$. Remarquons que

$$
\begin{aligned}
P(X = 0) &= C_2^0(\tfrac{1}{2})^0(\tfrac{1}{2})^2 = \tfrac{1}{4},\\
P(X = 1) &= C_2^1(\tfrac{1}{2})^1(\tfrac{1}{2})^1 = 2 \times \tfrac{1}{4} = \tfrac{1}{2},\\
P(X = 2) &= C_2^2(\tfrac{1}{2})^2(\tfrac{1}{2})^0 = \tfrac{1}{4}.
\end{aligned}
$$

La loi de probabilité de X est alors résumée dans le tableau suivant:

X	0	1	2
$P(X = x_i)$	1/4	1/2	1/4

► Pour l'exemple 4, X représente le nombre de jets pour obtenir pile. Donc l'ensemble fondamental de cette expérience est $\Omega = \{P, FP, FFP, FFFP, \ldots\}$. Par conséquent, Val$(X) = \{1, 2, \ldots\}$. On a $P(X = 1) = \frac{1}{2}$, $P(X = 2) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$, $P(X = 3) = (\frac{1}{2})^2(\frac{1}{2}) = \frac{1}{8}$, etc. D'où

$$P(X = k) = (\frac{1}{2})^{k-1}(\frac{1}{2}) = \frac{1}{2^k}.$$

Remarquons que la propriété (1) est vérifiée pour les exemples 1, 2 et 3. De plus, elle est aussi vérifiée pour l'exemple 4. En effet, on a bien

$$\sum_{k=1}^{\infty} p_i = \frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{2^{k-1}} = \frac{1}{2}\sum_{k=0}^{\infty}\frac{1}{2^k} = \frac{1}{2}\frac{1}{1-\frac{1}{2}} = 1.$$

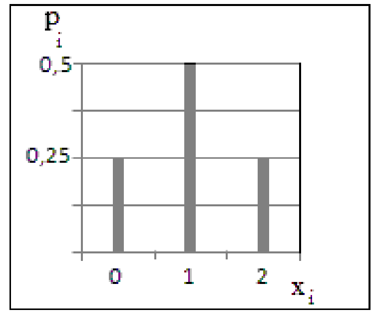


Figure 1. Diagramme en bâtons de de la loi de probabilité de X

*1) **Représentation graphique**:* On représente la loi de probabilité d'une VAD par un diagramme en bâtons: on met dans l'axe horizontal les valeurs x_i de la VA X et dans l'axe vertical, on met les différentes probabilités p_i. Puis pour chaque valeur x_i, on trace une ligne verticale proportionnelle à la probabilité p_i. Le diagramme en bâtons pour l'exemple 3 est représenté dans la figure 1.

*B. **Fonction de répartition d'une VAD***

Définition 3: La fonction de répartition d'une VAD est définie comme suit:
$F : R \to [0, 1]$
$x \mapsto F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$.
Propriétés:
La fonction de répartition d'une VAD possède les propriétés suivantes:

(i) F est croissante;
(ii) F est continue à droite;
(iii) $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$.

Si Val$(X) = \{x_1, x_2, \ldots, x_k\}$, alors les valeurs de la fonction de répartition peuvent être calculées comme suit:

$$F(x) = \begin{cases} 0, & \text{si } x < x_1;\\ F(x_{i+1}), & \text{si } x \in [x_i, x_{i+1}[, i = 1, \ldots, k-1;\\ 1, & \text{si } x \geq x_k, \end{cases}$$

avec

$$F(x_{i+1}) = \sum_{j=1}^{i} p_j.$$

Généralement, ces valeurs sont résumées dans le tableau suivant:

x	$F(x)$
$x < x_1$	0
$x_1 \leq x < x_2$	$F(x_1) = p_1$
$x_2 \leq x < x_3$	$F(x_2) = p_1 + p_2 = F(x_1) + p_2$
$\vdots$	$\vdots$
$x_{k-1} \leq x < x_k$	$F(x_{k-1}) = F(x_{k-2}) + p_{k-1}$
$x \geq x_k$	$F(x_k) = 1$

La fonction de répartition de la VAD X définie dans l'exemple 3 est comme suit:

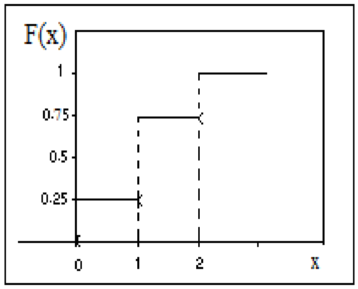


Figure 2. Courbe en escalier de la fonction de répartition de X

x	$F(x)$
$x < 0$	0
$0 \leq x < 1$	$F(0) = 0.25$
$1 \leq x < 2$	$F(1) = 0.25 + 0.5 = 0.75$
$x \geq 2$	$F(2) = 1$

1) Représentation graphique: La fonction de répartition d'une variable aléatoire discrète X est représentée par une courbe en escalier. Dans l'axe horizontal, on met les valeurs x_i de la VAD X et dans l'axe vertical, on met les valeurs $F(x_i)$. Puis pour chaque intervalle, $[x_i, x_{i+1}[$, on trace un segment horizontal, ouvert à droite, au niveau de la valeur $F(x_i)$. La courbe en escalier de la fonction de répartition pour la VAD définie dans l'exemple 3 est représentée dans la figure 2.

C. *Caractéristiques d'une VAD*

Considérons une variable aléatoire discrète X prenant les valeurs $x_1, x_2, \ldots, x_k$ et suivant la loi de probabilité $p_i = P(X = x_i),\ i = 1, 2, \ldots, k$.

*1) **Espérance mathématique**:*

Définition 4: L'espérance mathématique d'une variable aléatoire discrète X, notée $E(X)$, est donnée par la formule suivante:

$$E(X) = \sum_{i=1}^{k} p_i x_i. \tag{2}$$

L'espérance mathématique permet d'estimer la valeur moyenne de la loi de probabilité (le gain moyen espéré, le temps d'attente moyen espéré, etc.). De ce fait, elle est dite un paramètre de position.

*2) **Variance et écart type**:*

Définition 5: La variance et l'écart-type d'une variable aléatoire discrète X, notés respectivement $V(X)$ et σ_X, sont donnés par les formules suivantes:

$$V(X) = \sum_{i=1}^{k} p_i (x_i - E(X))^2 \text{ et } \sigma_X = \sqrt{V(X)}. \tag{3}$$

Ces paramètres permettent de mesurer la dispersion des différentes observations autour de l'espérance mathématique. Plus l'écart-type est faible, plus les différentes valeurs x_i sont regroupées autour de l'espérance mathématique. C'est pourquoi, la variance et l'écart-type sont appelés paramètres de dispersion.

Remarque 2: Pour des considérations pratiques, il est préférable de calculer la variance avec la formule équivalente suivante:

$$V(X) = \sum_{i=1}^{k} p_i x_i^2 - [E(X)]^2. \tag{4}$$

Exemple 5: Calculons l'espérance mathématique, la variance et l'écart-type de la VAD définie dans l'exemple 3:

x_i	0	1	2	$\sum$
p_i	0.25	0.5	0.25	1
$p_i x_i$	0	0.5	0.5	1
$p_i x_i^2$	0	0.5	1	1.5

D'où

$E(X) = \sum_{i=1}^{3} p_i x_i = 1,$
$V(X) = \sum_{i=1}^{3} p_i x_i^2 - [E(X)]^2 = 1.5 - (1)^2 = 0.5,$
$\sigma_X = \sqrt{V(X)} = \sqrt{0.5} = 0.25.$

D. *Lois discrètes usuelles*

*1) **Loi de Bernoulli**:* Considérons une expérience aléatoire ayant deux issues possibles (succès, échec; défectueux, non défectueux, etc). Soient p la probabilité de réalisation de la première issue et $q = 1 - p$ la probabilité de réalisation de la seconde issue. On définie la VAD X comme suit:

$$X = \begin{cases} 1, & \text{si la première issue s'est réalisée;} \\ 0, & \text{si la seconde issue s'est réalisée.} \end{cases}$$

On a $P(X = 1) = p$ et $P(X = 0) = q = 1 - p$. On dit alors que la variable aléatoire X suit la loi de Bernoulli de paramètre p. On note $X \rightsquigarrow B(p)$.

Espérance mathématique et variance:

Les caractéristiques d'une VAD suivant la loi de Bernoulli de paramètre p sont calculées comme suit:

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

Exemple 6: La VAD définie dans l'exemple 1 suit une loi de Bernoulli de paramètre $p = 1/2$. De plus, on a $E(X) = p = 1/2$ et $V(X) = p(1 - p) = 1/4$. La VAD définie dans l'exemple 2 suit aussi une loi de Bernoulli de paramètre $p = 1/6$, avec $E(X) = 1/6$ et $V(X) = 5/36$.

*2) **Loi binômiale**:* Si l'on répète une expérience aléatoire ayant deux issues possibles n fois d'une façon indépendante, et que l'on s'intéresse à la VAD X définie comme étant le nombre de fois où la première issue (succès) s'est réalisée, alors Val$(X) = \{0, 1, \ldots, n\}$ et la loi de probabilité de X aura la forme suivante:

$$P(X = x) = C_n^x p^x (1 - p)^{n-x},\ x = 0, 1, \ldots, n. \tag{5}$$

On dit alors que cette VAD suit une loi binômiale de paramètres n et p. On écrit $X \rightsquigarrow B(n, p)$.

Espérance mathématique et variance:

$$E(X) = np \text{ et } V(X) = np(1 - p). \tag{6}$$

Exercice 1: Démontrer que la loi binômiale est bien une loi de probabilité, puis démontrer les formules (6).

Exemple 7: La VAD définie dans l'exemple 3 suit une loi binômiale de paramètres $n = 2$ et $p = 1/2$: $X \rightsquigarrow B(2, 1/2)$. De plus $E(X) = np = 2/2 = 1$ et $V(X) = np(1 - p) = 1/2$.

*3) **Loi géométrique**:* Si l'on répète une expérience aléatoire ayant deux issues possibles un certain nombre de fois d'une façon indépendante, et que l'on s'intéresse à la VAD X définie comme étant le nombre de répétitions effectuées pour réaliser la première issue (succès), alors Val$(X) = \{1, 2, \dots\}$ et la loi de probabilité de X aura la forme suivante:

$$P(X = x) = (1-p)^{x-1}p, \; x = 0, 1, \dots \tag{7}$$

où p est la probabilité de réalisation de la première issue. On dit alors que cette VAD X suit une loi géométrique (ou loi de Pascal) de paramètre p. On écrit $X \rightsquigarrow G(p)$.

Espérance mathématique et variance:

$$E(X) = \frac{1}{p} \text{ et } V(X) = \frac{1-p}{p^2}. \tag{8}$$

Exercice 2: Démontrer les formules (8).

Exemple 8: La VAD définie dans l'exemple 4 suit une loi géométrique de paramètre $p = 1/2$: $X \rightsquigarrow G(1/2)$.

*4) **Loi de Poisson**:* On dit qu'une VAD X suit une loi de Poisson de paramètre λ si

$$\begin{aligned} &\text{Val}(X) = \{0, 1, 2, \dots\}, \text{ et} \\ &P(X = x) = e^{-\lambda}\tfrac{\lambda^x}{x!}, \; x = 0, 1, \dots . \end{aligned} \tag{9}$$

On écrit alors $X \rightsquigarrow \mathcal{P}(\lambda)$.

Espérance mathématique et variance:

$$E(X) = \lambda \text{ et } V(X) = \lambda. \tag{10}$$

Exercice 3: Démontrer que la loi de poisson est bien une loi de probabilité, puis démontrer les formules (10).

Exemple 9: La VAD définie comme étant le nombre de clients qui arrivent à l'instant t à une fille d'attente devant un guichet d'une banque suit généralement une loi de Poisson de paramètre λ: $X \rightsquigarrow \mathcal{P}(\lambda)$, où le paramètre théorique λ peut être estimé par le nombre moyen de clients qui arrivent à la file à l'instant t.

V. VARIABLE ALÉATOIRE CONTINUE (VAC)

A. *Densité de probabilité d'une VAC*

Définition 6: La densité de probabilité d'une variable aléatoire continue X, notée f, est une fonction continue par morceaux qui vérifie les propriétés suivantes:

(i) $\forall x \in \mathbb{R}, \; f(x) \geq 0$;
(ii) $\int_{-\infty}^{+\infty} f(x)dx = 1$.

Remarque 3: Soit X une VAC. S'il existe une densité de probabilité f, vérifiant:

$$P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \forall x \in \mathbb{R},$$

alors X est dite une variable aléatoire *absolument continue*.

Propriétés: soient $a, b \in \mathbb{R}$.

- $P(a < X \leq b) = \int_a^b f(x)dx$.
- $P(X = a) = 0$;
- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$;
- $P(X > a) = \int_a^{+\infty} f(x)dx = 1 - \int_{-\infty}^{a} f(x)dx$.

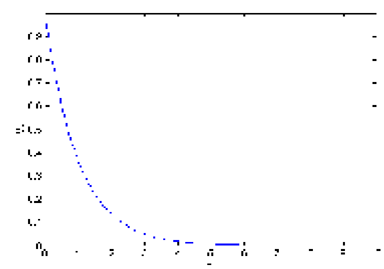


Figure 3. Graphe de la densité de probabilité f

Exemple 10: La fonction suivante:

$$f(x) = \begin{cases} e^{-x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

est une densité de probabilité. En effet, $\forall x \in \mathbb{R}$, $f(x) \geq 0$. De plus,

$$\int_{-\infty}^{+\infty} f(x)dx = \int_0^{+\infty} e^{-x}dx = lim_{t\to+\infty} - e^{-x}|_0^t = 1.$$

Le graphe de la densité de probabilité f est montré dans la figure 3.

B. *Fonction de répartition d'une VAC*

Définition 7: La fonction de répartition d'une VAC X est définie comme suit:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt. \tag{11}$$

Propriétés:

(i) F est continue et croissante sur $\mathbb{R}$;
(ii) Si f est continue au point x alors $F'(x) = f(x)$;
(iii) $P(a < X < b) = F(b) - F(a)$.

Exemple 11: Soit X une VAC ayant comme densité de probabilité la fonction f définie dans l'exemple 10. Calculons la fonction de répartition de X. Si $x \geq 0$, alors

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^x e^{-t}dt = -e^{-t}|_0^x = 1 - e^{-x}.$$

Si $x < 0$, alors $F(x) = \int_{-\infty}^{x} f(t)dt = 0$. D'où

$$F(x) = \begin{cases} 1 - e^{-x}, & x \geq 0; \\ 0, & x < 0. \end{cases}$$

Le graphe de la fonction de répartition de la VAC X est présenté dans la figure 4.

C. *Caractéristiques d'une VAC*

*1) **Espérance mathématique**:*

Définition 8: L'espérance mathématique d'une VAC X est définie comme suit:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx. \tag{12}$$

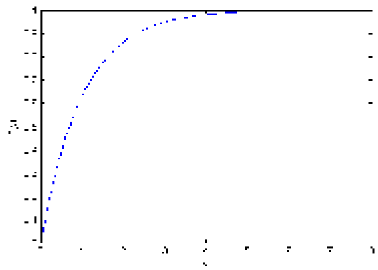


Figure 4. Graphe de la fonction de répartition de la VAC X

*2) **Variance et écart type**:*

Définition 9: La variance et l'écart-type d'une VAC X sont définis comme suit:

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x)dx \text{ et } \sigma_X = \sqrt{V(X)}. \quad (13)$$

Remarque 4: Pour calculer la variance dans la pratique, on utilise la formule suivante:

$$V(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - [E(X)]^2. \quad (14)$$

Exemple 12: Calculons l'espérance mathématique et la variance pour la VAC définie dans les exemples 10 et 11:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{+\infty} xe^{-x}dx,$$

et

$$V(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - [E(X)]^2 = \int_{-\infty}^{+\infty} x^2 e^{-x}dx - [E(X)]^2.$$

Posons

$$I_1 = \int_{0}^{+\infty} xe^{-x}dx \text{ et } I_2 = \int_{0}^{+\infty} x^2 e^{-x}dx.$$

Utilisons l'intégration par partie pour calculer les deux intégrales I_1 et I_2:

$$\begin{cases} g(x) = x \\ h'(x) = e^{-x} \end{cases} \Rightarrow \begin{cases} g'(x) = 1, \\ h(x) = -e^{-x}. \end{cases}$$

On a alors

$$\begin{aligned} \int xe^{-x}dx &= g(x)h(x) - \int g'(x)h(x)dx + c \\ &= -xe^{-x} - e^{-x} + c \\ &= -(x+1)e^{-x} + c, \end{aligned}$$

où $c \in \mathbb{R}$. D'où

$$I_1 = lim_{t \to +\infty} - (x+1)e^{-x}|_0^t = 1.$$

De la même manière, on trouve

$$\int x^2 e^{-x}dx = -(x^2 + 2x + 2)e^{-x} + c',$$

où $c' \in \mathbb{R}$. D'où

$$I_2 = lim_{t \to +\infty} - (x^2 + 2x + 2)e^{-x}|_0^t = 2.$$

Par conséquent,

$$E(X) = I_1 = 1 \text{ et } V(X) = I_2 - I_1^2 = 2 - 1^2 = 1.$$

D. *Variable aléatoire centrée réduite*

Définition 10: Soient X une VAC d'espérance μ et d'écart-type σ et Z une VAC définie comme suit:

$$Z = \frac{X - \mu}{\sigma}. \quad (15)$$

La VAC Z est appelée une VAC centrée réduite.

Propriété: Il est simple de vérifier que l'espérance mathématique et l'écart-type de la VAC Z sont

$$E(Z) = 0 \text{ et } \sigma_Z = 1. \quad (16)$$

E. *Lois continues usuelles*

*1) **Loi uniforme**:*

Définition 11: On dit qu'une VAC X suit une loi uniforme de paramètres a et b si sa densité de probabilité s'écrit sous la forme:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{si } x \in [a,b]; \\ 0, & \text{sinon.} \end{cases} \quad (17)$$

On note $X \rightsquigarrow U(a,b)$.

La fonction de répartition:

$$F(x) = \begin{cases} 0, & \text{si } x < a; \\ 1, & \text{si } x > b; \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b. \end{cases} \quad (18)$$

Espérance mathématique et variance:

$$E(X) = \frac{a+b}{2} \text{ et } V(X) = \frac{(b-a)^2}{12}. \quad (19)$$

Exercice 4: Démontrer les formules (18) et (19). Puis, représenter graphiquement les fonctions de densité et de répartition d'une VAC suivant une loi uniforme.

*2) **Loi exponentielle**:*

Définition 12: On dit qu'une VAC X suit une loi exponentielle de paramètres $\lambda > 0$ si sa densité de probabilité s'écrit sous la forme:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0; \\ 0, & x < 0. \end{cases} \quad (20)$$

On note $X \rightsquigarrow \text{Exp}(\lambda)$.

La fonction de répartition:

$$F(x) = \begin{cases} 0, & \text{si } x < 0; \\ 1 - e^{-\lambda x}, & \text{si si } x \geq 0. \end{cases} \quad (21)$$

Espérance mathématique et variance:

$$E(X) = \frac{1}{\lambda} \text{ et } V(X) = \frac{1}{\lambda^2}. \quad (22)$$

Exercice 5: Démontrer les formules (21) et (22), puis représenter graphiquement les fonctions de densité et de répartition d'une VAC suivant une loi exponentielle de paramètre λ.

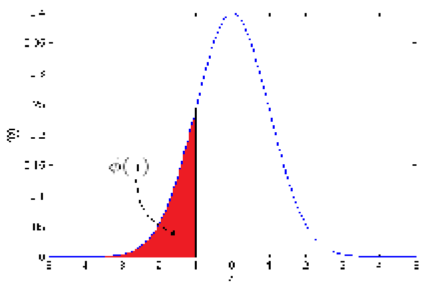


Figure 5. Graphe de la densité de probabilité de Z

*3) **Loi normale**:*

Définition 13: On dit qu'une VAC X suit une loi normale (dite aussi loi de Laplace-Gauss) de paramètres μ et $\sigma > 0$ si sa densité de probabilité s'écrit sous la forme:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ x \in]\infty, +\infty[. \tag{23}$$

On note $X \rightsquigarrow N(\mu, \sigma)$.
L'espérance mathématique et la variance de X sont données par:

$$E(X) = \mu \text{ et } V(X) = \sigma^2. \tag{24}$$

Soit X une VAC qui suit une loi normale de paramètres μ et σ. Dans la pratique, on est souvent amené à calculer les probabilités suivantes:

- $P(X \leq a)$,
- $P(X \geq a)$,
- $P(a \leq X \leq b)$.

Pour calculer $P(X \leq a)$, il faudrait calculer l'intégrale suivante:

$$\int_{\infty}^{a} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dz.$$

Or la primitive de la densité de probabilité d'une loi normale, i.e. sa fonction de répartition, ne peut pas être calculée implicitement. Pour cela, on effectue le changement de variable suivant:

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

Puis, on fait recours à la table de la loi normale centrée réduite pour trouver la valeur suivante:

$$\phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz. \tag{25}$$

On dit alors que la VAC Z suit une loi normale centrée réduite et on écrit: $Z \rightsquigarrow N(0, 1)$. La fonction $\phi(z)$ est alors la fonction de répartition de la VAC Z. La densité de la variable aléatoire Z:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ z \in]\infty, +\infty[$$

est représentée dans la figure 5. L'espérance mathématique et la variance de Z sont alors:

$$E(Z) = 0 \text{ et } V(Z) = 1. \tag{26}$$

La propriété intéressante suivante découle de la symétrie de la loi normale centrée réduite par rapport à l'axe des ordonnées:

$$\phi(-\alpha) = 1 - \phi(\alpha). \tag{27}$$

Ainsi, il est facile de remarquer que:

- $P(Z \leq \alpha) = \phi(\alpha)$,
- $P(Z \geq \alpha) = 1 - \phi(\alpha)$,
- $P(\alpha \leq Z \leq \beta) = \phi(\beta) - \phi(\alpha)$.

Par conséquent,

$$P(X \leq a) = P(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}) = \phi(\frac{a-\mu}{\sigma}),$$

$$P(X \geq a) = 1 - \phi(\frac{a-\mu}{\sigma})$$

et

$$\begin{aligned} P(a \leq X \leq b) &= P(\tfrac{a-\mu}{\sigma} \leq \tfrac{X-\mu}{\sigma} \leq \tfrac{b-\mu}{\sigma}) \\ &= \phi(\tfrac{b-\mu}{\sigma}) - \phi(\tfrac{a-\mu}{\sigma}). \end{aligned}$$

Remarque 5: Géométriquement, $P(a \leq Z \leq b)$ représente l'aire hachurée compris entre la courbe $y = f(z)$ et les deux droites d'équations $y = a$ et $y = b$.

Exemple 13: Soit $Z \rightsquigarrow N(10, 1.2)$. Calculons $P(X \leq 8)$, $P(X \geq 8)$ et $P(-8 \leq X \leq 10)$.
On a

$$\begin{aligned} P(X \leq 8) &= \phi(\tfrac{8-10}{1.2}) = \phi(-1.66) \\ &= 1 - \phi(1.66) \\ &= 1 - 0.9515 = 0.0485. \end{aligned}$$

$$P(X \geq 8) = 1 - \phi(\frac{8-10}{1.2}) = 1 - \phi(-1.66) = 0.9515$$

et

$$\begin{aligned} P(8 \leq X \leq 10) &= \phi(\tfrac{10-10}{1.2}) - \phi(\tfrac{8-10}{1.2}) \\ &= \phi(0) - \phi(-1.66) \\ &= 0.5 - 0.0485 = 0.4515. \end{aligned}$$

Dans certaines situations pratiques, on aura besoin de calculer x tel que:

$$P(X \leq x) = p,$$

où p est une probabilité connue.

Exemple 14: Soit $X \rightsquigarrow N(10, 1.2)$. Calculons x, tel que $P(X \leq x) = 0,8413$. On a

$$P(X \leq x) = P(\frac{X-10}{1.2} \leq \frac{x-10}{1.2}) = 0,8413.$$

En utilisant la table de la loi normale centrée réduite, on trouve

$$\frac{x-10}{1.2} = 1 \Rightarrow x = 11.2.$$

VI. EXERCICES DE TD

EXERCICE 1: On jette un dé une seule fois. On définit la variable aléatoire X par le nombre de fois où l'on a obtenu le chiffre "1".
1) Trouver l'ensemble des valeurs de la VA X.
2) Determiner la loi de probabilité de X et la représenter graphiquement.
3) Determiner la fonction de répartition de X et la représenter graphiquement.
4) Calculer l'espérance mathématique et la variance de X.

Maintenant on lance le dé deux fois successivement. On définit la variable aléatoire X par le nombre de fois où l'on a obtenu le chiffre "1".
Répondre aux mêmes questions 1), 2), 3) et 4).

EXERCICE 2: Le nombre d'ordinateurs vendus chaque jour dans un magasin spécialisé suit une loi de Poisson de paramètre 4. Calculer la probabilité que dans une journée:
1) on ne vende aucun ordinateur.
2) on vende 4 ordinateurs.
3) on vende au moins un ordinateur.
4) le nombre d'ordinateurs vendus est compris entre 2 et 6.

EXERCICE 3: Si X est une variable aléatoire continue de densité de probabilité

$$f(x) = \begin{cases} \frac{1}{a}x^2, & \text{si } 2 \leq x \leq 5; \\ 0, & \text{sinon.} \end{cases}$$

1) Pour quelle valeur de a, f est-elle une densité de probabilité?
2) Calculer l'espérance mathématique et la variance de cette variable aléatoire.

EXERCICE 4: Soit $X \rightsquigarrow U(1,2)$.
1) Déterminer la fonction de répartition F de X et la représenter graphiquement.
2) Calculer $P(X < 3/2)$, $P(1/2 < X \leq 5/4)$, $P(X < 4/3/X < 3/2)$.

EXERCICE 5: Une usine fabrique 9000 unités d'un certain produit en un temps t. Pour cette même période, la demande, en milliers d'unités, concernant ce produit peut être considérée comme une variable aléatoire D suivant une loi exponentielle de paramètre 1/3.
1) Quelle est la probabilité que la demande dépasse la production?
2) Quelle devrait être la production pour que cette demande ne dépasse pas 4%?

EXERCICE 6: Soit une variable aléatoire X qui suit une loi normale de paramètre $\mu = 60$ et $\sigma = 5$.
1) Quelle est la proportion des valeurs dépassant 65.
2) Quelle est la proportion des valeurs comprises entre 56 et 63.
3) Quelle est la valeur de x dépassé par 30% des valeurs.

VII. EXERCICES SUPPLÉMENTAIRES DE TD

EXERCICE 7: On jette une pièce de monnaie une seule fois. On définit la variable aléatoire X par le nombre de faces obtenu.
1) Trouver l'ensemble des valeurs de la VA X.
2) Determiner la loi de probabilité de X et la représenter graphiquement.
3) Déterminer la fonction de répartition de X et la représenter graphiquement.
4) Calculer l'espérance mathématique et la variance de X.
Maintenant on lance la pièce de monnaie trois fois successivement. On définit la variable aléatoire X par le nombre de faces obtenu après les trois jets.
Répondre aux mêmes questions 1), 2), 3) et 4).

EXERCICE 8: Soit f la fonction définie par:

$$f(x) = \begin{cases} ax(1-x), & \text{si } 0 \leq x \leq 1; \\ 0, & \text{sinon.} \end{cases}$$

1) Pour quelle valeur de a, f est-elle une densité de probabilité?
2) Calculer alors $E(X)$ et $V(X)$ pour une variable aléatoire X admettant cette densité.

EXERCICE 9: Une entreprise produit des cables électriques d'une résistance moyenne de 40 Watt et un écart-type de 2 Watt. Si la résistance des cables électriques suit une loi normale, alors quel est le pourcentage des cables qui ont une résistance dépassant 43 Watt.

REFERENCES

[1] S. Adjabi, Statistique Mathématique, *Editions Universitaires Européennes*, USA, 2011.

[2] JM. Bernabotto, Cours de statistiques d'IUT, polycopié de cours, http://eteaching.free.fr/Bts/Poly%20bts/coursstatprobaiut.pdf.

[3] A. Chibat, Cours de statistique, *Collection les mathématiques à l'université*, Université Mentouri de Constantine, 2000.

[4] A. Chibat, Notions sur le calcul des probabilités, *Collection les mathématiques à l'université*, Université Mentouri de Constantine, 2000.

[5] AM. Dussaix et JP. Indjehagopian, Statistiques pour la gestion, Editions Chihab, Alger, 1995.

[6] K. Khaldi, Méthodes statistiques et probabiltés, *Editions CASBAH*, Alger, 2000.

[7] B. Mareschal, Aide à la décision - Statistique, Polycopié de cours, Université Libre de Bruxelles, 2002.

[8] M. Gentes, Cours de Probababilités et Statistiques, Tome 1, Polycopié de cours, IUT d'Orsay, Département Mesures Physiques, 2009-2010.

[9] S. M. Mordjane, Operations Research, *Editions Maison du Livre National*, Benghazi, Lybie, 2002.

[10] D. Mouchiroud, Mathématiques : Outils pour la Biologie, Polycopié de cours, 2002.

[11] A. Moussaoui et B. Youssef, Statistique, Tomme 1 et 2, Maison des sciences pour l'édition et la distribution, Annaba, 2010.

[12] S. Redjel, Lois de probabilité, *Editions de l'université Mentouri de Constantine*, 1999.

[13] AS. Saib, Statistique Descriptive, Université de Tebessa, Polycopié de cours, 2008-2009.

[14] G. Strange, Linear Algebra and its Applications, *Third Edition, Thomson Learning, USA*, 1988.