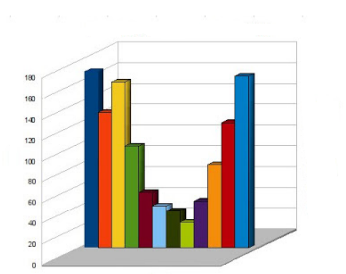
RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abdelhamid Ibn Badis de Mostaganem

Faculté des Sciences Exactes et de l'Informatique

Statistique descriptive et probabilités
Cours et exercices corrigés



Djemaia BENSIKADDOUR

2016/2017

Table des matières

Remerciements

Ce polycopié a été réalisé pendant la période dans laquelle j'ai exercé ma charge pédagogique au sein de la faculté des sciences de la nature et de la vie (FSNV) où j'ai fait connaissance d'un nombre de collègues de différentes spécialités. Il est temps pour les remercier.

Je remercie d'abord Monsieur *M. ZELMAT* ex-chef de département d'agronomie pour sa rigueur, sa simplicité et sa honnêteté qui m'ont beaucoup marqué.

J'exprime ma gratitude à Monsieur *Dj. BENABEDLMOUMENE* chef de département d'agronomie d'avoir été disponible, serviable et compréhensif.

Je remercie tous les collègues que j'ai rencontrés au sein de la faculté des sciences de la nature et de la vie, les responsables, les administratifs et les enseignants.

Notations

- ✓ $\mathbb{N}$: L'ensemble des nombres entiers positifs.
- ✓ $\mathbb{N}^*$: L'ensemble des nombres entiers positifs non nuls
- ✓ $\mathbb{R}$: L'ensemble des nombres réels.
- ✓ $n!$: n factoriel, $n \in \mathbb{N}$.
- ✓ $\sum$: Sigma : somme
- ✓ Π : Pi : produit
- ✓ P : Probabilité
- ✓ Ω : L'univers : l'ensemble de tous les évenements élémentaires d'une expérience aléatoire
- ✓ $\mathcal{F}$: La tribu
- ✓ $\bar{A}$: Le complémentaire d'un évènement A.
- ✓ $\emptyset$: L'ensemble vide

INTRODUCTION

La statistique descriptive est un outil mathématique permettant d'étudier les caractéristiques d'un ensemble d'unités appelé population (objets ou personnes) en fonction de critères de description bien déterminés. Elle sert à classer et à présenter graphiquement et numériquement les données observées de n'importe quel phénomène, afin de les traiter, les modéliser et interpréter les faits aux quels ces données sont relatives.

Le développement de la statistique s'appuie sur la théorie des probabilités, cette dernière est la branche des mathématiques créée depuis le moyen âge et développée par des nombreux mathématiciens notamment en $17^{ème}$ siècle à la suite d'un grand nombre de recherches menées sur les jeux de hasard, parmi les quelles, les travaux de Blaise Pascal et Pierre Fermat.

L'objetif de la probabilité est d'expliquer le hasard pour faire des prévisions, elle nécessite la connaissance de quelques notions d'algèbre et d'analyse mathématique.

Ce travail est un document élémentaire de statistique descriptive et de probabilités, il est destiné aux étudiants de la deuxième année de différentes filières universitaires (sciences exactes et sciences expérimentales). Son objectif est d'introduire les notions de base qui permettent d'organiser, présenter, et correctement analyser les résultats observés d'une épreuve aléatoire ainsi que les différentes lois de probabilités qui servent à faire des prévisions. On espère qu'il aidera les étudiants à mieux maitriser les notions introduites au début des études supérieures en statistique et probabilités.

Ce document est composé de deux parties :

✓ La première est consacrée à présenter des généralités sur les séries statistiques. Elle contient deux chapitres, le premier consiste à organiser les données d'une série statistique simple à une variable quantitative (discrète ou continue) dans des tableaux puis les présenter graphiquement et déterminer les différents paramètres de position, dispersion et de forme afin de les analyser. Dans le deuxième chapitre, on a introduit les séries statistiques à deux variables quantitatives discrètes ou continues et quelques paramètres qui mesurent la liaison entre elles (la covariance et le coefficient de corrélation).

- ✓ La deuxième partie du travail intitulée probabilités, est divisée en quatre chapitres. Dans le premier chapitre réside l'analyse combinatoire. Dans le deuxième, on s'intéresse à rappeler quelques notions de la théorie des ensembles pour introduire les probabilités. Le troisième chapitre est consacré à l'étude de cinq lois discrètes de probabilités, finalement dans le dernier chapitre on a exposé les lois continues de probabilité les plus utilisées dans les différents domaines scientifiques.

Chaque chapitre contient une section d'exercices, elle sert à aider l'étudiant à apprendre un certain nombre d'outils de statistique et à maîtriser le calcul des probabilités.

Première partie

Statistique descriptive

Chapitre 1

Généralités sur les séries statistiques

Le terme statistique désigne à la fois un ensemble de données d'observations, et l'activité qui consiste en leur traitement et leur interprétation.

1.1 Vocabulaire statistique

* **La population Statistique** : est l'ensemble d'individus (personnes ou objets) sur lequel des méthodes et des techniques de présentation et de description statistique sont appliquées.
Par exemple :

1/ L'ensemble des étudiants de la FSNV.

2/ Les arbres fruitiers d'un champ.

3/ La population d'un pays.

* **Les individus statistiques** ou **unités statistiques** : sont les éléments de la population statistique. On associe à ces éléments des propriétés appelées **"caractères"**.
Un **caractère** peut être :

1/ **Qualitatif :** couleur, sexe, efficacité d'un traitement,...

2/ **Quantitatif :** grandeur, âge, taille, poids,...

On distingue deux types de caractère quantitatif :

1/ **discret**,

2/ **continu**.

* Une étude statistique est en fait l'étude de propriétés caractéristiques "caractères" de la population considérée.
* Généralement le nombre d'individus d'une population est très grand, pour cela l'étude statistique se restreint à une partie de cette population supposée assez représentative qu'on appelle **"échantillon"**.
* Si l'échantillon est composé de n individus qui forment la population, alors $x = (x_1, x_2, ..., x_n)$ est la série statistique associée au caractère étudié. On appelle x aussi une variable statistique.
* **L'effectif partiel :** est le nombre d'apparitions d'une valeur du caractère.
* **L'effectif total (la taille de l'échantillon) :** est la somme des effectifs partiels .

Remarque 1.1.1 *Si la série statistique contient k valeurs différentes $x = (x_1, x_2, ..., x_k)$ telles que l'effectif partiel de chaque valeur x_i où $1 \leq i \leq k$ est noté par n_i, alors la taille de l'échantillon $n = \sum_{i=1}^{k} n_i$.*

Exemple 1.1.1 *On considère une population formée de vingt 20 familles numérotée de 1 à 20. Le caractère étudié est le nombre d'enfants de chaque famille.*

$F_1 : 2$	$F_2 : 0$	$F_3 : 2$	$F_4 : 1$
$F_5 : 3$	$F_6 : 2$	$F_7 : 4$	$F_8 : 0$
$F_9 : 5$	$F_{10} : 2$	$F_{11} : 2$	$F_{12} : 3$
$F_{13} : 3$	$F_{14} : 4$	$F_{15} : 5$	$F_{16} : 6$
$F_{17} : 2$	$F_{18} : 1$	$F_{19} : 1$	$F_{20} : 0$

*Les différentes valeurs prises par le caractère discret **"nombre d'enfants"** sont représentées par le tableau du type suivant :*

x_i $(1 \leq i \leq 7)$:	*les valeurs du caractère,*
n_i $(1 \leq i \leq 7)$:	*l'effectif partiel.*

tel que $\sum_{i=1}^{7} n_i = n = 20$: *l'effectif total.*

x_i	0	1	2	3	4	5	6
n_i	3	3	6	3	2	2	1

Remarque 1.1.2 *Le caractère (nombre d'enfants) de l'exemple (1.1.1) est un caractère **discret**.*

*** Représentations graphiques d'une série statistique discrète**

1/ Diagramme en bâtons : Il est construit dans un plan (xoy) muni d'un repère orthogonal, les valeurs de la série statistique sont portées sur l'axe des abscisses (ox), à partir de chaque valeur x_i, on trace le segment de droite verticale et dont la hauteur est proportionnelle à l'effectif correspond.

Exemple 1.1.2

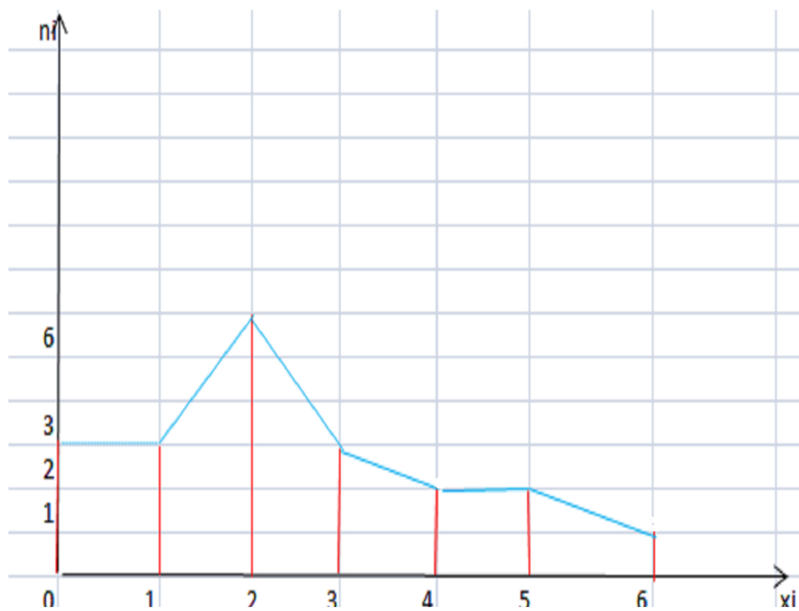


Figure 1

Figure 1 : Le diagramme en bâtons de la série statistique en exemple (1.1.1).

Remarque 1.1.3 *La courbe obtenue en joignant par des segments de droite les extrémités des bâtons est appelée* ***"le polygone des effectifs".***

2/ Diagramme à secteurs : Présenter chaque valeur x_i par un secteur d'angle α_i tel que

$$\alpha_i = \frac{360\ n_i}{n}.$$

Exemple 1.1.3

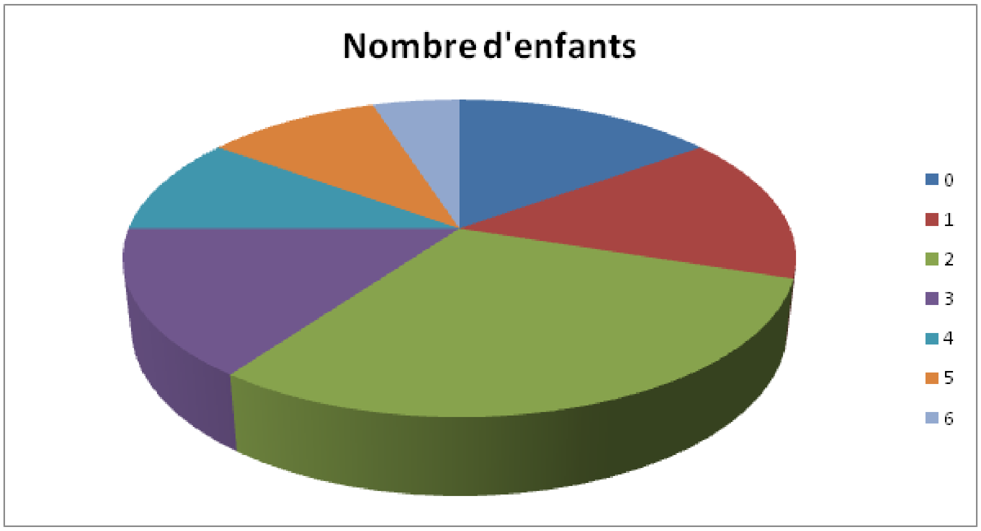


Figure 2

Figure 2 : Représentation graphique par un diagramme à secteurs de la série statistique en exemple (1.1.1).

* **Les fréquences** On appelle "**fréquence**" le rapport

$$f_i = \frac{n_i}{n}$$

tel que

$$\sum_{i=1}^{k} f_i = f_1 + f_2 + ... + f_k = 1.$$

Exemple 1.1.4 *Le tableau ci-dessous représente la distribution des salaires mensuels de* 100 *employés d'une entreprise*

Classe salaire (DA)	*Effectif partiel* (n_i)	*Centre de classe* (c_i)	*Fréquence* (f_i)
$[20000,\ 25000[$	13	22500	$0,13$
$[25000,\ 30000[$	29	27500	$0,29$
$[30000,\ 35000[$	45	32500	$0,45$
$[35000,\ 40000[$	8	37500	$0,08$
$[40000,\ 45000[$	5	42500	$0,05$
	$\sum_{i=1}^{7} n_i = 100$		$\sum_{i=1}^{7} f_i = 1$

*Le centre de la classe $[x_i,\ x_{i+1}[$: $c_i = \frac{x_i + x_{i+1}}{2}$.

*L'amplitude de la classe $[x_i,\ x_{i+1}[$: $a_i = x_{i+1} - x_i$, il n'est pas forcément constant.

Une série statistique classée est naturellement associée à une série à valeurs discrètes qui est la série des centres des différentes classes (voir l'exemple (1.1.4)).

Remarque 1.1.4 *Les séries statistiques continues étudiées sont supposées de même amplitude.*

*** Représentations graphiques d'une série statistique classée (continue)**

i/ Histogramme des effectifs : Est un graphique permettant de représenter les séries statistiques continues. Il est construit dans un plan (xoy) muni d'un repère orthogonal où les valeurs de la série statistique sont placées sur l'axe des abscisses (ox) et les effectifs partiels sont portées sur l'axe des ordonnées (oy). À partir de chaque classe $[x_i,\ x_{i+1}[$ on trace un rectangle, sa largeur est $x_{i+1} - x_i$ et sa hauteur est proportionnelle à l'effectif correspond.

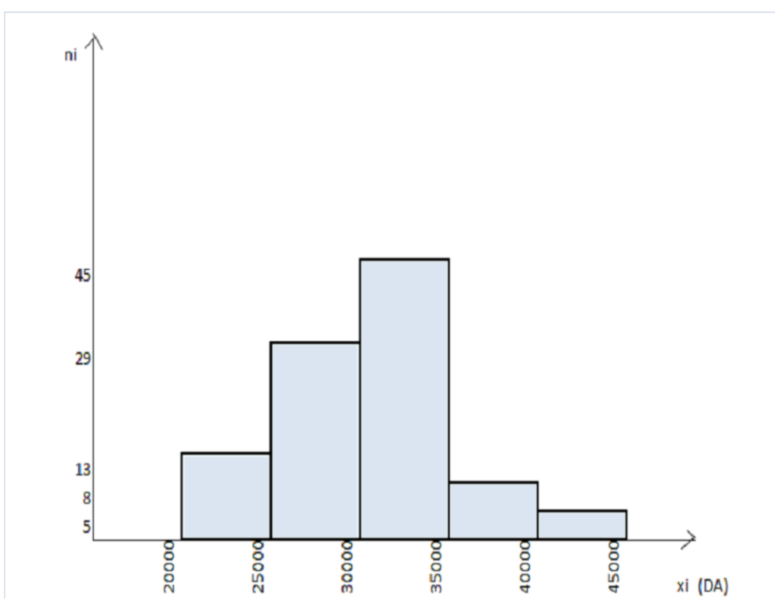


Figure 3

Figure 3 : L'histogramme des effectifs de la série statistique en exemple (1.1.4).

ii/ <u>Diagramme en escalier</u> : Dans un plan (xoy) muni d'un repère orthogonal, on trace à partir de chaque classe $[x_i,\ x_{i+1}[$ placée sur l'axe des abscisses un segment horizontal proportionnel à l'effectif partiel correspondant à cette classe.

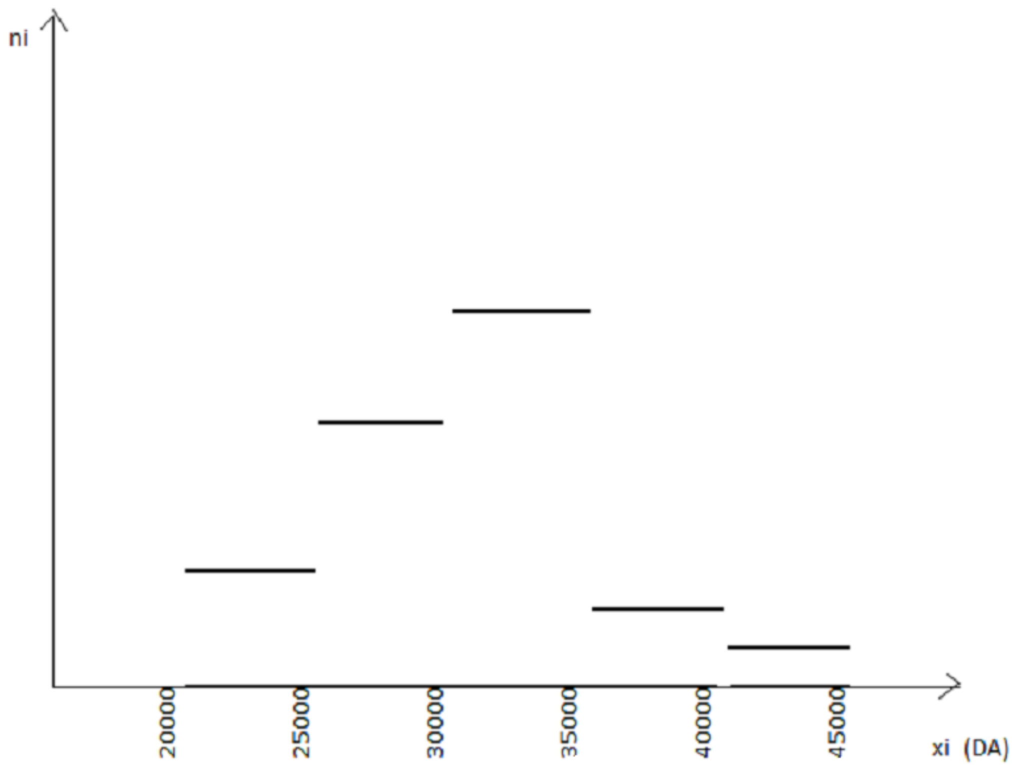


Figure 4

Figure 4 : Diagramme en escalier des effectifs de la série statistique en exemple (1.1.4).

***Effectifs cumulés et fréquences cumulées** Soit $(x_i,\ n_i)_{1\leq i\leq k}$ une série statistique à valeurs discrètes (isolées) telles que $x_1 < x_2 < ... < x_k$.

On appelle **effectif cumulé croissant** de la valeur x_i, $1 \leq i \leq k$, la somme des effectifs partiels de x_i et de toutes les valeurs de la série qui sont inférieures à x_i.

$$N_i \nearrow \ : \text{ effectif cumulé croissant de } x_i$$

$$N_i \nearrow \ = n_1 + n_2 + ... + n_i.$$

On appelle **fréquence cumulée croissante** de la valeur x_i le nombre

$$F_i \nearrow = f_1 + f_2 + ... + f_i.$$

$N_i \searrow$: **effectif cumulé décroissant** de x_i est la somme des effectifs partiels de x_i et de toutes les valeurs de la série qui sont supérieures à x_i

$$N_i \searrow = n_i + n_{i+1} + ... + n_k,$$

de même **la fréquence cumulée décroissante** de la valeur x_i est

$$F_i \searrow = f_i + f_{i+1} + ... + f_k.$$

Exemple 1.1.5 *Calculer $N_i \nearrow$, $f_i \nearrow$, $N_i \searrow$, $f_i \searrow$ de la série statistique de l'exemple (1.1.1) :*

x_i	n_i	$N_i \nearrow$	$f_i \nearrow$	$N_i \searrow$	$f_i \searrow$
0	3	3	$0,15$	20	1
1	3	6	$0,30$	17	$0,85$
2	6	12	$0,60$	14	$0,70$
3	3	15	$0,75$	8	$0,40$
4	2	17	$0,85$	5	$0,25$
5	2	19	$0,95$	3	$0,15$
6	1	20	1	1	$0,05$

Pour une série statistique classée (à caractère continu), on définit de la même manière l'effectif cumulé (la fréquence cumulée) d'une classe, voir le tableau ci-dessous de l'exemple (1.1.4).

$[x_i,\ x_{i+1}[$	n_i	$N_i \nearrow$	$f_i \nearrow$	$N_i \searrow$	$f_i \searrow$
$[20000,\ 25000[$	13	13	$0,13$	100	1
$[25000,\ 30000[$	29	42	$0,42$	87	$0,87$
$[30000,\ 35000[$	45	87	$0,87$	58	$0,58$
$[35000,\ 40000[$	8	95	$0,95$	13	$0,13$
$[40000,\ 45000[$	5	100	1	5	$0,05$

***Polygones cumulatifs** : Dans le même plan (xoy) muni d'un repère orthogonal, on trace la courbe des effectifs cumulés croissants en reliant les points $(x_i,\ N_i \nearrow)$ et la courbe des effectifs cumulés décroissants en reliant les points $(x_i,\ N_i \searrow)$.

Exemple 1.1.6 *On reprend l'exemple (1.1.1)*

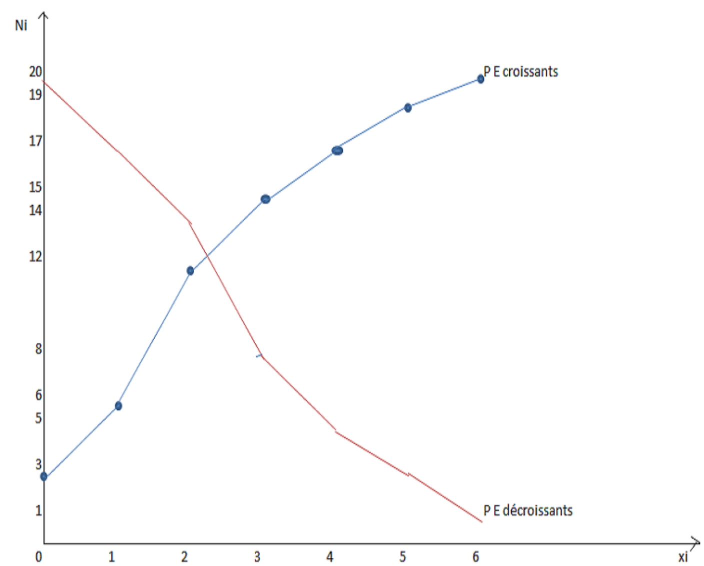


Figure 5

Figure 5 : Polygones des effectifs croissants et décroissants de la série statistique en exemple (1.1.1).

1.2 Paramètres de position

Les paramètres de position servent à caractériser l'ordre de grandeur des valeurs de la série statistique discrète ou continue, on commence par le cas discret.

1.2.1 Cas d'une série statistique discrète

Soit $x : (x_i, n_i)_{1 \leq i \leq k}$, $k \in \mathbb{N}^*$, une série statistique à valeurs discrètes telles que $x_1 < x_2 < ... < x_k$ et $\sum_{i=1}^{k} n_i = n$.

L'étendue "e"

Définition 1.2.1 *L'étendue est la différence des valeurs extrêmes* $e = x_k - x_1$.

Exemple 1.2.1 *Dans l'exemple (1.1.1)* $e = 6 - 0 = 6$.

Le mode "Mo"

Définition 1.2.2 *Le mode est le seul paramètre de position qui s'applique à tous les types de caractères (qualitatif ou quantitatif), il est la valeur* x_i *de plus grand effectif (ou fréquence). On le note par* M_o.

Remarque 1.2.1 *Le mode n'est pas nécessairement unique.*

Remarque 1.2.2 *Dans l'exemple (1.1.1)* $M_o = 2$ *(la valeur du caractère* x_3 *correspondante au plus grand effectif* $n_3 = 6$*).*

La médiane "Med"

Définition 1.2.3 *La médiane* "**Med**" *est la valeur du caractère qui divise les données de la série statistique en deux parties égales, celle qui lui sont plus petites et celles qui lui sont plus grandes.*

i/ *Si* n *est impair* $n = 2r + 1,\ r \in \mathbb{N}^*$ *alors* $Me = x_{r+1}$.

ii/ *Si* n *est pair* $n = 2r,\ r \in \mathbb{N}^*$ *alors* $Med = \frac{x_r + x_{r+1}}{2}$.

Exemple 1.2.2 *Dans l'exemple (1.1.1), on a* $n = 20 = 2 * r$ *alors la médiane est donnée par* $Me = \frac{x_{10} + x_{111}}{2} = 2$.

Remarque 1.2.3 *La médiane peut être déterminée graphiquement. Si l'histogramme des effectifs cumulés est tracé d'une manière précise, la médiane est l'abscisse du point dont l'ordonnée est égale à* $\frac{n}{2}$.

Règle pratique de détermination de la médiane la médiane est la valeur x_i de la série statistique qui correspond à l'effectif cumulé croissant $N_i \nearrow \geq \frac{n}{2}$.

Exemple 1.2.3 *On reprend l'exemple (1.1.1),* $N_3 \nearrow = 12 \geq \frac{n}{2}$ *alors* $Me = x_3 = 2$.

Les quantiles

Définition 1.2.4 *Les quantiles d'une série statistique sont les valeurs* x_K *qui divisent la série statistique en* K *parties égales, c'est-à-dire le quantile d'ordre* K *est par définition la valeur* x_K *qui correspond à l'effectif cumulé croissant supérieur ou égal à* Kn *où* n *est l'effectif total.*

Les quartiles

Définition 1.2.5 *Les quartiles* $Q_1,\ Q_2,\ Q_3$ *sont les valeurs qui correspondent aux effectifs cumulés croissants* $\frac{n}{4},\ \frac{n}{2},\ \frac{3n}{4}$ *respectivement, ils partagent la série statistique en* 4 *parties égales (les* 4 *parties ont toutes le même effectif).*

$$\underbrace{x_1....}_{25\%}Q_1 \quad \underbrace{.....}_{25\%}Q_2 \quad \underbrace{.....}_{25\%}Q_3 \quad \underbrace{.....}_{25\%}x_k$$

Remarque 1.2.4 *Il est clair que* Q_1 *est le premier quartile, le deuxième quartile est la médiane* $Q_2 = Med$ *et* Q_3 *et le troisième quartile.*

Définition 1.2.6 *On appelle interquartile le nombre* $Q_3 - Q_1$.

Les déciles

Définition 1.2.7 *Les déciles* $D_1,\ D_2, ..., D_9$ *sont les valeurs qui correspondent aux effectifs cumulés croissants* $\frac{n}{10},\ \frac{2n}{10}, ...,\ \frac{9n}{10}$ *respectivementsont , ils partagent la série statistique en* 10 *parties qui ont toutes le même effectif.*

Définition 1.2.8 *Le nombre* $D_9 - D_1$ *est l'interdécile.*

Remarque 1.2.5 *On note que* $D_5 = Q_2 = Me$.

Les centiles

Définition 1.2.9 *Les centiles* $C_1,\ C_2, ..., C_{99}$ *sont les valeurs qui correspondent aux effectifs cumulés croissants* $\frac{n}{100},\ \frac{2n}{100}, ...,\ \frac{99n}{100}$ *respectivementsont , ils partagent la série statistique en 100 parties qui ont toutes le même effectif.*

Définition 1.2.10 *Le nombre* $C_{99} - D_1$ *est l'intercentile.*

Remarque 1.2.6 *On note là aussi que* $C_{50} = D_5 = Q_2 = Me$.

Les moyennes

La moyenne arithmétique

Définition 1.2.11 *On appelle moyenne arithmétique de la variable statistique* $(x_i, n_i)_{1\leq i\leq k}$, $k \in \mathbb{N}^*$ *définie précédemment le nombre*

$$\begin{aligned}\bar{x} &= \frac{\displaystyle\sum_{i=1}^{k} n_i x_i}{\displaystyle\sum_{i=1}^{k} n_i} \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i x_i\end{aligned}$$

Proposition 1.2.1 *Soient* $x : (x_i, n_i)_{1\leq i\leq k}$, $y :\ (y_i, n_i)_{1\leq i\leq k}$ $k \in \mathbb{N}^*$ *deux séries statistiques dont les valeurs respectives ont les mêmes effectifs partiels, et* $a,\ b \in \mathbb{R}$. *On a les propriétés suivantes :*

1/ $\overline{x+y} = \bar{x} + \bar{y}$.

2/ $\overline{ax} = a\bar{x}$.

3/ $\overline{x+b} = \bar{x} + b$

Preuve.

1/ On a

$$\begin{aligned}\overline{x+y} &= \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i + y_i\right) \\ &= \frac{1}{n}\left[\sum_{i=1}^{k} n_i x_i + \sum_{i=1}^{k} n_i y_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i x_i + \frac{1}{n}\sum_{i=1}^{k} n_i y_i \\ &= \bar{x} + \bar{y}.\end{aligned}$$

2/ Soit $a \in \mathbb{R}$, alors

$$\begin{aligned}
\overline{ax} &= \frac{1}{n}\sum_{i=1}^{k} n_i\left(ax_i\right) \\
&= a\frac{1}{n}\sum_{i=1}^{k} n_i x_i \\
&= a\bar{x}.
\end{aligned}$$

3/ Soit $b \in \mathbb{R}$, alors

$$\begin{aligned}
\overline{x+b} &= \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i+b\right) \\
&= \frac{1}{n}\sum_{i=1}^{k} n_i x_i + \frac{1}{n}\sum_{i=1}^{k} n_i b \\
&= \overline{x} + b\frac{1}{n}\sum_{i=1}^{k} n_i \\
&= \bar{x} + b.
\end{aligned}$$

□

Remarque 1.2.7 *Les deux propriétés 1/ et 2/ de la proposition (1.2.1) sont équivalentes à*

$$\overline{ax+b} = a\bar{x} + b, \ \forall a, \ b \in \mathbb{R}.$$

La moyenne géométrique

Définition 1.2.12 *La moyenne géométrique de la série statistique* $(x_i, n_i)_{1\leq i\leq k}$, $k \in \mathbb{N}^*$ *est définie par*

$$\begin{aligned}
\bar{x}_G &= \left(\prod_{i=1}^{k} x_i^{n_i}\right)^{\frac{1}{n}} \\
&= \sqrt[n]{x_1^{n_1}...x_k^{n_k}}.
\end{aligned}$$

Remarque 1.2.8 *Par passage aux logarithmes (décimal ou népérien), on obtient deux expressions équivalentes de la moyenne géométrique*

$$\ln \bar{x}_G = \frac{1}{n}\sum_{i=1}^{k} n_i \ln x_i \quad ou \quad \log \bar{x}_G = \frac{1}{n}\sum_{i=1}^{k} n_i \log x_i.$$

Ces expressions montrent qu'il impossible de calculer la moyenne géométrique dés qu'une des valeurs de la variable statistique est nulle ou négative (voir l'exemple (1.1.1)).

La moyenne harmonique

Définition 1.2.13 *La moyenne harmonique de la variable statistique* $(x_i, n_i)_{1\leq i\leq k}$, $k \in \mathbb{N}^*$, *telle que* $x_i \neq 0, \ \forall i = 1, .., k$ *est définie par*

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

Classement des moyennes Les moyennes définies ci-dessus sont classées comme suit

$$\bar{x}_H < \bar{x}_G < \bar{x}.$$

Exemple 1.2.4 *Calculer les moyennes arithmétique, géométrique et harmonique de la variable statistique définie en exemple (1.1.1).*

1.2.2 Cas d'une variable statistique continue (classée)

Soit la série statistique continue $[x_i, \ x_{i+1}[, \ 1 \leq i \leq k-1$, on suppose que toutes les classes ont la même étendue.

Le mode "Mo"

Une variable statistique continue est définie par des intervalles, ce qui signifie qu'il n y a pas une valeur précise du mode, on parle ici d'une classe modale qui correspond au plus grand effectif (ou fréquence).

Exemple 1.2.5 *Dans l'exemple (1.1.4) la classe* $[30000, \ 35000[$ *est correspondante au plus grand effectif.*

On peut déterminer le mode dans le cas continu par deux méthodes différentes :

1/ La valeur du mode est égale au centre de la classe modale

$$\begin{aligned} Mo &= \frac{30000 + \ 35000}{2} \\ &= 32500DA. \end{aligned}$$

2/ Le mode peut être déterminé par la formule suivante :

$$Mo = x_m + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} a$$

où x_m est la borne inférieur de la classe modale, a son étendue et n_m est son effectif partiel, n_{m-1} et n_{m+1} sont les effectifs de classes qui précède et qui suit respectivement la classe modale.

Exemple 1.2.6 *Dans l'exemple (1.1.4), le mode est*

$$\begin{aligned} Mo &= 30000 + \frac{45 - 29}{(45 - 29) + (45 - 8)} 5000 \\ &= 30135,13 DA. \end{aligned}$$

Remarque 1.2.9 *Il n'est pas praticable de dire que le mode est le centre de la classe modale, la deuxième méthode permet de chercher une valeur dans la classe modale qui le présente mieux.*

La médiane "Med"

La détermination graphique de la médiane est très simple. Dans le diagramme des effectifs cumulés (ou fréquences cumulées), la médiane est l'abscisse du point d'intersection du polygone des effectifs (ou de fréquences) et l'horizontale à $\frac{n}{2}$ (ou 50% de la taille de l'échantillon) voir l'exemple ci-dessous :

Exemple 1.2.7 *Dans l'exemple (1.1.4), le salaire médiane correspond à l'effectif cumulé croissant* $\frac{n}{2} = 50$, *alors le* $50^{ème}$ *employé reçoit un salaire compris entre* 30000 *et* 35000 DA.

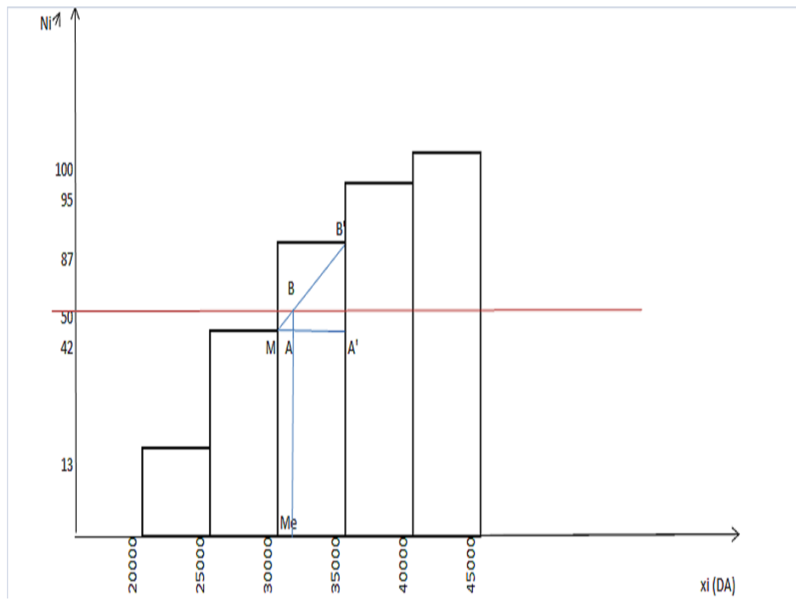


Figure 6

Figure 6 : Détermination graphique de la médiane de la série statistique en exemple (1.1.1).

On peut déterminer la médiane en procèdant par interpolation proportionnelle. Ceci suppose que la distribution des valeurs de la série dans la classe qui contient la médiane est uniforme. Les triangles MAB et $MA'B'$ sont semblables, ils ont donc leurs côtés homologues proportionnels ce qui donne

$$\frac{MA}{MA'} = \frac{MB}{MB'} \Rightarrow MA = MA' \frac{MB}{MB'}$$

cela implique

$$MA = (x_{m+1} - x_m) \frac{\frac{n}{2} - N_m}{N_{m+1} - N_m}$$

d'où la médiane est donnée par

$$Med = x_m + \frac{\frac{n}{2} - N_m}{N_{m+1} - N_m} a$$

telle que $[x_{m+1},\ x_m[$ est la classe qui contient la médiane, $a = (x_{m+1} - x_m)$ son amplitude et $N_m,\ N_{m+1}$ sont les effectifs cumulés correspondant à ces bornes.

Les moyennes

On définit les moyennes d'une variable statistique continue comme étant celles de la série des centres associée $(c_i, n_i)_{1 \leq i \leq k}$ (c'est-à-dire remplcer les classes par leurs centres).

Plus précisement, on a

$$\begin{aligned}
\text{1/ La moyenne arithmétique } \bar{x} &= \tfrac{1}{n}\sum_{i=1}^{k} n_i c_i, \\
\text{2/ La moyenne géométrique } \bar{x}_G &= \sqrt[n]{\prod_{i=1}^{k} c_i^{n_i}}, \\
\text{3/ La moyenne harmonique } \bar{x}_H &= \frac{n}{\sum_{i=1}^{k} \frac{n_i}{c_i}}.
\end{aligned}$$

où c_i : le centre de la classe $[x_i,\ x_{i+1}[$, n : l'effectif total et k : le nombre de classes.

Les quantiles

La formule donnée dans (1.2.2) pour déterminer la médiane d'une variable statistique continue, peut être généralisée au cas des différents quantiles qui sont toujours définis d'une manière analogue à celle du cas discret.

On note par $[x_i, x_{i+1}[$ la classe contenant le $K^{ème}$ quantile x_K où $0 < K < 1$, qui est la classe d'effectif cumulé croissant supérieur ou égal à nK, on a donc

$$\begin{aligned}
x_K &= x_i + \frac{nK - N(x_i)}{N(x_{i+1}) - N(x_i)}(x_{i+1} - x_i) \\
&= x_i + \frac{nK - N(x_i)}{N(x_{i+1}) - N(x_i)}a
\end{aligned}$$

où $N(x_{i+1})$ et $N(x_i)$ sont les effectifs cumulés croissants correspondant aux bornes de la classe $[x_i, x_{i+1}[$.

Exemple 1.2.8 *On a relevé le nombre de pièces déffaillantes constatées durant chaque jour de travail de deux machines A et B produisant des pièces mécaniques. On a alors deux séries statistiques (distributions) :*

a/ *La première* $(x_i, n_i)_{1\leq i\leq 8}$ *: représente les résultats de la machine A.*

b/ *La seconde* $(y_j, n_j)_{1\leq j\leq 6}$ *: représente les résultats de la machine B.*

Machine A

x_i	n_i	$N_i^{\nearrow}$	n_ix_i	$n_ix_i^2$	$n_i\|x_i-\bar{x}\|$
0	13	13	0	0	19,5
1	42	55	42	42	21
2	38	93	76	152	19
3	2	95	6	18	3
4	2	97	8	32	5
5	1	98	5	25	3,5
6	1	99	6	36	4,5
7	1	100	7	49	5,5
Total	100	/	150	354	81

Machine B

y_j	n_j	$N_j^{\nearrow}$	n_jy_j	$n_jy_j^2$	$n_j\|y_j-\bar{y}\|$
0	35	35	0	0	52,5
1	40	75	40	40	20
2	1	76	2	4	0,5
3	1	77	3	9	1,5
4	10	87	40	160	25
5	13	100	65	325	45,5
Total	100	/	150	538	145

On veut comparer les deux distributions à partir des caractéristiques de la tendance centrale. Pour cela on calcule la moyenne arithmétique ainsi que la médiane de chaque série statistique.

i/ *La moyenne arithmétique*

Machine A *Machine B*

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^{8} n_ix_i & \overline{y} &= \frac{1}{n}\sum_{j=1}^{6} n_jy_j \\ &= \frac{1}{100}(150) & &= \frac{1}{100}(150) \\ &= 1,5 & &= 1,5 \end{aligned}$$

ii/ *La médiane*

Machine *A* : *La mediane Me_A est la valeur de la distribution qui correspond à l'effectif cumulé croissant $N_i^{\nearrow} \geq \frac{n}{2} = 50$. Du tableau, on trouve $Me_A = 1$.*

Machine *B* : *La mediane Me_B est la valeur de la distribution qui correspond à l'effectif cumulé croissant $N_i^{\nearrow} \geq \frac{n}{2} = 50$. Du tableau, on trouve $Me_B = 1$.*

Remarque 1.2.10 *On remarque que $\overline{x} = \overline{y}$ et $Me_A = Me_B = 1$, alors on ne peut pas comparer les deux distributions (machine A et B) à partir des caractéristiques de la tendance centrale.*

Il arrive parfois de trouver deux séries statistiques totalement différentes qui possèdent les mêmes caractéristiques de la tendance centrale (paramètres de position), en particulier la moyenne arithmétique et la médiane, pour comparer ses deux séries on aura besoin de déterminer d'autres paramètres qui sont les paramètres de dispersion.

Dans tout ce qui suit on considère la variable statistique $x : (x_i, n_i)_{1\leq i\leq k}$ telle que $\sum_{i=1}^{k} n_i = n$ et les valeurs x_i représentent les centres des classes en cas d'un caractère quantitatif continu.

1.3 Paramètre de dispersion

Les paramètres de dispersion caractérisent l'étalement des observations (les valeurs de la série statistique) autour d'un paramètre de position. Les paramètres de dispersion ne s'appliquent pas aux variables qualitatifs.

1.3.1 Les écarts

L'écart

Définition 1.3.1 *On appelle écart e_i la différence entre une valeur particulière x_i et la moyenne arithmétique $\bar{x}$*

$$e_i = x_i - \bar{x}.$$

L'écart moyen

Puisque les valeurs x_i s'écartent plus ou moins de la moyenne arithmétique, on calcule la moyenne des écarts et on considère cette valeur comme mesure de dispersion

$$\bar{e} = \frac{1}{n}\sum_{i=1}^{k} n_i e_i.$$

Remarque 1.3.1 *La moyenne des écarts est toujours égale à zéro $\sum_{i=1}^{k} n_i e_i = 0$ quel que soit la variable statistique considérée car la somme des écarts négatifs compense exactement la somme des écarts positifs.*

L'écart absolu moyen

Définition 1.3.2 *L'écart absolu moyen est la moyenne des valeurs absolus des écarts*

$$\begin{aligned}\overline{|e|} &= \frac{1}{n}\sum_{i=1}^{k} n_i \left|e_i\right| \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i \left|x_i - \bar{x}\right|.\end{aligned}$$

1.3.2 La variance et l'écart type

La variance

Définition 1.3.3 *On appelle la variance d'une variable statistique l'écart carré moyen.*

$$\begin{aligned} Var(x) &= \frac{1}{n}\sum_{i=1}^{k} n_i e_i^2 \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^2. \end{aligned}$$

Remarque 1.3.2 *la variance mesure la dispersion autour de la moyenne de la variable x. Plus la variance est grande, plus les valeurs de la variable statistique sont dispersées autour de la moyenne.*

Proposition 1.3.1 *La variance d'une la variable statistique discrète ou continue vérifie les propriétés suivantes :*

1/ $Var(x) = \bar{x^2} - \left(\bar{x}\right)^2.$

2/ $Var(ax) = a^2 Var(x), \ \forall a \in \mathbb{R}.$

Preuve.

1/ On a

$$\begin{aligned} Var(x) &= \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i \left(x_i^2 - 2x_i\bar{x} + \left(\bar{x}\right)^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - 2\bar{x}\frac{1}{n}\sum_{i=1}^{k} n_i x_i + \left(\bar{x}\right)^2 \frac{1}{n}\sum_{i=1}^{k} n_i \\ &= \bar{x^2} - 2\left(\bar{x}\right)^2 + \left(\bar{x}\right)^2 \\ &= \bar{x^2} - \left(\bar{x}\right)^2. \end{aligned}$$

2/ Soit $a \in \mathbb{R}$,de 1/ on a

$$Var(ax) = \overline{(ax)^2} - \left(\overline{ax}\right)^2 = \overline{a^2x^2} - \left(\overline{ax}\right)^2$$

en utilisant la propriété 2/ de la moyenne arithmétique (proposition 1.2.1), on obtient

$$\begin{aligned} Var(ax) &= a^2\overline{x^2} - a^2\left(\overline{x}\right)^2 \\ &= a^2 Var(x),\ \forall a \in \mathbb{R}. \end{aligned}$$

□

L'écart type

Définition 1.3.4 *On appelle l'écart type d'une variable statistique la racine carée de la variance*

$$\sigma_x = \sqrt{Var(x)}$$

Remarque 1.3.3 *Il est clair que*

$$\sigma_{ax} = |a|\sqrt{Var(x)},\ \forall a \in \mathbb{R}.$$

Exemple 1.3.1 *On considère les résultats obtenus par* 2 *groupes de* 25 *étudiants à un examen de Statistiques. On note par x les résultats du premier groupe et par y les resultats obtenus par le deuxième groupe.*

Groupe 1

$Note\ sur\ 20\ (x_i)$	n_i	n_ix_i	$n_ix_i^2$
8	5	40	320
9	5	45	405
10	5	50	500
11	5	55	605
12	5	60	720

Groupe 2

$Note\ sur\ 20\ (y_i)$	n_i	n_iy_i	$n_iy_i^2$
0	5	0	0
5	5	25	125
10	5	50	500
15	5	75	1125
20	5	100	2000

Ces deux variables statistiques ont la même moyenne 10/20, il est clair qu'elles sont très différentes. On a $Var\,(x) = 2$ et $Var\,(y) = 50$.

La variance du premier groupe x est assez faible donc les notes obtenues par les étudiants sont très centrées (très proches de la moyenne), par contre la variance du deuxième groupe y est beaucoup plus grande ainsi cela signifie que les notes sont très espacées (étalées).

Coefficient de variation

Définition 1.3.5 *On appelle coefifcient de variation d'une variable statistique x le rapport*

$$C_v = \frac{\sigma_x}{\bar{x}}.$$

Exemple 1.3.2 *Comparer les deux distributions de l'exemple (1.3.1) en utilisant le coefficient de variation.*

$$\begin{array}{lll} \underline{\text{Groupe 1}} & & \\ C_{v_1} & = & \frac{\sigma_x}{\bar{x}} \\ & = & \frac{\sqrt{2}}{10} \\ & = & 0,14. \end{array} \qquad \begin{array}{lll} \underline{\text{Groupe 2}} & & \\ C_{v_2} & = & \frac{\sigma_y}{\bar{y}} \\ & = & \frac{\sqrt{50}}{10} \\ & = & 0,70. \end{array}$$

Comme $C_{v_2} > C_{v_1}$, alors la série statistique y représentant les résultats du groupe 2 est la plus dispersée.

1.3.3 Le moments

Définition 1.3.6 *Soit r un entier positif $(r \in \mathbb{N})$. On appelle moment simple d'ordre r de la la variable statistique x, la quantité*

$$m_r = \frac{1}{n} \sum_{i=1}^{k} n_i x_i^r.$$

1/ *Si $r = 0$, on a $m_0 = 1$: le moment d'ordre 0 est constatnt (égale à 1) pour toute variable statistique x.*

2/ *Si $r = 1$, on a $m_1 = \bar{x}$: le moment d'ordre 1 est la moyenne arithmétique de x.*

3/ *Si $r = 2$, on a $m_2 = \bar{x^2}$: le moment d'ordre 2 est la moyenne arithmétique de x^2.*

Définition 1.3.7 *Soit r un entier positif $(r \in \mathbb{N})$. On appelle moment centré d'ordre r de la variable statistique x, la quantité*

$$M_r = \frac{1}{n} \sum_{i=1}^{k} n_i \left(x_i - \bar{x}\right)^r.$$

1/ *Si $r = 0$, on a $M_0 = 1$: le moment centré d'ordre 0 est constatnt (égale à 1) pour toute variable statistique x.*

2/ *Si $r = 1$, on a $M_1 = 0$: le moment centré d'ordre 1 est constatnt (égale à 0) pour toute variable statistique x.*

3/ *Si* $r = 2$, *on a* $M_2 = Var(x)$: *le moment centré d'ordre* 2 *est la variance de la variable statistique* x.

Corollaire 1.3.1 *Centrer et réduire une variable statistique* x *consiste à la remplacer par la nouvelle variable*

$$X = \frac{x - \bar{x}}{\sigma_x}$$

centrée (a pour moyenne 0*) et réduite* ($\sigma_x = 1$).

1.4 Paramètres de forme

Les paramètres de forme sont définis pour une variable statistique quantitative disccrète ou continue.

1.4.1 Coefficient d'asymétrie

Il existe plusieurs coefficient d'asymétrie, les principaux sont les suivants :

Coefficient d'asymétrie de Pearson

Il fait intervenir le mode quand il existe, il est défini par

$$P = \frac{\bar{x} - Mo}{\sigma_x}$$

Coefficient d'asymétrie de Yule

Il fait intervenir la médiane et les quartiles, il est défini par

$$Y = \frac{Q_1 + Q_3 - 2Me}{2\left(Q_3 - Q_1\right)}$$

Coefficient d'asymétrie de Fisher

Il fait intervenir les moments centrés, il est défini par

$$F = \frac{M_3}{M_2^{3/2}} = \frac{M_3}{\sigma_x^3}$$

Remarque 1.4.1 *Lorsque le coefficient d'asymétrie est positif, la distribution est plus étalée à droite, on dit qu'il y a oblicité à gauche. Dans le cas contraire (le coefficient d'asymétrie est négatif) la distribution est étalée à gauche, on dit qu 'il y a oblicité à droite.*

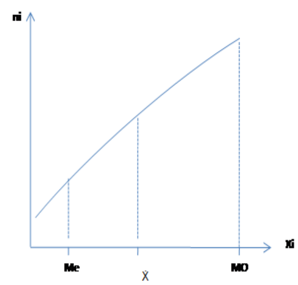


Figure 7 : Oblicit droite

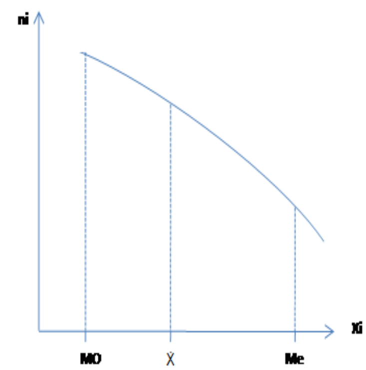


Figure 8 : Oblicité à gauche

Exemple 1.4.1 *Soit la variable statistique*

x_i	-1	0	1	2
n_i	3	2	4	1

Calculer les coefficient de Pearson et de Fisher.

1.4.2 Coefficient d'aplatissement

On a aussi plusieurs coefficients d'aplatissement, le coefficient de Pearson est le plus utilisé, il est donné par

$$\beta_2 = \frac{M_4}{M_2^2}$$

1.5 Ajustement graphique d'une variable statistique

En représentation graphique dans un plan muni d'un repère orthogonal, les valeurs x_i sont portées sur l'axe des abscisses et les effectifs n_i sont portés sur l'axe des ordonnées. La série $(x_i, n_i)_{1 \leq i \leq k}$ forme un nuage de points.

Exemple 1.5.1 *Représenter graphiquement la série statistique* $(x_i, n_i)_{1 \leq i \leq 6}$

x_i	1	2	3	4	5	6
n_i	3	5	4	1	7	11

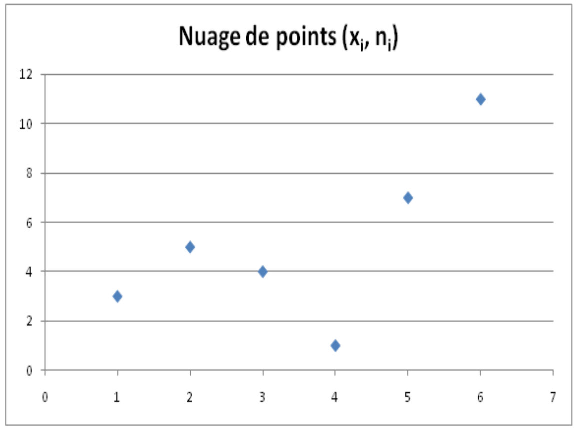


Figure 9

Figure 9 : Nuage de points de la série statistique de l'exemple (1.5.1).

1.5.1 Ajustement linéaire d'une variable statistique

On peut construire une droite d'ajustement qui passe au plus prés des points (x_i, n_i) et déterminer son équation : $y = ax + b$.

Exemple 1.5.2 *Trouver l'équation d'une droite d'ajustement de la variable statistique définie dans l'exemple (1.5.1). On choisit la droite qui passe par les deux points* $A(3, 4)$ *et* $B(6, 11)$. *Son équation est :* $y = \frac{7}{3}x - 3$.

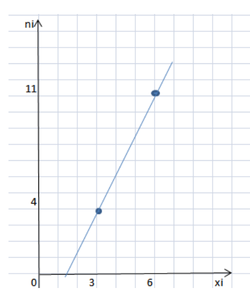


Figure 10

Figure 10 *: Ajustement linéaire de la série statistique de l'exemple (1.5.1).*

1.6 Exercices

Exercice 1 : Classer les caractères suivants dans le tableau ci-dessous

1/ Les différentes marques de voitures en Algérie.

2/ Le moyen d'âge de la population algérienne.

3/ La mortalité des nouveau-nés en Europe.

4/ Les pays de destinations des migrants africains.

5/ Nombre d'étages des gratte-ciels (les plus hautes tours) du monde.

6/ La taille des joueurs de l'équipe nationale de handball.

7/ Les moyens de transport utilisés par les étudiants.

8/ Le salaire mensuel des fonctionnaires d'un établissement.

9/ Nombre de communes par wilaya en Algérie.

10/ L'évolution des températures saisonnières en Algérie.

11/ La production des agrumes à Mostaganem.

12/ Les dépenses fiscales.

Caractère qualitatif	Caractère quantitatif discret	Caractère quantitatif continu

Exercice 2 : Les résultats d'observation de la séquence d'un brin d'ADN sont :

G, G, A, T, A, G, C, T, A, G, G, A, T, G, C, C, T, G, C, T, A, G, T, A, G, A, T, C, G, A, G, C, T, G , C, T, A, C, C, T, C, C, G, A, T, C, G, C, T,C.

A : Adénine, G : Guanine, C : Cytosine, T : Thymine.

1/ Identifier la série statistique présentée ci-dessus (la population, sa taille, le caractère étudié et sa nature).

2/ Donner la distribution des effectifs partiels.

3/ Tracer le diagramme en bâtons de cette série statistique.

4/ Calculer les fréquences.

Exercice 3 : On a relevé le moyen de transport utilisé par les étudiants de la faculté SEI pour se rendre en cours

Transport	Bus	Taxi	A pieds	Total
Effectif	300	165	45	
Fréquence				
Angle				

1/ Quelle est la poulation étudiée ? quel est le caractère étudié ? quelle est sa nature ?

2/ Représenter les données du tableau par un diagramme à secteurs.

Exercice 4 : Soit la série statistique 23, 23, 23, 34, 39, 39, 39, 39, 39, 39, 39, 39, 45, 45, 45, 45, 48, 48, 52, 52, 52, 52, 52, 55, 62, 62, 62, 62, 62, 62.

1/ Donner la distribution des effectifs partiels.

2/ Calculer les fréquences.

3/ Calculer les effectifs cumulés.

4/ Déterminer le mode, la médiane, et les quartiles de cette statistique.

5/ Calculer le $D3$ et $C90$.

Exercice 5 : Le tableau ci-dessous représente la vitesse de 200 véhicules enregistrés par un radar lors d'un contrôle routier.

La vitesse (Km/h)	$[75, 80[$	$[80, 90[$	$[90, 95[$	$[95, 100[$	$[100, 105[$	$[105, 110[$	$[110, 115[$
Nombre de véhicules	10	38	12	96	17	22	5

1/ Identifier la série statistique.

2/ Déterminer l'étendue de cette série statistique et l'amplitude des classes.

3/ Tracer l'histogramme des effectifs.

4/ Calculer les centres de classes puis tracer les courbes des effectifs cumulés.

Exercice 6 : Le taux de glycémie en g/l déterminé chez 32 personnes est représenté dans le tableau suivant :

$0,79$	$0,83$	$0,84$	$0,87$	$0,90$	$0,90$	$0,92$	$0,94$
$0,95$	$0,97$	$0,97$	$0,97$	$1,00$	$1,01$	$1,01$	$1,03$
$1,03$	$1,03$	$1,03$	$1,03$	$1,04$	$1,04$	$1,05$	$1,08$
$1,09$	$1,11$	$1,11$	$1,11$	$1,12$	$1,14$	$1,18$	$1,20$

1/ Quel le type du caractère étudié ?

2/ Mettre cette série statistique sous forme d'une série classée dont toutes les classes ont la même amplitude égale à 0.06.

3/ Présenter graphiquement cette série statistique.

4/ Calculer les fréquences.

5/ Déterminer la classe modale, puis calculer le mode.

6/ Calculer la médiane de cette série statistique.

7/ Calculer les moyennes arithmétique ($\bar{x}$), géométrique ($\bar{x}_G$) et harmonique ($\bar{x}_H$) de cette série statistique.

Exercice 7 : Dans une agence d'Algérie télécom, on a relevé les montants des factures de téléphone délivrées en une journée donnée. Les résultats sont consignés dans le tableau ci-dessous

Mont fct	$[500, 800[$	$[800, 1100[$	$[1100, 1400[$	$[1400, 1700[$	$[1700, 2000[$	$[2000, 2300[$
Nbr fct	3	10	27	12	26	22

Mont fct : montants des factures en DA, Nbr fct : nombre de factures.

1/ Identifier la série statistique x présentée ci-dessus (la population, sa taille, le caractère étudié et sa nature).

2/ Déterminer la classe modale, puis calculer le mode.

3/ Trouver la médiane de cette série statistique.

4/ Calculer le 75 ^{ème} centile C_{75}, la moyenne arithmétique ($\bar{x}$), la variance $Var(x)$, et le coefficient de variation Cv de x.

<u>Exercice 8 :</u> On a effectué un contrôle de la qualité pendant 100 heures de travail sur deux machines produisant des pièces mécaniques.

Certaines pièces présentent des défauts qui les rendent inutilisables. On a relevé le nombre de pièces inutilisables constatées durant chaque heure, les résultats obtenus sont les suivants :

<u>Machine A</u>

Nombre de pièces inutilisables	0	1	2	3	4	5	6	7
Nombre d'heures	13	42	38	2	2	1	1	1

<u>Machine B</u>

Nombre de pièces inutilisables	0	1	2	3	4	5
Nombre d'heures	35	40	1	1	10	13

1/ Comparer les deux distributions (machine A et B) à partir des caractéristiques de la tendance centrale (moyenne arithmétique et médiane).

2/ Comparer les deux distributions (machine A et B) en utilisant les caractéristiques de la dispersion (étendue, écart absolu moyen, écart type).

3/ Quel paramètre semble le plus intéressant à adopter pour comparer ces deux machines ?

<u>Exercice 9 :</u> On a pesé les olives récoltées dans 150 fermes, les résultats sont exprimés en tonnes.

Poids (tonnes) $[x_i, x_{i+1}[$	Effectifs n_i
[5.00, 5.01[	4
[5.01, 5.02[	18
[5.02, 5.03[	25
[5.03, 5.04[	36
[5.04, 5.05[	30
[5.05, 5.06[	22
[5.06, 5.07[	11
[5.07, 5.08[	3
[5.08, 5.09[	1

1/ Identifier la série statistique x présentée ci-dessus (la population, le caractère étudié et sa nature).

2/ On considère une nouvelle variable statistique $y = (x - 5.045)/0.01$, calculer la moyenne arithmétique et la variance de y.

3/ En déduire la moyenne et la variance de x.

4/ Calculer le coefficient de variation de y, en déduire le coefficient de variation de x.

<u>**Exercice 10 :**</u> On a étudié la durée nécessaire pour récolter les premiers fruits des arbres plantés par un fermier. Les résultats sont consignés dans le tableau suivant :

La durée (en mois)	Nombre d'arbres
$[18, \ 22[$	4
$[22, \ 26[$	6
$[26, \ 30[$	5
$[30, \ 34[$	8
$[34, \ 38[$	7

1/ Déterminer la série statistique étudiée (la population, sa taille, le caractère étudié et sa nature).

2/ Calculer la moyenne arithmétique, la variance et le moment centré d'ordre 3 de cette distribution.

3/ Calculer le coefficient de Fisher de cette distribution.

4/ En déduire le type d'oblicité de cette distribution.

1.7 Corrigés

<u>**Exercice 1 :**</u>

Caractère qualitatif	Caractère quantitatif discret	Caractère quantitatif continu
1, 4, 7	3, 5, 9	2, 6, 8, 10, 11, 12

<u>**Exercice 2 :**</u> Les résultats d'observation de la séquence d'un brin d'ADN sont :

G, G, A, T, A, G, C, T, A, G, G, A, T, G, C, C, T, G, C, T, A, G, T, A, G, A, T, C, G, A, G, C, T, G , C, T, A, C, C, T, C, C, G, A, T, C, G, C, T,C.

1/ <u>**Identification de la série statistique**</u>

* La population : les bases azotées de l'ADN.

* Sa taille : 50.

* Le caractère étudié : type de base (A, G, C, T).

* Sa nature : qualitatif.

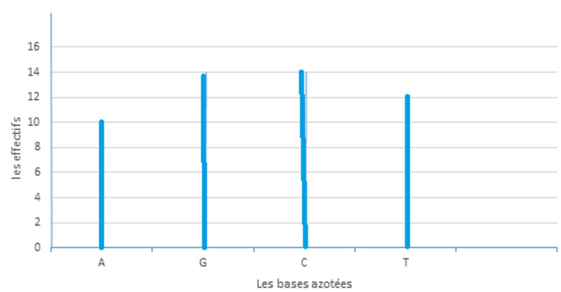
2/ La distributions des effectifs partiels

x_i	n_i	$f_i = \frac{n_i}{n}$
A	10	$0,2$
G	14	$0,28$
C	14	$0,28$
T	12	$0,24$
$Total$	$\sum_{i=1}^{4} n_i = 50$	$\sum_{i=1}^{4} f_i = 1$

3/ Diagramme en bâtons

Sur l'axe (ox), on représente les différentes valeurs du caractère.

Sur l'axe (oy), on représente les effectifs partiels associés.



Exercice 3 :

1/ La population : des étudiants de la faculté SEI, sa taille $n = 510$. Le caractère étudié : le moyen de transport utilisé pour se rendre à la faculté, sa nature qualitatif.

2/ Tableau des fréquences

Transport	Bus	Taxi	A pieds	Total
Effectif	300	165	45	$\sum_{i=1}^{3} n_i = 510$
Fréquence	$0,59$	$0,32$	$0,09$	$\sum_{i=1}^{3} f_i = 1$
Angle	$212,4$	$115,2$	$32,4$	$\sum_{i=1}^{3} \alpha_i = 360^{\circ}$

3/ Diagramme à secteurs

Exercice 4 :

1/ La distribution des effectifs partiels $(x_i)_{1\leq i\leq 8}$: les différentes valeurs du caractère et n_i : nombre d'apparition de chaque valeur (effectif partiel).

x_i	n_i	f_i	$N_i \nearrow$	$N_i \searrow$
23	3	$3/30$	3	30
34	1	$1/30$	4	27
39	8	$8/30$	12	18
45	4	$4/30$	16	18
48	2	$2/30$	18	14
52	5	$5/30$	23	12
55	1	$1/30$	24	7
62	6	$6/30$	30	6
$Total$	$\sum_{i=1}^{8} n_i = 30$	$\sum_{i=1}^{8} f_i = 1$	/	/

2/ Les fréquences $f_i = \frac{n_i}{n}$, (voir le tableau ci-dessus).

3/ Les effectifs cumulés $N_i \nearrow = n_1 + ... + n_i$ et $N_i \searrow = n_i + ... + n_8$, (voir le tableau ci-dessus).

4/ a/ Le mode la valeur de la série du plus grand effectif partiel. Du tableau, on trouve $M_o = 39$.

b/ La mediane la valeur de la série correspondante à $N_i \nearrow \geq \frac{n}{2} = 15$. Du tableau, on trouve $M_e = 45$.

c/ Les quartiles

i/ Q_1 : la valeur de la série correspondante à $N_i \nearrow \geq \frac{n}{4} = 7,5$. Du tableau, on trouve $M_e = 39$.

ii/ $Q_2 = M_e = 45$.

iii/ Q_3 : la valeur de la série correspondante à $N_i \nearrow \geq \frac{3n}{4} = 22,5$. Du tableau, on trouve $Q_3 = 52$.

5/ **Les quantiles**

a/ $3^{ème}$ décile : la valeur de la série correspondante à $N_i \nearrow \geq \frac{3n}{10} = 9$. Du tableau, on trouve $D_3 = 39$.

b/ $90^{ème}$ centile : la valeur de la série correspondante à $N_i \nearrow \geq \frac{90n}{100} = 27$. Du tableau, on trouve $C_{90} = 62$.

Exercice 5 :

1/ **Identification la série statistique** : la population : des véhicules sa taille $n = 200$. Le caractère étudié : la vitesse de chaque véhicule, sa nature quantitatif continu.

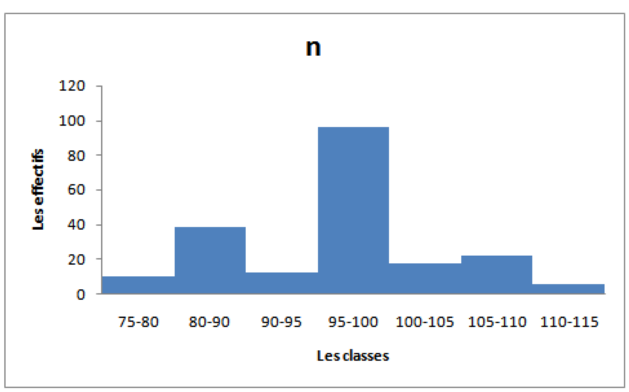
2/ L'étendue de cette série statistique

$$\begin{aligned} e &= x_{\max} - x_{\min} \\ &= 115 - 75 \\ &= 40\ Km/h \end{aligned}$$

et l'amplitude des classes

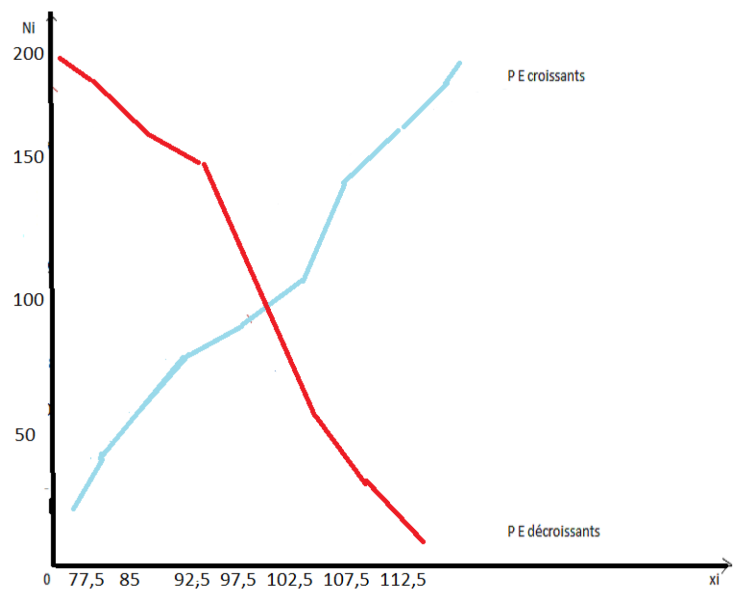
$$a_i = x_{i+1} - x_i$$

3/ L'histogramme des effectifs



4/ Les centres de classes $c_i = \frac{x_{i+1}, + x_i}{2}$

$[x_{i+1},\ x_i[$	n_i	a_i	c_i	$N_i \nearrow$	$N_i \searrow$
$[75,\ 80[$	10	5	$77,5$	10	200
$[80,\ 90[$	38	10	85	48	190
$[90,\ 95[$	12	5	$92,5$	60	152
$[95,\ 100[$	96	5	$97,5$	156	140
$[100,\ 105[$	17	5	$102,5$	173	44
$[105,\ 110[$	22	5	$107,5$	195	27
$[110,\ 115[$	5	5	$112,5$	200	5
$Total$	$\sum_{i=1}^{7} n_i = 200$	/	/	/	/



Courbes des effectifs cumuls croissants et dcroissants

Exercice 6 : L'étude statistique concerne un échantillon de 32 personnes

1/ X : la variable statistique étudié ici est le taux de glycémie en g/l déterminé chez 32 personnes, donc elle représente un caractère quantitatif continu.

2/ Le nombre de classes

$$\begin{aligned} k &= \frac{e}{a} \\ &= \frac{x_{\max} - x_{\min}}{0.06} \\ &= \frac{1,20 - 0,79}{0,06} \\ &= 6,83. \end{aligned}$$

On prend donc $k = 7$. Les classes $[x_i;\ x_{i+1}[$ où $1 \leq i \leq 7$ sont représentées dans le tableau ci-dessous :

$Classes$	n_i	c_i	f_i	$N_i \nearrow$	$n_i c_i$	$n_i \ln c_i$	n_i/c_i
$[0,79;\ 0,85[$	3	0,82	3/32	3	2,46	−0,595	3,66
$[0,85;\ 0,91[$	3	0,88	3/32	6	2,64	−0,383	3,40
$[0,91;\ 0,97[$	3	0,94	3/32	9	2,82	−0,185	3,19
$[0,97;\ 1,03[$	6	1,00	6/32	15	6,00	0	6
$[1,03;\ 1,09[$	9	1,06	9/32	24	9,54	0,524	8,49
$[1,09;\ 1,15[$	6	1,12	6/32	30	6,72	0,679	5,36
$[1,15;\ 1,21[$	2	1,18	2/32	32	2,36	0,331	1,69
$Total$	32	\	1	\	32,54	0,371	31,79

3/ On peut présenter graphiquement cette série statistique par histogramme des effectifs ou un diagramme en excalier.

4/ Les fréquences

$$f_i = \frac{n_i}{n}$$

où n est la taille de l'échantillon. Les résultats sont dans le tableau ci-dessus.

5/ La classe modale $[x_m;\ x_{m+1}[$ est la classe du plus grand effectif partiel. Du tableau, $[1,03;\ 1,09[$ est la classe modale. Le mode M_o est donné par :

$$\begin{aligned} M_o &= x_m + a\frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} \\ &= 1,03 + 0,06\frac{9-6}{(9-6)+(9-6)} \\ &= 1,06\ g/l \end{aligned}$$

6/ Pour calculer la médiane de cette série statistique, il faut d'abord déterminer la classe $[x_m;\ x_{m+1}[$ qui contient la médiane M_e. Cette dernière correspond à $N_i \nearrow \geq \frac{n}{2} = 16$. Du tableau, on trouve $[x_m;\ x_{m+1}[\ = [1,03;\ 1,09[$. La médiane est donc

$$\begin{aligned} M_e &= x_m + a\frac{\frac{n}{2} - N_m}{N_{m+1} - N_m} \\ &= 1,03 + 0,06\frac{16-15}{24-15} \\ &= 1,036\ g/l \end{aligned}$$

7/ Les moyennes de la série statistique

a/ La moyenne arithmétique :

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{7} n_i c_i \\ &= \frac{32,54\ g/l}{32} \\ &= 1,016\ g/l.\end{aligned}$$

b/ La moyenne géométrique :

$$\begin{aligned}\ln \bar{x}_G &= \frac{1}{n}\sum_{i=1}^{7} n_i \ln c_i \\ &= \frac{0,371\ g/l}{32} \\ &= 0,01159.\end{aligned}$$

Alors, $\bar{x}_G = 1,011\ g/l$.

c/ La moyenne harmonique :

$$\begin{aligned}\bar{x}_H &= \frac{n}{\sum_{i=1}^{7} \frac{n_i}{c_i}} \\ &= \frac{32}{31,79}\ g/l \\ &= 1,006\ g/l.\end{aligned}$$

Exercice 7 : Dans une agence d'Algérie télécom, on a relevé les montants des factures de téléphone délivrées en une journée donnée. Les résultats obtenus sont consignés dans le tableau ci-dessous :

$[x_i, x_{i+1}[$	n_i	c_i	$N_i^{\nearrow}$	$n_i c_i$	$n_i c_i^2$
[500, 800[	3	650	3	1950	1267500
[800, 1100[	10	950	13	9500	9025000
[1100, 1400[	27	1250	40	33750	42187500
[1400, 1700[	12	1550	52	18600	28830000
[1700, 2000[	26	1850	78	48100	88985000
[2000, 2300[	22	2150	100	47300	101695000
$Total$	100	\	\	159200	271990000

Ici $(c_i)_{1\leq i\leq 6}$ sont les centres des classes $c_i = \frac{x_i + x_{i+1}}{2}$.

1/ Identifier la série statistique

i/ La population : les factures de téléphone, sa taille $n = \sum_{i=1}^{6} n_i = 100$.

ii/ Le caractère étudié : les montants des factures, sa nature quantitatif continu (classé).

2/ La classe modale $[x_m, x_{m+1}[$ est la classe du plus grand effectif. De la table, il est clair que la classe modale est $[1100,\ 1400[$ et le mode

$$\begin{aligned} Mo &= x_m + \frac{n_m - n_{m-1}}{(n_m - n_{m-1}) + (n_m - n_{m+1})} a \\ &= 1100 + \frac{27 - 10}{(27 - 10) + (27 - 12)} 300 \\ &= 1259,375 DA. \end{aligned}$$

3/ Il faut tout d'abord, trouver la classe qui contient la médiane $[x_m, x_{m+1}[$, elle correspond à l'effectif cumulé croissant $\geq \frac{n}{2} = 50$. De la table, on trouve que la classe qui contient la médiane est $[1400,\ 1700[$, d'où la médiane est

$$\begin{aligned} Med &= x_m + \frac{\frac{n}{2} - N_m}{N_{m+1} - N_m} a \\ &= 1400 + \frac{\frac{100}{2} - 40}{52 - 40} 300 \\ &= 1649,99 \\ &\simeq 1650\ DA. \end{aligned}$$

4/ Calculer C_{75}, $\bar{x}$, $Var(x)$, et Cv de x

i/ <u>Le $75^{ème}$ centile C_{75} :</u> La classe qui contient C_{75} est la classe d'effectif cumulé croissant $\geq \frac{75n}{100} = 75$. De la table, on trouve $[1700,\ 2000[$ et

$$\begin{aligned} C_{75} &= x_i + \frac{n\frac{75}{100} - N(x_i)}{N(x_{i+1}) - N(x_i)} a \\ &= 1700 + \frac{75 - 52}{78 - 52} 300 \\ &= 1965,384 \\ &= 1965,4\ DA. \end{aligned}$$

ii/ La moyenne arithmétique $\bar{x}$

$$\begin{aligned}\bar{x} &= \frac{1}{n}\sum_{i=1}^{6} n_i c_i \\ &= \frac{1}{100}(159200) \\ &= 1592\ DA.\end{aligned}$$

iii/ La variance $Var(x)$:

$$\begin{aligned}Var(x) &= \overline{x^2} - (\bar{x})^2 \\ &= \frac{1}{n}\sum_{i=1}^{6} n_i c_i^2 - (1592)^2 \\ &= \frac{1}{100}(271990000) - 2534464 \\ &= 185436\ \ (DA)^2.\end{aligned}$$

iv/ Coefficient de variation de x

$$\begin{aligned}C_v &= \frac{\sigma_x}{\bar{x}} \\ &= \frac{\sqrt{Var(x)}}{\bar{x}} \\ &= \frac{\sqrt{185436}}{1592} \\ &= 0,27.\end{aligned}$$

Exercice 8 : On a relevé le nombre de pièces inutilisables constatées durant chaque heure de travail de deux machines A et B produisant des pièces mécaniques. On a alors deux séries statistiques (distributions) :

a/ La première $(x_i, n_i)_{1\leq i\leq 8}$: représentes les résultats de la machine A.

b/ La seconde $(y_j, n_j)_{1\leq j\leq 6}$: représente les résultats de la machine B.

Machine A

x_i	n_i	$N_i^{\nearrow}$	$n_i x_i$	$n_i x_i^2$	$n_i \lvert x_i - \bar{x} \rvert$
0	13	13	0	0	19,5
1	42	55	42	42	21
2	38	93	76	152	19
3	2	95	6	18	3
4	2	97	8	32	5
5	1	98	5	25	3,5
6	1	99	6	36	4,5
7	1	100	7	49	5,5
$Total$	100	/	150	354	81

Machine B

y_j	n_j	$N_j^{\nearrow}$	$n_j y_j$	$n_j y_j^2$	$n_j \lvert y_j - \bar{y} \rvert$
0	35	35	0	0	52,5
1	40	75	40	40	20
2	1	76	2	4	0,5
3	1	77	3	9	1,5
4	10	87	40	160	25
5	13	100	65	325	45,5
$Total$	100	/	150	538	145

1/ Comparer les deux distributions à partir des caractéristiques de la tendance centrale

i/ La moyenne arithmétique

Machine A Machine B

$$\begin{aligned} \overline{x} &= \tfrac{1}{n}\sum_{i=1}^{8} n_i x_i \\ &= \tfrac{1}{100}(150) \\ &= 1,5 \end{aligned} \qquad \begin{aligned} \overline{y} &= \tfrac{1}{n}\sum_{j=1}^{6} n_j y_j \\ &= \tfrac{1}{100}(150) \\ &= 1,5 \end{aligned}$$

ii/ La médiane

Machine A : La mediane Me_A est la valeur de la distribution qui correspond à l'effectif cumulé croissant $N_i^{\nearrow} \geq \frac{n}{2} = 50$. Du tableau, on trouve $Me_A = 1$.

Machine B : La mediane Me_B est la valeur de la distribution qui correspond à l'effectif cumulé croissant $N_i^{\nearrow} \geq \frac{n}{2} = 50$. Du tableau, on trouve $Me_B = 1$.

On remarque que $\overline{x} = \overline{y}$ et $Me_A = Me_B = 1$, alors on ne peut pas comparer les deux distributions (machine A et B) à partir des caractéristiques de la tendance centrale.

2/ Comparer les deux distributions à partir des caractéristiques de la dispersion

i/ L'étendue

Machine A Machine B

$$\begin{aligned} e_A &= x_{\max} - x_{\min} \\ &= 7 - 0 \\ &= 7 \end{aligned} \qquad \begin{aligned} e_B &= y_{\max} - y_{\min} \\ &= 5 - 0 \\ &= 5 \end{aligned}$$

ii/ L'écart absolu moyen

Machine A Machine B

$$\begin{aligned} \overline{|e_A|} &= \tfrac{1}{n}\sum_{i=1}^{8} n_i \lvert x_i - \bar{x} \rvert \\ &= \tfrac{1}{100}(81) \\ &= 0,81 \end{aligned} \qquad \begin{aligned} \overline{|e_B|} &= \tfrac{1}{n}\sum_{j=1}^{6} n_j \lvert y_j - \bar{y} \rvert \\ &= \tfrac{1}{100}(145) \\ &= 1,45 \end{aligned}$$

iii/ L'écart type

Machine A

$$\begin{aligned}\sigma_x &= \sqrt{Var(x)} \\ &= \sqrt{\overline{x^2} - (\bar{x})^2} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{8} n_i x_i^2 - 2.25} \\ &= \sqrt{\frac{1}{100}(354) - 2.25} \\ &= 1,135\end{aligned}$$

Machine B

$$\begin{aligned}\sigma_y &= \sqrt{Var(y)} \\ &= \sqrt{\overline{y^2} - (\bar{y})^2} \\ &= \sqrt{\frac{1}{n}\sum_{j=1}^{8} n_j y_j^2 - 2.25} \\ &= \sqrt{\frac{1}{100}(538) - 2.25} \\ &= 1,769\end{aligned}$$

Les écarts (l'écart absolu moyen et l'écart type) de la distribution de la machine B sont supérieurs par rapport à ceux de la distribution de la machine A, alors la distribution de la machine B est plus dispersèe (étalée) par rapport à celle de la machine A.

3/ Les paramètres les plus intéréssants à adopter pour comparer ces deux machines sont les écarts : l'écart absolu moyen et l'écart type.

Remarque 1.7.1 *Les deux derniers exercices (9 et 10) à traiter par les étudiants.*

Chapitre 2

Séries statistiques doubles

Il arrive fréquement d'étudier deux caractères quantitatifs différents x et y d'un échantillon d'une même population pour déterminer s'il existe une relation entre eux (par exemple : la taille et le poids des bébés prématurés) dans le sens que les valeurs de l'un peuvent être obtenues à partir de l'autre à laide d'une correspondance. On dit alors que x et y sont deux variables statistiques dépendantes.
Ce chapitre est consacré à préciser le sens de cette dépendance et à mesurer la liaison eventuelle entre ces deux variables.
En passant aux séries des centres (i.e remplacer chaque classe par son centre), on pourra toujours supposer que les séries considérées sont à valeurs quantitatives discrètes (isolées).

2.1 Traitement des données et représentation graphique

2.1.1 Traitement des données

Pour chaque individu, on relève la valeur de deux caractères x et y, on obtient alors une liste de couples de nombres $(x_i, y_j)_{\substack{1\leq i\leq k\\ 1\leq j\leq l}}$, $k,\ l\ \in \mathbb{N}^*$ que l'on peut présenter sous forme d'un tableau à deux entrées

$x \setminus y$	y_1	y_2	....	y_j	......	y_l	$n_{i.}$	
x_1	n_{11}	n_2	....	n_{1i}	......	n_{1l}	$n_{1.}$	Effectif marginal de x
x_2	n_{21}	n_{22}	....	n_{2j}	....	n_{2l}	$n_{2.}$	
.	.	.	....	......	.....	......	.	
.	.	.	....	.	.	.	.	
.	.	.	....	.	.	.	.	
x_i	n_{i1}	n_{i2}	....	n_{ij}	.	n_{il}	$n_{i.}$	
.	.	.	....	.	.	.	.	
.	.	.	....	.	.....	.	.	
x_k	n_{k1}	n_{k2}	.....	n_{kj}	......	n_{kl}	$n_{k.}$	
$n_{.j}$	$n_{.1}$	$n_{.2}$	....	$n_{.1j}$	....	$n_{.l}$	n	
			Effectif marginal	de y				

* $x_1,\ x_2, \ldots.., x_k$ sont les valeurs du caractère x.

* $y_1,\ y_2, \ldots.., y_l$ sont les valeurs du caractère y.

* n_{ij} est l'effectif du couple (x_i, y_j) pour tout $1 \leq i \leq k$ et $1 \leq j \leq l$.

* $n_{i\ .}$ est l'effectif marginal de x_i, $\ n_{i\ .} = \sum_{j=1}^{l} n_{ij}$.

* $n_{.\ j}$ est l'effectif marginal de y_j, $\ n_{.\ j} = \sum_{i=1}^{k} n_{ij}$.

* n est l'effectif total, $n = \sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij} = \sum_{i=1}^{k} n_{i\ .} = \sum_{j=1}^{l} n_{.\ j}$.

Définition 2.1.1 ***"Fréquences partielles et marginales"*** *Pour tout* $1 \leq i \leq k$ *et* $1 \leq j \leq l$*, on a*

1/ La fréquence partielle du couple (x_i, y_j) est donnée par

$$f_{ij} = \frac{n_{ij}}{n}$$

2/ Les fréquences marginales de la valeur x_i (respectivement y_j) sont

$$f_{i.} = \frac{n_{i.}}{n} \text{ et } f_{.j} = \frac{n_{.j}}{n}$$

vérifiant les relations

$$f_{i.} = \sum_{j=1}^{l} f_{ij} \quad \text{et} \quad f_{.j} = \sum_{i=1}^{k} f_{ij}$$

en plus on a

$$\sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij} = 1$$

2.1.2 Représentation graphique

Soit la série statistique double $(x_i, y_j, n_{ij})_{\substack{1\leq i\leq k\\ 1\leq j\leq l}}$ où $x_1,\ x_2, ..., x_k$ sont les differentes valeurs du premier caractère x et $y_1,\ y_2, ..., y_l$ sont les valeurs du second caractère y.

Définition 2.1.2 ***"Nuage de points"*** *Dans un repère orthogonal du plan, l'ensemble des points M_{ij} de coordonnées $(x_i,\ y_j)$ constitue le nuage de points associé à la série statistique double donnée ci-dessu.*

:/Users/User/AppData/Local/Temp/graphics/$QYEVW515_{19.pdf}$

Définition 2.1.3 ***"Le point moyen"*** *On appelle le point moyen du nuage de points associé à la série statistique double $(x_i, y_j, n_{ij})_{\substack{1\leq i\leq k\\ 1\leq j\leq l}}$ le point G de coordonnées $(\bar{x},\ \bar{y})$ où $\bar{x}$ et $\bar{y}$ sont les moyennes arithmétiques des séries statistiques $(x_i, n_{i.})_{1\leq i\leq k}$ et $(y_j, n_{.j})_{1\leq j\leq l}$ respectivement.*

Exemple 2.1.1 *Le tableau suivant donne la moyenne y des maxima de tension artérielle en fonction de l'âge x d'une population donnée.*

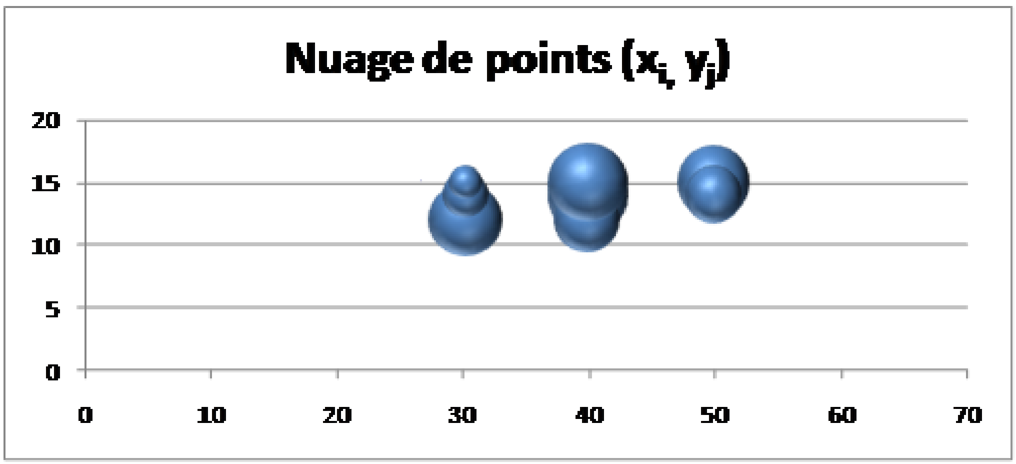
x \ y	12	14	15	ni,
30	5	2	1	
40	4	6	6	
50	0	3	5	
60	0	0	8	
n,j				

$n_{1.}$ est le nombre d'individus ayant 30 ans, alors

$$\begin{aligned} n_{1.} &= n_{11} + n_{12} + n_{13} \\ &= 5 + 2 + 1 \\ &= 8 \end{aligned}$$

de même on a $n_{.1}$ est le nombre d'individus ayant une tension artérielle égal à 12, alors

$$\begin{aligned} n_{.1} &= n_{11} + n_{21} + n_{31} + n_{41} \\ &= 5+4+0+0 \\ &= 9. \end{aligned}$$



On peut en déduire donc les tables des deux séries simples x et y.

La variable x

x_i	$n_{i.}$
30	8
40	16
50	8
60	8
\	$\sum_{i=1}^{4} n_{i.} = 40$

La variable y

y_j	$n_{.j}$
12	9
14	11
15	20
\	$\sum_{j=1}^{3} n_{.j} = 40$

2.2 Covariance d'une série statistique double

Définition 2.2.1 *On appelle covariance d'une série statistique double* $(x_i, y_j, n_{ij})_{\substack{1\leq i\leq k \\ 1\leq j\leq l}}$ *le nombre* $Cov(x,y)$ *donné par :*

$$\begin{aligned} Cov(x,y) &= \overline{xy} - \bar{x}\ \bar{y} \\ &= \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_i y_j\right) - \bar{x}\ \bar{y} \end{aligned}$$

où $\bar{x}$ et $\bar{y}$ sont les moyennes arithmétiques des deux séries simples x et y respectivement.

Remarque 2.2.1 *La covariance mesure les dispersions des deux variables x et y autour de leurs moyennes :*

1/ Si $Cov(x, y)$ est nulle, on dit que les deux variables x et y se produisent indépendamment.

2/ Si $Cov(x, y) > 0$, les deux variables x et y sont liées positivement $\Rightarrow$ l'une augmente, l'autre augmente aussi.

3/ Si $Cov(x, y) < 0$, les deux variables x et y sont liées négativement $\Rightarrow$ l'une augmente, l'autre diminue.

Proposition 2.2.1 ***"Propriétés de la covariance"*** *Soient x et y deux variables statistiques et a, $b, a\prime, b\prime$ des constantes réelles, alors :*

1/ $Cov(x, x) = var(x)$.

2/ $Cov(x, y) = cov(y, x)$.

3/ $Cov(ax + b, a'y + b') = aa'Cov(x, y)$.

Preuve.

1/

$$\begin{aligned} Cov(x, x) &= \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{i=1}^{k} n_{ii}x_i^2\right) - (\bar{x})^2 \\ &= \bar{x^2} - (\bar{x})^2 \\ &= Var(x). \end{aligned}$$

2/

$$\begin{aligned} Cov(x, y) &= \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j\right) - \bar{x}\ \bar{y} \\ &= \frac{1}{n}\left(\sum_{j=1}^{l}\sum_{i=1}^{k} n_{ji}y_jx_i\right) - \ \bar{y}\bar{x} \quad \text{(le produit est commutatif)} \\ &= Cov(y,\ x). \end{aligned}$$

3/ On utilise les propriétés de la moyenne arithmétique, on a

$$\begin{aligned}
Cov(x,y) &= \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(ax_i+b)(a'y_j+b')\right) - \overline{(ax+b)}\ \overline{(a'y+b')} \\
&= \frac{aa'}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j + \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}ba'y_j\right) + \\
&\quad \frac{1}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}ax_ib'\right) + \frac{bb'}{n}\left(\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}\right) - (a\ \bar{x}+b)\,(a'\ \bar{y}+b') \\
&= \frac{aa'}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j + ba'\bar{y} + ab'\ \bar{x} + bb' - aa'\bar{x}\bar{y} - ab'\ \bar{x} - ba'\ \bar{y} - bb' \\
&= aa'\left[\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_iy_j - \bar{x}\bar{y}\right] \\
&= aa'Cov(x,y).
\end{aligned}$$

□

2.3 Ajustement affine d'une série statistique double

On considère la série statistique double $(x_i, y_j, n_{ij})_{\substack{1\le i\le k\\ 1\le j\le l}}$, avec un nuage de points $M_{ij}(x_i, y_j)$ associé.

Si les points de nuage paraissent presque alignés, on peut chercher une relation de la forme $y = ax + b$ où a et b sont deux constantes réelles telles que a est non nulle qui exprime de façon approchée les valeurs de la variable statistique y en fonction des valeurs de la variable x. Alors, graphiquement, il s'agit de déterminer l'équation de la droite qui passe au plus prés de la majorité des points du nuage.

2.3.1 Méthodes d'ajustement affine d'une série statistique double

Il existe de nombreuses méthodes d'obtenir un ajustement affine d'une série statistique double, les plus utilisées sont :

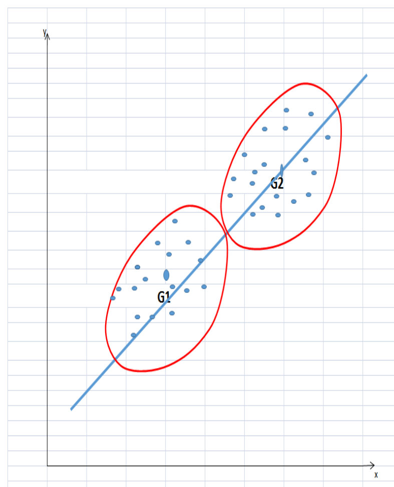
Méthode de Mayer

La détermination d'un ajustement affine d'une série statistique double par la méthode de Mayer se fait en trois étapes :

i/ Découper la série statistique double en deux sous-séries de même effectif, si n l'effectif total de la population est impair, on peut mêttre la valeur surnuméraire du caractère dans n'importe laquelle des deux sous-séries.

ii/ Calculer les coordonnées des deux points moyens G_1 et G_2 associés à ces deux sous-séries.

iii/ Trouver l'équation de la droite qui passe par les deux points G_1 et G_2. Cette droite est dite la droite de Mayer, elle constitue un ajustement affine de la série statistique double.



Droite de mayer

Méthodes de moindres carrés

La méthode de moindres carrés consiste à trouver l'équation : $y = ax + b$ de la droite (D) dite la droite de régression y en x, telle que

$$a = \frac{Cov(x, y)}{Var(x)} \text{ et } b = \bar{y} - a\ \bar{x}.$$

De façon similaire on peut chercher l'équation $y = a' * x + b'$ de la droite (D) dite la droite de régression x en y, où

$$a' = \frac{Cov(x, y)}{Var(y)} \text{ et } b' = \bar{x} - a'\ \bar{y}.$$

Remarque 2.3.1 *Les deux droites de régression données ci-dessous passent par le point moyen* $G\left(\bar{x},\ \bar{y}\right)$.

2.4 Coefficient de corélation linéaire

Définition 2.4.1 *On appelle coefficient de corrélation linéaire du couple* (x, y)*, le nombre réel* $r(x, y)$ *déduit de la covariance et donné par :*

$$r(x,y) \ = \frac{Cov(x,y)}{\sigma_x \sigma_y}.$$

1/ Le coefficient de corrélation ne dépend pas des unités de mesure des variables statistiques étudiées.

2/ La valeur du coéfficient de corrélation est comprise entre -1 et 1 :

$$|r(x,y)| \leq 1.$$

3/ L'absence de corrélation $(r(x, y) = 0)$ implique l'absence de relation linéaire entre les variables statsitiques.

4/ Si $r(x, y) > 0$, alors les variables aléatoires x et y varient dans le même sens, et elles varient dans le sens contraire si $r(x, y) < 0$.

5/ Lorsque la corrélation est forte $r(x, y) \geq \frac{3}{4}$ les deux droites de régression sont très proches et l'ajustement linéaire du nuage de points sera possible, dans le cas contraire le nuage de points ne peut pas être ajusté par une droite, mais il se peut trouver une fonction qui donne la variabilité de x par rapport à y.

2.5 Exercices

Exercice 1 : Soit la série statistique double $(x_i, y_j)_{\substack{1\leq i\leq n \\ 1\leq j\leq n}}$, $n \ \in \mathbb{N}^*$. Montrer les propriétés suivantes :

1/ $Cov\,(x + a, y) = Cov(x, y), \ \forall a \in \mathbb{R}.$

2/ $Var\,(ax + by) = a^2 Var(x) + b^2 Var(y) + 2abCov(x, y).$

Exercice 2 : Soient x et y deux variables statistiques telles que $3y+x = 5, \ \ \bar{x} = 2, \ \ \sigma_x = 0,6.$

1/ Choisir les bonnes réponses.

a/ Les deux variables x et y sont

i/ indépendantes.

ii/ corrélées.

iii/ linéairement dépendantes.

b/ Le coefficient de corrélation entre les deux variables x et y est

i/ négatif.

ii/ nul.

iii/ positif.

c/ La moyenne de la variable y est égale à

i/ 1.

ii/ 2.

iii/ 3.

2/ Calculer $Cov(x, y)$.

3/ En déduire le coefficient de corrélation, sachant que $\overline{y^2} = 3.89$.

Exercice 3 : Dans une entreprise, on appelle y le pourcentage du personnel à temps partiel par rapport au nombre total de salariés en fonction du nombre x d'années écoulées depuis l'an 2000.

x_i	2	4	6	8	10	12	14
y_i	27	26	19	20	8	5	11

1/ Identifier les caractères étudiés.

2/ Représenter par un nuage de points les données de cette étude.

3/ Quelles sont les coordonnées du point moyen G ?

4/ Calculer la covariance de cette série ?

5/ Peut- on affiner ce nuage de points par une droite ? Si cela est possible, trouver l'équation de la droite de régression Y en x.

6/ Quel est le pourcentage du personnel à temps partiel en 2016 ?

Exercice 4 : Une entreprise commerciale consacre une certaine somme à des opérations publicitaires au début de chaque mois.

Les résulats mensuls sont récapitulés dans le tableau ci-dessous :

* les sommes consacrées aux opérations publicitaires,

* les montants des ventes obtenus.

Mois	Dépenses de publicités (en millions de DA) x_i	Ventes (en millions de DA) y_i
Janvier	2.5	38
Février	2.1	42
Mars	2.4	42
Avril	2.4	39
Mai	2.5	40
juin	1.9	45
juillet	2.1	35
Aout	1.8	24
Septembre	3	38
Octobre	3.1	40
Novembre	3.3	44
Décembre	2.5	53

1/ Identifier les caractères étudiés et leurs natures (types) ?

2/ Trouver les coordonnées du point moyen « G» de cette série statistique double.

3/ Les deux variables x et y sont -elles indépendantes ?

4/ Peut-on ajuster linéairement les deux variables x et y ? Justifier.

Exercice 5 : On veut exprimer la taille y en (m) d'un joueur de football en fonction de son poids x en (kg). On a effectué cette étude sur un échantillon composé de 24 joueurs. Les résultats numériques obtenus sont les suivants :

$$\begin{aligned}\sum_{i=1}^{24} x_i &= 1528\ kg,\ \sum_{i=1}^{24} x_i^2 = 97830\ kg^2,\ \sum_{i=1}^{24} y_i = 40,09\ m,\\ \sum_{i=1}^{24} y_i^2 &= 67,0079\ m^2 \text{ et } \sum_{i=1}^{24} x_i y_i = 2555,89\ kg.m.\end{aligned}$$

1/ Calculer les moyennes $\bar{x}$ et $\bar{y}$, les variances $Var(x)$ et $Var(y)$ et la covariance $Cov(x,y)$.

2/ Déterminer la droite des moindres carrés de la taille des joueurs en fonction de leurs poids.

3/ Calculer une mesure de la qualité de l'ajustement des données. Commenter le résultat.

2.6 Corrigés

<u>Exercice 1 :</u> Soit la série statistique double $(x_i, y_j)_{\substack{1\le i\le n\\ 1\le j\le n}}$, $n \in \mathbb{N}^*$. Montrer les propriétés suivantes :

1/ $Cov\,(x+a,y) = Cov(x,y),\ \forall a \in \mathbb{R}$.

Soit $a \in \mathbb{R}$. En utilisant les propriétés de la moyenne arithmétique, on a

$$\begin{aligned} Cov\,(x+a,y) &\overset{déf}{=} \overline{(x+a)\,y} - \overline{(x+a)}\ \bar{y}\\ &\overset{pma}{=} \overline{(xy+ay)} - (\bar{x}+a)\,\bar{y}\\ &\overset{pma}{=} \overline{xy} +\ \overline{ay} - \bar{x}\bar{y} - a\bar{y}\\ &\overset{pma}{=} = \overline{xy} + a\ \overline{y} - \bar{x}\bar{y} - a\bar{y}\\ &= \overline{xy} - \bar{x}\bar{y}\\ &\overset{déf}{=} Cov(x,y).\end{aligned}$$

2/ $Var\,(ax+by) = a^2Var(x) + b^2Var(y) + 2abCov(x,y)$.

Soient $a,\ b \in \mathbb{R}$

$$\begin{aligned} Var\,(ax+by) &\overset{pvar}{=} \overline{(ax+by)^2} - \left(\overline{ax+by}\right)^2\\ &\overset{pma}{=} \overline{a^2x^2 + 2abxy + b^2y^2} - (a\ \bar{x} + b\ \bar{y})^2\\ &\overset{pma}{=} a^2\ \overline{x^2} + 2ab\ \overline{xy} + b^2\overline{y^2} - a^2\,(\bar{x})^2 - b^2\,(\bar{y})^2 - 2ab\bar{x}\bar{y}\\ &= a^2\left(\overline{x^2} - (\bar{x})^2\right) + b^2\left(\overline{y^2} - (\bar{y})^2\right) + 2ab\,(\overline{xy} - \bar{x}\bar{y})\\ &\overset{déf}{=} a^2Var(x) + b^2Var(y) + 2abCov(x,y).\end{aligned}$$

pma : propriété de la moyenne arithmétique. pvar : propriété de la variance.

Remarque 2.6.1 *On peut démontrer les deux propriétés en utilisant les définitions de la moyenne arithmétique et de la variance.*

Exercice 2 : Soient x et y deux variables statistiques telles que $3y+x=5$, $\bar{x}=2$, $\sigma_x=0,6$.

1/ a/ ii/ et iii/ (il y a une relation linéaire entre y et x, l'un dépend de l'autre).

b/ i/ $(y=-\frac{1}{3}x+\frac{5}{3},\ la\ pente\ a=\frac{Cov(x,y)}{Var(x)}=-\frac{1}{3}<0\Rightarrow Cov(x,y)<0\Rightarrow r(x,y)<0)$.

c/ i/ $(\bar{y}=-\frac{1}{3}\bar{x}+\frac{5}{3}=1)$.

2/ $Cov\,(x,y)=-\frac{1}{3}Var(x)=-\frac{1}{3}\ \sigma_x^2=-\frac{1}{3}\,(0,6)^2=-0,12$.

3/ Le coefficient de corrélation

$$r(x,y)=\frac{Cov(x,y)}{\sigma_x\sigma_y}$$

où

$$\begin{aligned}\sigma_y &= \sqrt{Var\,(y)}\\ &= \sqrt{\overline{y^2}-(\bar{y})^2}\\ &= \sqrt{3,89-1}\\ &= 1,7.\end{aligned}$$

Alors

$$\begin{aligned}r(x,y) &= \frac{-0,12}{0,6.1,7}\\ &= -0,1176.\end{aligned}$$

Exercice 3 :

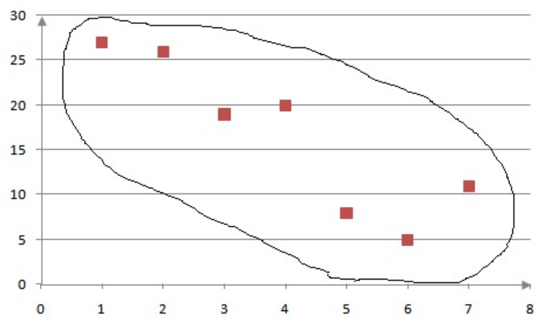
1/ Les caractères étudiés :

x : le nombre d'années écoulées depuis l'an 2000. Sa nature : un caractère quantitatif discrèt.

y : le pourcentage du personnel à temps partiel. Sa nature : un caractère quantitatif continu.

2/ Le nuage de points dans un plan muni d'un repère orthogonal :

L'axe (ox) représente les valeurs du caractère "x" et L'axe (oy) représente les valeurs du caractère "y".



Nuage de points

3/ Le point moyen $G(\overline{\mathbf{x}}, \overline{\mathbf{y}})$:

$$\begin{aligned}\overline{x} &= \frac{1}{n}\sum_{i=1}^{7} n_{i.}x_i,\\ &= \frac{1}{7}(56),\\ &= 8.\end{aligned}$$

Et

$$\begin{aligned}\overline{y} &= \frac{1}{n}\sum_{j=1}^{7} n_{.j}y_j\\ &= \frac{1}{7}(116),\\ &= 16,57.\end{aligned}$$

4/ La covariance Cov(x,y) :

$$\begin{aligned}Cov(x,y) &= \left(\frac{1}{n}\sum_{i=1}^{7}\sum_{j=1}^{7} n_{ij}x_iy_j\right) - \overline{x}\ \overline{y}\\ &= \frac{1}{7}(726) - 132,56,\\ &= -28,84.\end{aligned}$$

5/ Comme $Cov(x,y) \neq 0$, alors, on peut affiner le nuage de points de cette série statistique double par une droite de régression y en x d'équation $y = ax + b$ où $a = \frac{Cov(x,y)}{Var(x)}$.

$$\begin{aligned}
Var(x) &= \overline{x^2} - (\overline{x})^2, \\
&= \left(\frac{1}{n} \sum_{i=1}^{7} n_{i.} (x_i)^2 \right) - 64, \\
&= \frac{1}{7} \left((2)^2 + (4)^2 + (6)^2 + (8)^2 + (10)^2 + (12)^2 + (14)^2 \right) - 64, \\
&= \frac{1}{7} (560) - 64, \\
&= 16.
\end{aligned}$$

Alors,

$$a = -\frac{28,84}{16} = -1,8.$$

et

$$b = \overline{y} - a\overline{x} = 16,57 + 1,8 \times 8 = 30,97.$$

Finalement,

$$\left(D_{y/x}\right) : y = -1,8x + 30,97.$$

6/ Il faut d'abord calculer le coefficient de corrélation linéaire

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y},$$

pour confirmer si l'ajustement linéaire est accepté ou non.

On a

$$\sigma_x = \sqrt{Var(x)} = \sqrt{16} = 4,$$

et

$$\begin{aligned}
Var(y) &= \overline{y^2} - (\overline{y})^2, \\
&= \left(\frac{1}{n} \sum_{j=1}^{7} n_{.j} (y_j)^2 \right) - (16,57)^2, \\
&= \frac{1}{7} (2376) - (16,57)^2, \\
&= 64,86.
\end{aligned}$$

et

$$\sigma_y = \sqrt{Var(y)} = \sqrt{64,86.} = 8,05.$$

Donc,

$$|r_{xy}| = \frac{28,84}{4 \times (8,05)} = 0,895 \simeq 0,9 > 0,75$$

d'où l'ajustement linéaire est accepté. Comme $x_i = 17$ ans en 2016, alors

$$y = -1,8\,(17) + 30,97 = 0,37 = 37\%.$$

Exercice 4 :

1/ Les caractères étudiés sont :

i/ Les dépenses de publicités en millions de DA : caractère X quantitatif continu.

ii/ Les ventes en en millions de DA : caractère Y quantitatif continu.

2/ Les coordonnées du point moyen $G(\bar{X}, \bar{Y})$: Soit $n = \sum_{i=1}^{8} n_{i.} = \sum_{i=1}^{9} n_{.j} = 12$, la taille de l'échantillon.

Caractère X (MDA)

x_i	$n_{i.}$	$n_{i.}x_i$	$n_{i.}x_i^2$
1.8	1	1.8	3.24
1.9	1	1.9	3.61
2.1	2	4.2	8.82
2.4	2	4.8	11.52
2.5	3	7.5	18.75
3	1	3	9
3.1	1	3.1	9.61
3.3	1	3.3	10.89
/	12	29.6	75.44

Caractère Y (MDA)

y_j	$n_{.j}$	$n_{.j}y_j$	$n_{.j}y_j^2$
24	1	24	576
35	1	35	1225
38	2	76	2888
39	1	39	1521
40	2	80	3200
42	2	84	3528
43	1	43	1849
44	1	44	1936
45	1	45	2025
/	12	470	18748

La moyenne arithmétique du caractère X :

$$\begin{aligned} \bar{X} &= \frac{1}{n}\sum_{i=1}^{8} n_{i.}x_i \\ &= \frac{1}{12}29,6\ MDA \\ &= 2,466\ MDA \\ &\simeq 2,47\ MDA. \end{aligned}$$

La moyenne arithmétique du caractère Y :

$$\begin{aligned}\bar{X} &= \frac{1}{n}\sum_{j=1}^{9} n_{.j}y_j\\ &= \frac{1}{12}470 \quad MDA\\ &= 39,166\ MDA\\ &\simeq 39,17 \quad MDA.\end{aligned}$$

Alors, le point moyen de la série statistique doube est $G(2,47MDA,\ 39,17MDA)$.

3/ Deux variables statistiques sont dites linéairement indépendantes si la covariance $Cov(X,Y)$ est nulle.

$x_i \backslash y_j$	24	35	38	39	40	42	43	44	45	$EM\ de\ X$
1.8	1	0	0	0	0	0	0	0	0	1
1.9	0	0	0	0	0	0	0	0	1	1
2.1	0	1	0	0	0	0	1	0	0	2
2.4	0	0	0	1	0	1	0	0	0	2
2.5	0	0	1	0	1	1	0	0	0	3
3	0	0	1	0	0	0	0	0	0	1
3.1	0	0	0	0	1	0	0	0	0	1
3.3	0	0	0	0	0	0	0	1	0	1
$EM\ de\ Y$	1	1	2	1	2	2	1	1	1	12

Calcul de la covariance

$$\begin{aligned}Cov(X,Y) &= \frac{1}{n}\sum_{i=1}^{8}\sum_{j=1}^{9} n_{ij}x_iy_j - \bar{X}\ \bar{Y}\\ &= \left[\frac{1}{12}(1170,1) - 2,47.39,17\right]\ (MDA)^2\\ &= [97.5 - 2,47.39,17]\ (MDA)^2\\ &= 0,7051\ (MDA)^2.\end{aligned}$$

Comme $Cov(X,Y) \neq 0$, alors les deux variables X et Y ne sont pas indépendantes.

4/ La droite de régression Y en X : $Y = aX + b$ où

$$a = \frac{Cov(x,y)}{Var(X)} \; et \; b = \bar{Y} - a \; \bar{X}.$$

On calcule $Var(X)$:

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

et

$$\begin{aligned}
\bar{X}^2 &= \frac{1}{n}\sum_{i=1}^{8} n_{i.}x_i^2 \\
&= \frac{1}{12}75,44 \;\; (MDA)^2 \\
&= 6,286 \;\; (MDA)^2 \\
&\simeq 6,29 \;\; (MDA)^2 .
\end{aligned}$$

Donc

$$\begin{aligned}
Var(X) &= \left[6,29 \;\; - (2,47)^2\right](MDA)^2 \\
&= 0,1891\,(MDA)^2 ,
\end{aligned}$$

d'où

$$\begin{aligned}
a &= \frac{0,7501 \;\; (MDA)^2}{0,1891 \;\; (MDA)^2} \\
&= 3,966 \\
&\simeq 3,97
\end{aligned}$$

et

$$\begin{aligned}
b &= [39,17 - 3,97.2,47]\,(MDA) \\
&= 29,3641 \;\; (MDA) \\
&\simeq 29,36 \;\; (MDA) .
\end{aligned}$$

Finalement, l'équation de la droite de régression Y en X

$$Y = 3,97X + 29,36.$$

5/ <u>Le coefficient de corrélation linéaire</u> :

$$r(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

où

$$\sigma_X = \sqrt{Var(X)} = \sqrt{0,1891 \ \ (MDA)^2} = 0,43\,(MDA)$$

et

$$\begin{aligned}
\sigma_Y &= \sqrt{Var(Y)} \\
&= \sqrt{\frac{1}{n}\sum_{j=1}^{9} n_{.j} y_i^2 - \left(\bar{Y}\ \right)^2} \\
&= \sqrt{\left[\frac{1}{12} 18784 \ \ - (39,17)^2\right] (MDA)^2} \\
&= \sqrt{[1562,33 - 1534,2889]\,(MDA)^2} \\
&= \sqrt{28,0411\,(MDA)^2} \\
&= 5,295 \ \ (MDA) \\
&\simeq 5,3 \ \ (MDA)\,.
\end{aligned}$$

Par conséquent

$$\begin{aligned}
r(X,Y) &= \frac{0,7501\,(MDA)^2}{0,43.5,3 \ \ (MDA)^2} \\
&= 0,329 \\
&\simeq 0.33 \\
&\simeq 33\%.
\end{aligned}$$

6/ $r(X,Y) \simeq 33\% < 75\%$ l'ajustement linéaire du nuage de points n'est pas accepté. Alors on ne peut pas déterminer le montant des ventes si les dépenses de publicités dépassent $3,3$ MDA.

Exercice 5 : On veut exprimer la taille y en (m) d'un joueur de football en fonction de son poids x en (kg). On a effectué cette étude sur un échantillon composé de 24 joueurs, $n = 24$.

$$\begin{aligned}\sum_{i=1}^{24} x_i &= 1528\ kg,\ \sum_{i=1}^{24} x_i^2 = 97830\ kg^2,\ \sum_{i=1}^{24} y_i = 40,09\ m,\\ \sum_{i=1}^{24} y_i^2 &= 67,0079\ m^2 \text{ et } \sum_{i=1}^{24} x_i y_i = 2555,89\ kg.m.\end{aligned}$$

1/ Calculer

i/ Les moyennes $\bar{x}$ et $\bar{y}$

Le poids x	La taille y
$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{24}(1528) = 63,67\ kg$	$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{24}(40,09) = 1,67\ m$

ii/ Les variances $Var(x)$ et $Var(y)$

Le poids x	La taille y
$Var(x) = \overline{x^2} - (\overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (63,67)^2 = \frac{1}{24}(97830) - 4053,8689 = 22,3811\ \ (kg)^2.$	$Var(y) = \overline{y^2} - (\overline{y})^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - (1,67)^2 = \frac{1}{24}(67,0079) - 2,7889 = 0,003\ m^2.$

iii/ La covariance $Cov(x, y)$

$$\begin{aligned}Cov(x,y) &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\ \bar{y}\\ &= \frac{1}{24}(2555,89) - 63,67 * 1,67\\ &= 0,1665\ kg.m\end{aligned}$$

2/ La droite des moindres carrés de la taille des joueurs en fonction de leurs poids

$$y = ax + b$$

où

$$\begin{aligned}a &= \frac{Cov(x,y)}{Var(x)}\\ &= \frac{0,1665}{22,3811}\\ &= 0,0074\ m/kg\end{aligned}$$

et

$$\begin{aligned} b &= \bar{y} - a\ \bar{x} \\ &= 1,67 - 0,0074 * 63,67 \\ &= 1,198842 \\ &\simeq 1,2\ m. \end{aligned}$$

Enfin, la droite des moindres carrés (régression) de y en x est

$$(D) : y = 0,0074x + 1,2.$$

3/ La mesure de la qualité de l'ajustement des données est le coéfficient de corrélation

$$\begin{aligned} r(x,y) &= \frac{Cov(x,y)}{\sigma_x \sigma_y} \\ &= \frac{0,1665}{\sqrt{Var(x)}\sqrt{Var(y)}} \\ &= \frac{0,1665}{\sqrt{Var(x)Var(y)}} \\ &= \frac{0,1665}{\sqrt{22,3811 * 0,003}} \\ &= 0,64255 \\ &\simeq 64,26\ \%. \end{aligned}$$

Comme $r(x,y) < 75\%$, alors l'ajustement des données n'est pas accepté.

Deuxième partie

Probabilités

Chapitre 3

Analyse combinatoire

L'objectif de l'analyse combinatoire et de dénombrer les manières distinctes de grouper les éléments d'un ensemble E de n $(n \in \mathbb{N}^*)$ éléments suivant des lois déterminées.

Notations : Soit E un ensemble de n éléments,

1/ On appelle cardinal de l'ensemble E et on le note par $Card\ E$, le nombre d'éléments de E.

2/ Pour tout $n \in \mathbb{N}^*$, le factoriel de n noté $n!$ est défini par :

$$n! = n \times (n-1) \times (n-2)...3 \times 2 \times 1$$

et on a : $0! = 1! = 1$.

3/ On appelle partition de E l'ensemble de tous les sous-ensembles de E, qu'on note par $\mathcal{P}(E)$.

3.1 Arrangements, permutations et combinaisons

On considère dans tout ce qui suit, E un ensemble de n $(n \in \mathbb{N}^*)$ éléments, et p un entier positif non nul vérifiant $1 \leq p \leq n$.

3.1.1 Arrangements

Définition 3.1.1 *On appelle un arrangement de p éléments de E, toute suite ordonnée de p éléments parmi les n éléments de E. Le nombre des arrangements est noté par : A_n^p.*

Remarque 3.1.1 *Deux arrangements sont dites distinctes, s'ils diffèrent par la nature des éléments* $\Rightarrow$ *ils sont formés des éléments différents qui les composent ou par leur ordre dans la suite.*

Exemple 3.1.1

1/ Le code d'une carte bancaire est composé de 4 chiffres choisis parmi les 10 chiffres de l'ensemble $E = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, alors les différentes manières de composer ce code sont dites les arrangements de $p = 4$ éléments d'un ensemble de $n = 10$ éléments, et on note A_{10}^4.

2/ Choisir une délégation formé de 2 étudiants d'une section de 150 étudiants constitu des arrangements A_{150}^2.

1/ Dans le premier cas de l'exemple (3.1.1), le même élément peut se retrouver plusieurs fois (code 9039 ou 4404...ect) $\Rightarrow$ chaque chiffre du code a 10 possiblités d'apparition $\underset{10}{X}\ \underset{10}{X}\,\underset{10}{X}\,\underset{10}{X}$ $\Rightarrow$ le nombre des arrangements $A_{10}^4 = 10 \times 10 \times 10 \times 10 = 10^4$.

2/ Dans le deuxième cas, les deux étudiants sont forcéments différents, alors le premier est choisi par les 150 étudiants de la section or le deuxième sera choisi parmi les 149 qui restent $\underset{150}{X}\ \underset{149}{X}$ $\Rightarrow$ le nombre des arrangements $A_{150}^2 = 150 \times 149 = 22350$.

On distingue alors, deux types d'arrangements : arrangements avec répétition et arrangements sans répétition.

Arrangement avec répétition

Définition 3.1.2 *Lorsqu'un élément peut être choisi ou observé plusieurs fois, le nombre d'arrangements avec répétition de* p *éléments d'un ensemble de* n *éléments est :*

$$A_n^p = n^p.$$

Exemple 3.1.2 *Voir le premier cas de l'exepmle (3.1.1).*

Arrangement sans répétition

Définition 3.1.3 *Lorsqu'un élément peut être choisi ou observé une seule fois, le nombre d'arrangements sans répétition de* p *éléments d'un ensemble de* n *éléments est :*

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

Exemple 3.1.3 *Voir le deuxième cas de l'exepmle (3.1.1).*

3.1.2 Permutations

Définition 3.1.4 *On appelle une permutation toute suite ordonnée de n éléments de l'ensemble E ou tout arrangement de n éléments de E. Le nombre des permutations est noté par : P_n.*

Exemple 3.1.4

1/ On veut placer 5 étudiants autour d'une table ronde de 5 chaises.

2/ On cherche le nombre de mots possibles (avec ou signification) que l'on peut écrire en permutant les 9 lettres du mot pellicule.

Remarque 3.1.2 *Dans le premier cas de l'exemple (3.1.4) chaque étudiant est placé dans une chaise, donc les étudiants sont forcéments différents* $\underset{5}{X}\ \underset{4}{X}\,\underset{3}{X}\,\underset{2}{X}\,\underset{1}{X} \Rightarrow$ *le nombre de manières de placer les étudiants autour de la table est : $P_5 = 5 \times 4 \times 3 \times 2 \times 1 = 5!$. Par contre dans le deuxième cas la lettre "l" est répétée 3 fois et la lettre "e" est répétée 2 fois, alors on peut écrire des mots contenant la lettre "l" dans trois positions différentes et la lettre"e" dans deux positions différentes $\Rightarrow$ le nombre des mots qu'on peut écrire des 9 lettres du mot "pellicule" doit être raporté au nombre de permutations de 3 lettres identiques "l" et 2 lettres identiques "e". Par conséquent le nombre de mots dans ce cas est :*

$$\begin{aligned} P_9 &= \frac{9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(3 \times 2 \times 1)(2 \times 1)} \\ &= \frac{9!}{3! \times 2!} \end{aligned}$$

Cela conduire à distinguer deux types de permutations :

Permutation sans répétition

Définition 3.1.5 *On appelle une permutation sans répétition toute suite ordonnée de n éléments* ***distincts*** *de l'ensemble E, le nombre de permutations sans répétition de n éléments d'un ensemble de n éléments est :*

$$P_n = n!.$$

Exemple 3.1.5 *Voir le premier cas de l'exepmle (3.1.4).*

Arrangement avec répétition

Définition 3.1.6 *Lorsqu'un élément existe k fois, le nombre de permutations avec répétition de n éléments d'un ensemble de n éléments est :*

$$P_n = \frac{n!}{k!}.$$

En cas où chaque élément x_i de l'ensemble E existe k_i fois tel que $1 \leq i \leq m$ et $\sum_{i=1}^{m} k_i = k_1 + k_2 + ... + k_m = n$,

$$\boxed{E = \left\{ \underbrace{x_1...x_1}_{k_1 \ fois} \underbrace{x_2...x_2}_{k_2 \ fois} .... \underbrace{x_m...x_m}_{k_m \ fois} \right\}},$$

le nombre des permutations est donné par :

$$P_n = \frac{n!}{k_1! \times k_2! \times ... \times k_m!}.$$

Exemple 3.1.6 *Voir le deuxième cas de l'exepmle (3.1.4).*

3.1.3 Combinaisons

Définition 3.1.7 *Une combinaison est un arrangement de p éléments d'un ensemble E de n éléments dans lequel l'ordre ne compte pas. Il ya deux tupes de combinaisons : des combinaisons sans remises et des combinaisons avec remise.*

Combinaison sans remise

Définition 3.1.8 *On appelle combinaison sans remise tout ensemble de p éléments pris sans remise parmi les n éléments de E. Le nombre de combinaison sans remise est :*

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

Exemple 3.1.7 *Former une délégation de 3 employés au sein d'une entreprise de 50 employés $\rightarrow$ nombre de délégations est :*

$$C_{50}^3 = \frac{50!}{3! \times 47!} = \frac{50 \times 49 \times 48 \times 47!}{6 \times 47!} = 19600.$$

Combinaison avec remise

Définition 3.1.9 *On appelle combinaison avec remise tout ensemble de p éléments pris avec remise parmi les n éléments de E. Le nombre de combinaison avec remise est :*

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

Exemple 3.1.8 *Former des mots de 3 lettres (avec ou sans signification) à partir d'un ensemble de 5 lettres $\rightarrow$ nombre de mots est :*

$$C_7^3 = \frac{7!}{3! \times 4!} = \frac{7 \times 6 \times 5 \times 4!}{6 \times 4!} = 35.$$

Proposition 3.1.1 ***"Propriétés de combinaisons"***

1/

$$C_n^0 = C_n^n = 1.$$

2/

$$C_n^1 = C_n^{n-1} = n.$$

3/

$$C_n^2 = C_n^{n-2} = \frac{n(n-1)}{2}.$$

4/

$$C_n^p = C_n^{n-p}.$$

5/

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

Preuve. Pour démontrer les propriétés ci-dessus, on utilise la formule de la cobinaison donnée en définition (3.1.8).

1/

$$C_n^0 = \frac{n!}{0! * n!} = \frac{n!}{n! * 0!} = C_n^n = 1.$$

2/

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

3/

$$C_n^2 = \frac{n!}{2! * (n-2)!} = \frac{n!}{(n-2)! * 2!} = C_n^{n-2} = \frac{n*(n-1)}{2}.$$

4/

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

5/

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

□

Développement du binôme et triangle de Pascal

Le développement du polynôme

$$(a+b)^n = \underbrace{(a+b) \times (a+b) \times ... \times (a+b)}_{n\ fois}$$

nécessite calculer pour chaque terme a^k (ou b^k) avec $0 \leq k \leq n$, le nombre de façon distinctes de choisir k fois a (ou b) parmi n possibilités. Ce nombre est le coéfficient binomiale donné par :

$$C_n^k = \frac{n!}{k! \times (n-k)!}$$

et on a

$$\begin{array}{lccccccccccccc}
(a+b)^0 = & 1 & & & & & & & & & & & & \\
(a+b)^1 = & 1 \times a & + & 1 \times b & & & & & & & & & & \\
(a+b)^2 = & a^2 & + & 2 \times a \times b & + & b^2 & & & & & & & & \\
(a+b)^3 = & a^3 & + & 3 \times a^2 \times b & + & 3 \times a \times b^2 & + & b^3 & & & & & & \\
(a+b)^4 = & a^4 & + & 4 \times a^3 \times b & + & 6 \times a^2 \times b^2 & + & 4 \times a \times b^3 & + & b^4 & & & & \\
. & . & & . & . & . & & & & & & & & \\
. & . & & . & . & . & & & & & & & & \\
 & . & & . & . & . & & & & & & & & \\
(a+b)^n = & a^n & + & C_n^1 \times a^{n-1} \times b & + & C_n^2 \times a^{n-2} \times b^2 & + & .... & & .. & .. & + & C_n^n \times b^n &
\end{array}$$

Les coefficients binomiaux forment un triangle dont les lignes correspondent à n constant, alors la somme de deux coefficients consécutifs d'une ligne est égale au terme de la ligne suivante en dessous du deuxième terme.

Exemple 3.1.9 *Triangle de Pascal en cas $n = 7$.*

$$\begin{array}{cccccccc}
1 \\
1 & 1 \\
1 & 2 & 1 \\
1 & 3 & 3 & 1 \\
1 & 4 & 6 & 4 & 1 \\
1 & 5 & 10 & 10 & 5 & 1 \\
1 & 6 & 15 & 20 & 15 & 6 & 1 \\
1 & 7 & 21 & 35 & 35 & 21 & 7 & 1
\end{array}$$

3.2 Exercices

Exercice 1 : Soit l'ensemble $E = \{2,\ 3,\ 5,\ 6,\ 7,\ 9\ \}$. Avec les chiffres de cet ensemble

1/ Combien peut on avoir de nombres de 3 chiffres (avec et sans répétition) ?

2/ Parmi ceux-ci, combien sont inférieurs à 400 ?

3/ Parmiceux-ci, combien sont pairs ?

4/ Parmi ceux-ci, combien sont multiples de 5 ?

Exercice 2 : De combien de manières différentes peut-on mettre 3 personnes en rang.

Exercice 3 : Une urne contient 5 boules blanches et 3 boules rouges.

1/ Combien y-a-t-il de possibilités d'extraire de l'urne 3 boules dont 2 sont blanches et l'autre est rouge ?

2/ Combien y-a-t-il de possibilités d'extraire successivement de l'urne une boule blanche, une boule rouge et une autre boule blanche.

Exercice 4 : Calculer $\sum_{k=0}^{n} C_n^k$ en fonction de n, puis en déduire $\sum_{k=0}^{10} C_{10}^k$.

Exercice 5 : Résoudre l'équation

$$C_{2x+2}^{11-x} = C_{2x+2}^{2x-1}.$$

Exercice 6 : Une commission de 5 membres comprenant 3 économistes et 2 juristes doit être constituée à partir de 13 condidats se divisant en 7 économistes et 6 juristes.

1/ De combien de façons différentes cette commission peut être constituée.

2/ Même question, un économiste nomément désigné parmi les 7 économistes condidats devant absolument faire parti de la commission.

3/ Même question, un de 6 juristes condidats devant être écarté de la commission.

4/ Même question, un économiste et un juriste ne pouvant pas faire partie ensemble de la commission.

3.3 Corrigés

Exercice 1 :

On considère l'ensemble $E = \{2, 3, 5, 6, 7, 9\}$. Il est clair que le cardinal de l'ensemble E est :
$Card\ E = 6$.

1. **Le nombre des nombres composés de 3 chiffres :**

   Un nombre composé de 3 chiffres est écrit sous la forme ABC.

   (a) **Avec répétition :**

   Soit N_1 le nombre des nombres composés de 3 chiffres avec répétition. Ainsi, chaque chiffre du nombre ABC possède 6 propositions :

$$\begin{array}{ccc} A & B & C \\ 6 & 6 & 6 \end{array}$$

   Donc, c'est un arrangement A_n^p avec répétition où $p = 3$ et $n = 6$.

   D'où,

$$\begin{aligned} N_1 &= A_6^3 \\ &= 6^3 \\ &= 216. \end{aligned}$$

En dernier, on a 216 nombres composés de 3 chiffres avec répétition.

(b) **Sans répétition :**

Soit N_2 le nombre des nombres composés de 3 chiffres sans répétition. Ainsi, le premier chiffre A possède 6 propositions, le second B possède 5 propositions et le troisième C a 4 propositions :

$$\begin{array}{ccc} A & B & C \\ 6 & 5 & 4 \end{array}$$

Donc, c'est un arrangement A_n^p sans répétition où $p = 3$ et $n = 6$.

D'où :

$$\begin{aligned} N_2 &= A_6^3 \\ &= \frac{6!}{3!} \\ &= 6 \times 5 \times 4 \\ &= 120. \end{aligned}$$

En dernier, on a 120 nombres composés de 3 chiffres sans répétition.

2. **Le nombre des nombres inférieurs à 400 :**

Un nombre composé de 3 chiffres inférieur à 400 est écrit sous la forme ABC avec $A = 2$ ou $A = 3$.

(a) **Avec répétition :**

Soit N_3 le nombre des nombres composés de 3 chiffres avec répétition inférieurs à 400. Ainsi, le premier chiffre A possède 2 propositions et les deux autres chiffres B et C possèdent 6 propositions :

$$\begin{array}{ccc} A & B & C \\ 2 & 6 & 6 \end{array}$$

Donc, c'est un arrangement A_n^p avec répétition où $p = 2$ et $n = 6$ compté deux fois.

D'où :

$$\begin{aligned} N_3 &= 2 \times A_6^2 \\ &= 2 \times 6^2 \\ &= 72. \end{aligned}$$

En dernier, on a 72 nombres composés de 3 chiffres inférieurs à 400 avec répétition.

(b) **Sans répétition :**

Soit N_4 le nombre des nombres composés de 3 chiffres sans répétition inférieurs à 400. Ainsi, le premier chiffre A possède 2 propositions, le second B possède 5 propositions et le troisième C a 4 propositions :

$$\begin{array}{ccc} A & B & C \\ 2 & 5 & 4 \end{array}$$

Donc, c'est un arrangement A_n^p sans répétition où $p = 2$ et $n = 5$ compté deux fois.

D'où :

$$\begin{aligned} N_4 &= 2 \times A_5^2 \\ &= 2 \times \frac{5!}{3!} \\ &= 2 \times 5 \times 4 \\ &= 40. \end{aligned}$$

En dernier, on a 40 nombres composés de 3 chiffres sans répétition inférieurs à 400.

3. **Le nombre des nombres pairs :**

Un nombre composé de 3 chiffres est pair s'il est écrit sous la forme ABC avec $C = 2$ ou $C = 6$.

(a) **Avec répétition :**

Soit N_5 le nombre des nombres composés de 3 chiffres avec répétition pairs. Ainsi, les deux premiers chiffres A et B possèdent 6 propositions et le dernier chiffre C possède 2 propositions.

$$\begin{array}{ccc} A & B & C \\ 6 & 6 & 2 \end{array}$$

Donc, c'est un arrangement A_n^p avec répétition où $p = 2$ et $n = 6$ compté deux fois.

D'où,

$$\begin{aligned} N_5 &= 2 \times A_6^2 \\ &= 2 \times 6^2 \\ &= 72. \end{aligned}$$

En dernier, on a 72 nombres composés de 3 chiffres inférieurs pairs.

(b) **Sans répétition :**

Soit N_6 le nombre des nombres composés de 3 chiffres sans répétition pairs. Ainsi, le dernier chiffre C possède 2 propositions, le premier chiffre A possède 5 propositions et le deuxième chiffre B a 4 propositions (ici, on a commencé par le dernier chiffre vu que les nombres doivent être pairs) :

$$\begin{array}{ccc} A & B & C \\ 5 & 4 & 2 \end{array}$$

Donc, c'est un arrangement A_n^p sans répétition où $p = 2$ et $n = 5$ compté deux fois.

D'où :

$$\begin{aligned} N_6 &= 2 \times A_5^2 \\ &= 2 \times \frac{5!}{3!} \\ &= 2 \times 5 \times 4 \\ &= 40. \end{aligned}$$

En dernier, on a 40 nombres composés de 3 chiffres sans répétition pairs.

4. **Le nombre des nombres multiples de 5 :**

Un nombre composé de 3 chiffres est multiple de 5 s'il est écrit sous la forme ABC avec $C = 5$.

(a) **Avec répétition :**

Soit N_7 le nombre des nombres composés de 3 chiffres avec répétition multiple de 5. Ainsi, les premiers chiffres A et B possèdent 6 propositions et le dernier chiffre C possède une seule proposition :

$$\begin{array}{ccc} A & B & C \\ 6 & 6 & 1 \end{array}$$

Donc, c'est un arrangement A_n^p avec répétition où $p = 2$ et $n = 6$.

D'oÃ¹ :

$$\begin{aligned} N_7 &= A_6^2, \\ &= 6^2, \\ &= 36. \end{aligned}$$

En dernier, nous avons 36 nombres composés de 3 chiffres multiples de 5.

(b) **Sans répétition :**

Soit N_8 le nombre des nombres composés de 3 chiffres sans répétition multiples de 5. Ainsi, le dernier chiffre C possède une proposition, le premier chiffre A possède 5 propositions et le deuxième chiffre B a 4 propositions (ici, on a commencé par le dernier chiffre vu que les nombres doivent être multiples de 5).

$$\begin{array}{ccc} A & B & C \\ 5 & 4 & 1 \end{array}$$

Donc, c'est un arrangement A_n^p sans répétition où $p = 2$ et $n = 5$.

D'où :

$$\begin{aligned} N_8 &= A_5^2 \\ &= \frac{5!}{3!} \\ &= 5 \times 4 \\ &= 20. \end{aligned}$$

En dernier, on a 20 nombres composés de 3 chiffres sans répétition multiples de 5.

Exercice 2 : On veut mettre 3 personnes en rang

Le nombre de manières différentes pour mettre 3 personnes en rang :

Soit N le nombre de manières différentes pour mettre 3 personnes XYZ en rang. Ainsi, X possède 3 propositions, Y a 2 proposions et une proposition pour Z :

$$\begin{array}{ccc} X & Y & Z \\ 3 & 2 & 1 \end{array}$$

Donc, c'est une permutation P de $n = 3$ éléments sans répétition.

D'où :

$$\begin{aligned} N &= P_3 \\ &= 3! \\ &= 6. \end{aligned}$$

En dernier, on a 6 différentes manières pour mettre 3 personnes en rang.

Exercice 3 :

Une urne contient 5 boules blanches et 3 boules rouges.

1/ **Le nombre de possibilités d'extraire de l'urne 3 boules dont 2 blanches et une rouge :**

Soit N_1 le nombre de possibilités d'extraire de l'urne 3 boules dont 2 blanches et une rouge. Dans ce cas, c'est le produit de deux combinaisons l'une pour les boules blanches et l'autre pour la boule rouge.

- Boules blanches : C'est une combinaison C_n^p de $p = 2$ boules prise sans remise parmi les $n = 5$ boules blanches
- Boule rouge : C'est une combinaison C_n^p de $p = 1$ boule prise sans remise parmi les $n = 3$ boules rouges.

Ainsi,

$$\begin{aligned} N_1 &= C_5^2 \times C_3^1 \\ &= \frac{5!}{2!3!} \times \frac{3!}{1!2!} \\ &= 30. \end{aligned}$$

Donc, on a 30 possibilités d'extraire de l'urne 3 boules dont 2 blanches et une rouge.

2/ **Le nombre de possibilités d'extraire successivement de l'urne une boule blanche, une boule rouge et une boule blanche**

Soit N_2 le nombre de possibilités d'extraire successivement de l'urne une boule blanche, une boule rouge et une autre boule blanche. Dans ce cas, c'est le produit de trois combinaisons une pour la première boule blanche, la deuxième pour la boule rouge et la dernière pour la boule blanche.

- Boule blanche : C'est une combinaison C_n^p de $p = 1$ boule prise sans remise parmi les $n = 5$ boules blanches

- Boule rouge : C'est une combinaison C_n^p de $p=1$ boule prise sans remise parmi les $n=3$ boules rouges.
- Boule blanche : C'est une combinaison C_n^p de $p=1$ boule prise sans remise parmi les $n=4$ boules blanches (ici, le nombre de boules blanches devient 4).

  Ainsi,

$$\begin{aligned} N_2 &= C_5^1 \times C_3^1 \times C_4^1 \\ &= \frac{5!}{1!4!} \times \frac{3!}{1!2!} \times \frac{4!}{1!3!} \\ &= 60. \end{aligned}$$

Alors, on a 60 possibilités d'extraire successivement de l'urne une boule blanche, une boule rouge et une autre boule blanche.

Exercice 4 :

1/ **Calculer $\sum_{k=0}^{n} C_n^k$:**

On sait que le développement d'un binôme est donné par :

$$\forall n \in \mathbb{N}^* \;:\; (a+b)^n = \sum_{k=0}^{n} C_n^k \, a^{n-k} \, b^k.$$

Ainsi, pour $a=1$ et $b=1$, on aura

$$\begin{aligned} \forall n \in \mathbb{N}^* \;:\; \sum_{k=0}^{n} C_n^k &= (1+1)^n \\ &= 2^n. \end{aligned}$$

2/ **Déduire $\sum_{k=0}^{10} C_{10}^k$:**

De la première question, on aura :

$$\begin{aligned} \sum_{k=0}^{10} C_{10}^k &= 2^{10} \\ &= 1024. \end{aligned}$$

Exercice 5 : On a

$$C_{2x+2}^{11-x} = C_{2x+2}^{2x-1},$$

ce qui implique

$$11 - x = 2x - 1.$$

D'où $x = 4$.

Exercice 6 :

Une commission de 5 membres comprenant 3 économistes et 2 juristes doit être constituée à partir de 13 candidats se divisant en 7 économistes et 6 juristes.

1/ **Le nombre des différentes façons possibles pour construite la commission :**

Soit N_1 le nombre des différentes façons possibles pour construite la commission. Dans ce cas, c'est le produit de deux combinaisons l'une pour les économistes et l'autre pour les juristes.

- Économistes : C'est une combinaison C_n^p de $p = 3$ économistes prises sans remise parmi les $n = 7$ économistes.
- Juristes : C'est une combinaison C_n^p de $p = 2$ juristes prises sans remise parmi les $n = 6$ juristes.

Ainsi,

$$\begin{aligned} N_1 &= C_7^3 \times C_6^2 \\ &= \frac{7!}{3!4!} \times \frac{6!}{2!4!} \\ &= 525. \end{aligned}$$

Alors, on a 525 possibilités de construire cette commission.

2/ **Le nombre des différentes façons possibles pour construire la commission où un économiste nomément désigné parmi les 7 économistes condidats devant absolument faire parti de la commission :**

Soit N_2 le nombre des différentes façons possibles pour construite cette commission. Dans ce cas, c'est le produit de deux combinaisons l'une pour les économistes et l'autre pour les juristes.

- Économistes : C'est une combinaison C_n^p de $p = 2$ économistes prises sans remise parmi les $n = 6$ économistes vu qu'un économiste fait partie de la commission.
- Juristes : C'est une combinaison C_n^p de $p = 2$ juristes prises sans remise parmi les $n = 6$ juristes.

Ainsi,

$$\begin{aligned} N_2 &= C_6^2 \times C_6^2 \\ &= \frac{6!}{2!4!} \times \frac{6!}{2!4!} \\ &= 225. \end{aligned}$$

Donc : on a 225 possibilités de construire cette commission.

3/ **Le nombre des différentes façons possibles pour construite la commission où l'un des** 6 **juristes candidats devait être écarté de la commission :**

Soit N_3 le nombre des différentes façons possibles pour construite cette commission. Dans ce cas, c'est le produit de deux combinaisons l'une pour les économistes et l'autre pour les juristes.

- Économistes : C'est une combinaison C_n^p de $p = 3$ économistes prises sans remise parmi les $n = 7$ économistes.
- Juristes : C'est une combinaison C_n^p de $p = 2$ juristes prises sans remise parmi les $n = 5$ juristes vu qu'un juriste a été écarté de la commission.

Ainsi,

$$\begin{aligned} N_3 &= C_7^3 \times C_5^2 \\ &= \frac{7!}{3!4!} \times \frac{5!}{2!3!} \\ &= 350. \end{aligned}$$

Donc, on a 350 possibiliés de construire cette commission pour ce cas.

4/ **Le nombre des différentes façons possibles pour construite la commission où un économiste et un juriste ne peuvent pas faire partie ensemble de la commission :**

Soit N_4 le nombre des différentes façons possibles pour construite cette commission. On suppose que l'économiste écarté est X et le juriste écarté est Y.

Comme on a écarté un économiste et un juriste, alors, le nombre des économistes devient 6 et celui des juristes devient 5.

- Première méthode : N_4 est égale au nombre de toutes les commissions possible $-$ le nombre des commissions contenant X et Y.

  Le le nombre des commissions contenant X et Y est $C_6^2\,C_5^1$ (car, on a déjà pris l'éonomiste X et le juriste Y, donc le nombre des membres de la commission est 5). Ainsi,

$$\begin{aligned} N_4 &= 525 - C_6^2\,C_5^1 \\ &= 525 - 15 \times 5 \\ &= 450. \end{aligned}$$

Alors, on a 450 possibilités de construire cette commission.

- Deuxième méthode : N_4 est égale à la somme du nombre de toutes les commissions contenant l'économiste X mais pas le juriste Y + le nombre de toutes les commissions contenant le juriste Y mais pas l'économiste X + le nombre de toutes les commissions qui ne contiennent ni l'économiste X ni le juriste Y, donc,

$$\begin{aligned} N_4 &= C_6^2\,C_5^2 + C_6^3\,C_5^1 + C_6^3\,C_5^2 \\ &= 15 \times 10 + 20 \times 5 + 20 \times 10 \\ &= 450. \end{aligned}$$

Alors, on a 450 possibilités de construire cette commission.

Chapitre 4

Calcul de Probabilités

Dans les différents domaines scientifiques (biologie, médecine, sociologie...), on s'intéresse à de nombreux phénomènes dans les quels apparait souvent l'effet du hasard.
Ces phénomènes sont caractérisés par le fait que les résultats varient d'une ecpérience à l'autre.
Une expérience est appelée aléatoire s'il est impossible de prévoir son résultat c'est-à-dire, elle peut donner des résultats différents même si on la répète dans des conditions identiques.

4.1 Définitions

Définition 4.1.1 ***"l'espace probabilisé"*** *On appelle espace probabilisé le triplet* $(\Omega,\ \mathcal{F}, P)$ *où*

i/ Ω est l'univers : l'ensemble des évenements élémentaires d'une expérience aléatoire.

ii/ $\mathcal{F}$ est la tribu, elle est composée de tous les sous ensembles de Ω appelés les événements et vérifiant :

1/ $\Omega \in \mathcal{F}$.

2/ $\forall A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$.

3/ $\forall A,\ B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.

4/ $\forall \left(A_n\right)_{n\in I} \in \mathcal{F},\ I \subseteq \mathbb{N} \Rightarrow \underset{n\in I}{U} A_n \in \mathcal{F}$.

iii/ P est une application définie par :

$$\begin{array}{rccl} P: & \mathcal{F} & \to & [0,1] \\ & A & \mapsto & P(A) \end{array}$$

telle que

1/ $\forall A \in \mathcal{F}/\ (A \neq \emptyset \text{ } et \text{ } A \neq \Omega),\ \ 0 < P(A) < 1.$

2/ $P(\Omega) = 1$, l'événement Ω est dit certain.

3/ $P(\emptyset) = 0$, l'événement $\emptyset$ (l'ensemble vide) est dit impossible.

4/ $\forall A \in \mathcal{F},\ P\left(\bar{A}\right) = 1 - P(A)$ où $\bar{A}$ est le complémentaire de A dans Ω vérifiant $A \cup \bar{A} = \Omega$ et $A \cap \bar{A} = \emptyset$.

5/ $\forall\, (A_i)_{1 \leq i \leq n} \in \mathcal{F}$ une suite d"événements disjoints deux à deux, on a

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

6/ $\forall A \in \mathcal{F}, \forall B \in \mathcal{F},\ P(A \cup B) = P(A) + P(B) - P(A \cap B).$

7/ Deux événements A et B sont dits équiprobables si $P(A) = P(B)$.

Remarque 4.1.1 *Si l'ensemble Ω est fini ou dénombrable, on parle de probabilité discrète, le cas contraire est celui de probabilité continue.*

Définition 4.1.2 *Soient $(\Omega,\ \mathcal{F}, P)$ un espace probabilisé discret et $A \in \mathcal{F}$ un événement. La probabilité de A est :*

$$P(A) = \frac{Card\ A}{Card\ \Omega}.$$

Exemple 4.1.1 *Lorsqu'on lance une pièce de monnaie, on ne sait pas si c'est pile ou face qui sera visible et chacun des deux (pile, face) a une chance sur deux d'apparaitre $\Rightarrow$ $\Omega = \{pile,\ face\}$ et la tribu $\mathcal{F} = \mathcal{P}(\Omega) = \{\emptyset,\ \{pile\},\ \{face\},\ \Omega\}$ et la probabilité P définie pour tout évènement A par :*

$$\begin{array}{rccl} P: & \mathcal{F} & \to & [0,1] \\ & A & \mapsto & P(A) \end{array}$$

vérifie :

1/ Si $A = \emptyset$: $P(A) = 0$.

2/ Si $A = \{pile\} : P(A) = \frac{1}{2}$.

3/ Si $A = \{face\} : P(A) = \frac{1}{2}$.

4/ Si $A = \Omega : P(A) = 1$.

Exemple 4.1.2 *Le lancer d'un dé repose sur un principe similaire, les 6 faces du dé ont autant de chances d'apparaitre* $\frac{1}{6} \Rightarrow \Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$ *et* $P\{1\} = P\{2\} = ... = P\{6\} = \frac{1}{6}$. *On considère les évènements suivants :*

1/ Si A :"avoir un nombre pair"$\rightarrow A = \{2,\ 4,\ 6\} \rightarrow P(A) = \frac{CardA}{Card\Omega} = \frac{3}{6} = \frac{1}{2}$.

2/ Si A :"avoir un nombre impair"$\rightarrow A = \{1,\ 3,\ 5\} \rightarrow P(A) = \frac{1}{2}$.

3/ Si A :"avoir un nombre supérieur à 5"$\rightarrow A = \{\ 6\} \rightarrow P(A) = \frac{1}{6}$.

4.2 Probabilité conditionnelle

Soient $(\Omega,\ \mathcal{F}, P)$ un espace probabilisé et $A, B \in \mathcal{F}$ tel que $P(B) > 0$.

Définition 4.2.1 *La probabilité conditionnelle de l'évènement A sachant l'évènement B est définie par :*

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

Proposition 4.2.1 *La probabilité conditionnelle vérifie les propriétés suivantes :*

1/ $P(\Omega/B) = 1$.

2/ $P(A \cap B) = P(A/B) \times P(B)$.

3/ $P(\bar{A}/B) = 1 - P(A/B)$.

Preuve. Les deux premières propriétés sont évidentes, il suffit de démontrer la troisième. On a

$$\begin{aligned} A \cup \bar{A} = \Omega \quad &\Rightarrow \quad (A \cup \bar{A}) \cap B = \Omega \cap B \\ &\Rightarrow \quad (A \cap B) \cup (\bar{A} \cap B) = B \end{aligned}$$

d'où

$$P\left[(A \cap B) \cup (\bar{A} \cap B)\right] = P(B).$$

Comme $(A \cap B) \cap \left(\bar{A} \cap B\right) = \emptyset$, alors

$$P\left[(A \cap B) \cup \left(\bar{A} \cap B\right)\right] = P(A \cap B) + P\left(\bar{A} \cap B\right) = P(B)$$

en divisant par $P(B)$, on obtient

$$\frac{P(A \cap B)}{P(B)} + \frac{P\left(\bar{A} \cap B\right)}{P(B)} = 1$$

ce qui implique

$$P\left(\bar{A}/B\right) = 1 - P(A/B).$$

$\square$

Définition 4.2.2 *On dit que A et B sont deux évènements indépendants si $P(A/B) = P(B)$, c'est à dire $P(A \cap B) = P(A) \times P(B)$.*

Exemple 4.2.1 *On lance une pièce de monnaie 3 fois et on considère les évènements suivants :*

A : "on obtient face en premier lancement".

B : "on obtient face en deuxième lancement".

C : "on obtient face en deux lancements successifs pas plus".

En calculant les probabilités des 3 évènements cités ci-dessus, on obtient

1/ $P(A) = \frac{1}{2} \times 1 \times 1 = \frac{1}{2}$.

2/ $P(B) = 1 \times \frac{1}{2} \times 1 = \frac{1}{2}$.

3/ $P(C) = \frac{1}{2} \times \frac{1}{2} \times 1 + 1 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$.

Comme $P(A \cap B) = \frac{1}{2} \times \frac{1}{2} \times 1 = \frac{1}{4} = P(A) \times P(B)$, alors les deux évènements A et B sont indépendants. De même on a $P(A \cap C) = \frac{1}{2} \times \frac{1}{2} \times 1 = \frac{1}{4} = P(A) \times P(C) \Rightarrow$ les deux évènements A et C sont indépendants et $P(B \cap C) = 1 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = P(B) \times P(C) \Rightarrow$ les deux évènements B et C sont indépendants.

4.3 Exercices

Exercice 1 : Le code d'une carte bancaire est composé de 4 chiffres. Calculer la probabilité des évènements suivants :

1/ A : "avoir un code composé de 2 chiffres pairs".

2/ B : "avoir un code composé de 4 chiffres égaux".

3/ C : "avoir un code contient au moins 1 chiffre premier".

Exercice 2 : Un agriculteur a acheté 15 petits d'arbres d'agrumes dont 8 sont des orangers, 5 sont des citronniers et le reste des pamplemoussiers. Il a pris 3 arbres au hasard pour les planter dans le jardin de sa maison. Calculer la probabilité que les trois arbres plantés dans le jardin de la maison soient :

1/ A : "un oranger, un citronnier et un pamplemoussier".

2/ B : "des orangers".

3/ C : "deux orangers et un citronnier".

Exercice 3 : On prend au hasard et en même temps, 3 néons dans un lot de 15 dont 5 sont défectueuses. Calculer la probabilité des évènements :

1/ A : "au moins un est défectueux".

2/ B : "les 3 sont défectueux".

3/ C : "exactement un néon est défectueux".

Exercice 4 : Un industriel doit vérifier l'état de marche de ses machines. D'après des statistiques récentes, il a évalué à 15% la probabilité pour qu'une machine tombe en panne en 3 ans parmi toutes ees machines.
La probabilité de devenir hors d'usage suite à une panne grave est évaluée à 80%, cette probabilité est de 10% pour une machine n'ayant jamais eu de panne.

1/ Calculer la probabilité pour une machine donnée de plus de 3 ans d'être hors d'usage ?

2/ En déduire la probabilité pour une machine donnée de plus de 3 ans ne tombe pas en panne (reste en usage) ?

Exercice 5 : Un fumeur décide d'arrêter de fumer, après avoir lu une série de statistiques effrayantes sur les risques de santé liés au tabac.
D'après des statistiques, on estime les probabilités suivantes :

i/ Si cette personne n'a pas fumé un jour J_n, alors la probabilité pour qu'elle ne fume pas le jour suivant J_{n+1} est $0,3$.

ii/ Si elle a fumé un jour J_n, alors la probabilité pour qu'elle ne fume pas le jour suivant J_{n+1} est $0,9$.

1/ Quelle est la probabilité P_{n+1} pour qu'elle fume le jour J_{n+1} en fonction de la probabilité P_n pour qu'elle fume le jour J_n?

2/ Calculer la limite de P_n.

3/ Va-t-elle finir par s'arrêter de fumer ?

4.4 Corrigés

Exercice 1 :

On considère l'ensemble $E = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Il est clair que le cardinal de l'ensemble E est : $CardE = 10$.

1/ **La probabilité $P(A)$ pour avoir un code composé de deux chiffres paires :**

On a

$$\begin{aligned} P(A) &= \frac{\text{nombre des codes composés de deux chiffres paires}}{\text{nombre de tous les codes possibles}} \\ &= \frac{n}{N} \end{aligned}$$

où

$$N = A_{10}^4 = 10^4, \text{ un arrangement avec répétition,}$$

et

$$\begin{aligned} n &= A_5^2 A_5^2 \text{ un arrangement avec répétition pour les chiffres pairs et impairs} \\ &= 5^2.5^2 \\ &= 625. \end{aligned}$$

car le nombre de chiffres pairs de l'ensemble E = au nombre de chiffres impairs = 5.

$$\begin{aligned} P(A) &= \frac{n}{N} \\ &= \frac{625}{10000} \\ &= 0,0625. \end{aligned}$$

2/ **La probabilité $P(B)$ pour avoir un code composé de 4 chiffres égaux :**

On a

$$\begin{aligned} P(B) &= \frac{\text{nombre des codes composés de 4 chiffres égaux}}{\text{nombre de tous les codes possibles}} \\ &= \frac{10}{10000} \\ &= 0,001. \end{aligned}$$

1. **La probabilité $P(C)$ pour avoir un code contenant au moins un chiffre premier :**

Les chiffres premiers de l'ensemble E sont : $\{2, 3, 5, 7\}$.

On a

$$\begin{aligned} P(C) &= \frac{\text{nombre des codes contenant au moins un chiffre premier}}{\text{nombre de tous les codes possibles}} \\ &= \frac{n}{N}, \end{aligned}$$

avec $N = 10000$ et

$$\begin{aligned} n &= A_4^1 \times A_6^3 + A_4^2 \times A_6^2 + A_4^3 \times A_6^1 + A_4^4 \\ &\qquad \text{(car l'ensemble } E \text{ a 4 chiffres premiers et 6 chiffres non premiers)} \\ &= 4 \times 6^3 + 4^2 \times 6^2 + 4^3 \times 6 + 4^4 \\ &= 2080. \end{aligned}$$

Donc

$$\begin{aligned} P(C) &= \frac{2080}{10000} \\ &= 0,2080. \end{aligned}$$

Exercice 2 :

On calcule, d'abord, le nombre de possibilités N des 3 arbres :

On prend trois arbres d'un ensemble de 15 arbres, alors

$$\begin{aligned} N &= C_{15}^{3} \\ &= \frac{15!}{3!12!} \\ &= 455. \end{aligned}$$

1/ **La probabilité $P(A)$ pour planter un oranger, un citronnier et un pamplemoussier :**

On a

$$\begin{aligned} P(A) &= \frac{C_8^1 C_5^1 C_2^1}{C_{15}^3} \\ &= \frac{8 \times 5 \times 2}{455} \\ &= 0,1758. \end{aligned}$$

2/ **La probabilité $P(B)$ pour planter des orangers :**

On a

$$\begin{aligned} P(B) &= \frac{C_8^3}{C_{15}^3} \\ &= \frac{56}{455} \\ &= 0,1231. \end{aligned}$$

3/ **La probabilité $P(C)$ pour planter deux orangers et un citronnier :**

On a

$$\begin{aligned} P(C) &= \frac{C_8^2 C_5^1}{C_{15}^3} \\ &= \frac{28.5}{455} \\ &= 0,3077. \end{aligned}$$

Exercice 3 :

Tout d'abord, on calcule le nombre de possibilités N des 3 néons. On prend trois néons d'un lot de 15 néons.

$$\begin{aligned} N &= C_{15}^3 \\ &= 455. \end{aligned}$$

1/ **La probabilité $P(A)$ pour qu'au moins un néon soit défectueux :**

On a

$$\begin{aligned} P(A) &= \frac{C_5^1 C_{10}^2 + C_5^2 C_{10}^1 + C_5^3}{C_{15}^3} \\ &= \frac{5 \times 45 + 10 \times 10 + 10}{455} \\ &= 0,7363. \end{aligned}$$

2/ **La probabilité $P(B)$ pour que les 3 néons soient défectueux :**

On a

$$\begin{aligned} P(B) &= \frac{C_5^3}{C_{15}^3} \\ &= \frac{10}{455} \\ &= 0,0220. \end{aligned}$$

3/ **La probabilité $P(C)$ pour avoir exactement un néon défectueux :**

On a

$$\begin{aligned} P(C) &= \frac{C_5^1 C_{10}^2}{C_{15}^3} \\ &= \frac{5\times 45}{455} \\ &= 0,4945. \end{aligned}$$

Exercice 4 **:** On a

- $15\% = 0,15$ est la probabilité de l'évènement "Panne", qu'on note $panne$. Parmi ces 15%, on a
  - ☐ $80\% = 0,80$ est la probabilité de l'évènement "devenir hors d'usage suite à une panne grave", qu'on note $HU/panne$.
  - ☐ $10\% = 0,10$ est la probabilité de l'évènement "n'ayant jamais tomber en panne", qu'on note $HU/non\,panne$.
- $85\% = 0,85 = 100\% - 15\%$ est la probabilité de l'évènement " Non Panne", qu'on note $non\,panne$.

1/ La probabilité $P(HU)$ d'une machine donnée de plus de 3 ans d'être hors d'usage **:**

$$\begin{aligned} P(HU) &= P(HU/panne)P(panne) + P(HU/non\ panne)P(non\ panne) \\ &= 0,80\times 0,15 + 0,10\times 0,85 \\ &= 0,205. \end{aligned}$$

2/ La probabilité $P(U)$ d'une machine donnée de plus de 3 ans reste en usage **:**

$$\begin{aligned} P(U) &= 1 - P(HU) \\ &= 1 - 0,205 \\ &= 0,795. \end{aligned}$$

Exercice 5 **:** Soit l'évènement F_n : "*fumer le* $n^{\text{ème}}$ *jour*", $n\in\mathbb{N}^*$ tel que $P_n = P(F_n)$.

1/ Il est clair que

$$P\left(\overline{F_n}\right) = 1 - P_n,\ \forall n\in\mathbb{N}^*.$$

On a

$$P\left(\overline{F_{n+1}}/F_n\right) = 0,9 \ et \ P\left(\overline{F_{n+1}}/\overline{F_n}\right) = 0,3$$

et

$$\begin{aligned} P\left(\overline{F_{n+1}}\right) &= P\left(\overline{F_{n+1}}/F_n\right) P\left(F_n\right) + P\left(\overline{F_{n+1}}/\overline{F_n}\right) P\left(\overline{F_n}\right) \\ &= 0,9P_n + 0,3\left(1 - P_n\right). \end{aligned}$$

Alors

$$1 - P_{n+1} = 0,9P_n + 0,3\left(1 - P_n\right)$$

D'où

$$P_{n+1} = -0,6P_n + 0,7.$$

2/ La limite l de P_n qunad $n \to +\infty$ (si elle existe) vérifie l'équation

$$\begin{aligned} l &= -0,6l + 0,7 \\ &\Rightarrow l = 0,4375. \end{aligned}$$

Donc

$$\lim_{n \to +\infty} P_n = 0,4375 = 43,75\%.$$

3/ Comme $P_n = P\left(F_n\right) = 43,75\%$, cette personne ne va pas finir par s'arrêter de fumer ($\lim\limits_{n \to +\infty} P_n \neq 0$).

Chapitre 5

Variables aléatoires discrètes

5.1 Définitions et propriétés

Définition 5.1.1 *Soit* $(\Omega,\ \mathcal{F}, P)$ *un espace probabilisé, une variable aléatoire discrète* X *associée à cet espace probabilisé est une application de* Ω *dans* $\mathbb{R}$ *qui prend un nombre de valeurs fini ou dénombrable satisfaisant* $\sum_{x \in X} P(X = x) = 1$. *On note les éléments de la variable aléatoire discrète* X *par* $x_1,\ x_2,\ ...,\ x_i, ...,\ x_n$ *où* $n \in \mathbb{N}$. *Pour tout* $i \in \mathbb{N}$, *on pose*

$$P_X(X = x_i) = P_i = P\left(X^{-1}(x_i)\right),$$

et

$$\sum_{i=1}^{n} P_i = 1.$$

Exemple 5.1.1 *On lance 3 pièces de monnaie et on observe le nombre de faces obtenues. Il est clair que* X *prend les valeurs : 0, 1, 2, 3. Alors l'application*

$$X: \ \Omega \ \rightarrow \ \mathbb{R}$$

est une variable aléatoire telle que

$$\Omega = \{(ppp),\ (ppf),\ (pfp),\ (fpp),\ (pff),\ (fpf),\ (ffp),\ (fff)\}$$

et

$$X(\Omega) = \{0,\ 1,\ 2,\ 3\}$$

en plus, on a

$$P(X=0)=P\left[(ppp)\right]=\left(\frac{1}{2}\right)^3,$$
$$P(X=1)=P\left[(ppf)\,,\ (pfp)\,,\ (fpp)\right]=3\left(\frac{1}{2}\right)^3,$$
$$P(X=2)=P\left[(pff)\,,\ (fpf)\,,\ (ffp)\right]=3\left(\frac{1}{2}\right)^3,$$
$$P(X=3)=P\left[(fff)\right]=\left(\frac{1}{2}\right)^3.$$

Ces résultats sont résumés dans le tableau suivant :

X	0	1	2	3	$\sum_{x=0}^{3} P\left(X=x\right)$
P_X	$\left(\frac{1}{2}\right)^3$	$3\left(\frac{1}{2}\right)^3$	$3\left(\frac{1}{2}\right)^3$	$\left(\frac{1}{2}\right)^3$	1

Remarque 5.1.1 *Si X et$Y$ sont deux variables alétoires discrètes définies sur le même espace preobabilisé, alors $X+Y$ est aussi une variable aléatoire discrète définie sur le même espace.*

Dans tous ce qui suit, on considère la variable alétoire discrète $X=(x_1,\ x_2,\ ...,\ x_n)\,,\, n\in\mathbb{N}$.

Définition 5.1.2 ***L'espérance mathématique*** *de la variable aléatoire discrète X est le nombre*

$$E(X)=\sum_{i=1}^{n} P_i x_i.$$

Remarque 5.1.2 *L'espérance mathématique d'une variable aléatoire discrète X est la version probabiliste de la moyenne arithmétique.*

Proposition 5.1.1 *Soient X et Y deux variables aléatoires discrètes définies sur le même espace probabilisé. Alors :*

1/ $E(X+Y)=E(X)+E(Y)$.

2/ $E(aX+b)=aE(X)+b, \forall (a,b)\in\mathbb{R}^2$.

Preuve. Soient $X=(x_1,\ x_2,\ ...,\ x_n)$ et $Y=(y_1,\ y_2,\ ...,\ y_n)\,,\ n\in\mathbb{N}$ deux variables alétoires, on a :

1/

$$\begin{aligned} E(X+Y) &= \sum_{i=1}^{n} P_i\left(x_i+y_i\right) \\ &= \sum_{i=1}^{n} P_i x_i + \sum_{i=1}^{n} P_i y_i \\ &= E(X)+E(Y). \end{aligned}$$

2/ Soient $a, b \in \mathbb{R}$,

$$\begin{aligned} E(aX+b) &= \sum_{i=1}^{n} P_i\left(a x_i+b\right) \\ &= \sum_{i=1}^{n} P_i a x_i + \sum_{i=1}^{n} P_i b \\ &= a\sum_{i=1}^{n} P_i x_i + b\sum_{i=1}^{n} P_i \\ &= aE(X)+b \end{aligned}$$

□

Remarque 5.1.3 *De la proposition (5.1.1), on peut déduire la propriété suivante :*

$$E(aX+bY)=aE(X)+bY,\ \forall(a,b)\in\mathbb{R}^2.$$

Définition 5.1.3 ***La variance*** *de la variable aléatoire discrète X est le nombre positif $Var(X)$, donné par :*

$$Var(X)=\sum_{i=1}^{n} P_i\left[x_i-E(X)\right]^2$$

s'il existe.

Proposition 5.1.2 *Soient X une variable aléatoire discrète et a, b deux nombre réels.*

1/ $Var(X)=E\left(X^2\right)-\left[E(X)\right]^2.$

2/ $Var\left(aX+b\right)=a^2Var(X).$

Preuve. Soit la variable aléatoire discrète $X=\left(x_1,\ x_2,\ ...,\ x_n\right),\ n\in\mathbb{N}$. On a

1/

$$\begin{aligned} Var\,(X) &= \sum_{i=1}^{n} P_i\,[x_i - E(X)]^2 \\ &= \sum_{i=1}^{n} P_i\left[x_i^2 - 2E(X)x_i + (E(X))^2\right] \\ &= \sum_{i=1}^{n} P_i x_i^2 - 2E(X)\sum_{i=1}^{n} P_i x_i + (E(X))^2 \sum_{i=1}^{n} P_i \\ &= E\left(X^2\right) - 2\,(E(X))^2 + (E(X))^2 \\ &= E\left(X^2\right) - [E(X)]^2 . \end{aligned}$$

2/ En appliquant la première propriété, on obtient

$$\begin{aligned} Var\,(aX+b) &= E\left((aX+b)^2\right) - [E(aX+b)]^2 \\ &= E\left(a^2X^2 + 2abX + b^2\right) - [E(aX+b)]^2 \end{aligned}$$

de la proposition (5.1.1), on a :

$$\begin{aligned} Var\,(aX+b) &= a^2E(X^2) + 2abE\,(X) + b^2 - [aE(X)+b]^2 \\ &= a^2E(X^2) + 2abE\,(X) + b^2 - a^2\,(E(X))^2 - 2abE(X) - b^2 \\ &= a^2\left[E(X^2) - (E(X))^2\right] \\ &= a^2 Var\,(X)\,. \end{aligned}$$

$\square$

Définition 5.1.4 ***L'ecart type*** *$\delta\,(X)$ de la variable aléatoire discrète X est la racine carrée de la variance de X*

$$\delta\,(X) = \sqrt{Var(X)}.$$

Définition 5.1.5 ***La covariance*** *des deux variables aléatoires discrètes X et Y est :*

$$Cov\,(X,Y) = E\,(XY) - E(X)E(Y).$$

5.2 Quelques lois de probabilités discrètes

Dans cette section, on présente les définitions et les propriétés de quelques lois de probabilité discrète les plus usuelles dans les sciences expérimentales.

5.2.1 Loi de Bernoulli

Définition 5.2.1 *Une epreuve de Bernoulli est une expérience aléatoire dont le résultat peut être soit un succés noté par "1" de probabilité "p" ou un echec qu'on note par "0" de probabilité "q" mais pas les deux simultanément, tel que* $p + q = 1$. *La loi de Bernoulli est résumé dans le tableau ci-dessous :*

X	0	1
P_X	q	p

En plus, On a

1/ l'espérance mathématique de la loi de Bernoulli est :

$$E(X) = p.$$

2/ Sa variance est :

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= p - p^2 \\ &= p(1-p) \\ &= pq. \end{aligned}$$

3/ Finalement, son écart type est :

$$\delta(X) = \sqrt{pq}.$$

Exemple 5.2.1 *On lance une pièce de monnaie, la variable alétoire X est le nombre de faces obtenues. Alors X prend que deux valeurs* 0 *et* 1

X	0	1
P_X	$\frac{1}{2}$	$\frac{1}{2}$

et on a

1/ $E(X) = \frac{1}{2}$.

2/ $Var(X) = \frac{1}{4}$.

3/ $\delta(X) = \frac{1}{2}$.

5.2.2 Loi binomiale

Définition 5.2.2 *On effectue n $(n \in \mathbb{N}^* - \{1\})$ répétitions indépendantes d'une épreuve de Bernoulli dont la probabilité de succés est p. On définit la variable aléatoire X par le nombre de succés parmi les n résultas obenus. Alors X suit une loi binomiale de paramètres $n,\ p,\ q$ tels que $q = 1 - p$ est la probabilité d'echec. on note $X \sim B(n,\ p,\ q)$ où*

$$X = \{0, 1, 2, ..., n\}$$

et

$$\underset{0 \leq k \leq n}{P(X = k)} = C_n^k p^k q^{n-k}$$

avec

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

On peut vérifier aisément que

$$\sum_{k=0}^{n} P(X = k) = 1.$$

Proposition 5.2.1 *Soit X une variable aléatoire discrète suit la loi binomilale $B(n,\ p,\ q)$.*

1/ L'espérance mathématique de X est :

$$E(X) = np.$$

2/ La variance de X est :

$$Var(x) = npq.$$

3/ L'ecart type de X est :

$$\delta(X) = \sqrt{npq}$$

Preuve. Soit la variable aléatoire discrète $X \sim B(n,\ p,\ q)$,

1/ L'espérance mathématique de X

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} kP(X=k) \\
&= \sum_{k=0}^{n} kC_n^k p^k q^{n-k} \\
&= \sum_{k=0}^{n} k\frac{n!}{k!(n-k)!} p^k q^{n-k} \\
&= np\sum_{k=0}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} \\
&= np\left(p+q\right)^{n-1} \\
&= np.
\end{aligned}$$

2/ La variance de X

$$Var(x) = E(X^2) - \left(E(X)\right)^2.$$

Comme

$$\begin{aligned}
E\left(X^2\right) &= \sum_{k=0}^{n} k^2 P(X=k) \\
&= \sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k} \\
&= \sum_{k=0}^{n} k^2 \frac{n!}{k!(n-k)!} p^k q^{n-k} \\
&= \sum_{k=0}^{n} \frac{kn!}{(k-1)!(n-k)!} p^k q^{n-k} \\
&= \sum_{k=0}^{n} \frac{(k-1+1)\,n!}{(k-1)!(n-k)!} p^k q^{n-k} \\
&= \sum_{k=0}^{n} \frac{(k-1)\,n!}{(k-1)!(n-k)!} p^k q^{n-k} + \sum_{k=0}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} \\
&= n(n-1)p^2 \sum_{k=0}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} + np\sum_{k=0}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} \\
&= n(n-1)p^2 \left(p+q\right)^{n-2} + np\left(p+q\right)^{n-1} \\
&= n(n-1)p^2 + np
\end{aligned}$$

alors

$$\begin{aligned} Var(x) &= n(n-1)p^2 + np - (np)^2 \\ &= -np^2 + np \\ &= np\,(1-q) \\ &= npq. \end{aligned}$$

3/ Il est clair que

$$\delta\,(X) = \sqrt{Var(X)} = \sqrt{npq}.$$

$\square$

Exemple 5.2.2 *Voir l'exemple (5.1.1), $X \sim B\left(3, \frac{1}{2}, \frac{1}{2}\right)$ donc*

1/ $E(X) = np = \frac{3}{2}$.

2/ $Var(X) = npq = \frac{3}{4}$.

3/ $\delta\,(X) = \sqrt{npq} = \frac{\sqrt{3}}{2}$.

5.2.3 Loi géométrique

Définition 5.2.3 *On répète continuellement et de façon indépendante une épreuve de Bernoulli dont la probabilité de succés est p (non nulle). On définit la variable alétoire discrète X par le nombre d'épreuves nécéssaires pour obtenir un premier succés. Alors X prend les valeurs*

$$X = \{1, 2, 3, ...\}$$

et suit une loi géométrique de paramètre p. On note $X \sim géo\,(p)$ où

$$\underset{k\geq 1}{P(X=k)} = q^{k-1}p.$$

tel que $q = 1-p$ est la probabilité d'échec. Pour tout $n \in \mathbb{N}$, on a

$$\begin{aligned} \sum_{k=1}^{n} P(X=k) &= \sum_{k=1}^{n} q^{k-1}p \\ &= p\sum_{k=1}^{n} q^{k-1} \\ &= p\frac{1-q^n}{1-q}. \end{aligned}$$

Comme $0 < q < 1$, *alors quand* n *tend vers* $+\infty$, *on obteient*

$$\sum_{k=1}^{n} P(X = k) = \frac{p}{1-q} = 1.$$

Proposition 5.2.2 *Soit* X *une variable aléatoire discrète suit la loi géométrique* $géo\,(\,p)$.

1/ L'espérance mathématique de X est :

$$E(X) = \frac{1}{p}.$$

2/ La variance de X est :

$$Var(x) = \frac{q}{p^2}.$$

3/ L'ecart type de X est :

$$\delta\,(X) = \frac{\sqrt{q}}{p}.$$

Preuve. Soit la variable alétoire discrète $X \sim géo\,(\,p)$.

1/ L'espérance mathématique de X

$$\begin{aligned} E(X) &= \sum_{k=1}^{n} kP(X = k) \\ &= \sum_{k=1}^{n} kq^{k-1}p \\ &= p\sum_{k=1}^{n} \left(q^k\right)' \\ &= p\left(\sum_{k=1}^{n} q^k\right)' \\ &= p\left(\frac{1-q^n}{1-q}\right)' \end{aligned}$$

Comme $0 < q < 1$, alors quand $n \to +\infty$, on a

$$\begin{aligned} E(X) &= p\left(\frac{1}{1-q}\right)' \\ &= \frac{p}{(1-q)^2} \\ &= \frac{1}{p}. \end{aligned}$$

2/ La variance de X

$$Var(x) = E(X^2) - [E(X)]^2 \tag{5.2.1}$$

où

$$\begin{aligned}
E(X^2) &= \sum_{k=1}^{n} k^2 P(X=k) \\
&= p \sum_{k=1}^{n} k^2 q^{k-1} \\
&= pq \sum_{k=1}^{n} k\,(k-1+1)\,q^{k-2} \\
&= pq \left[\sum_{k=2}^{n} k\,(k-1)\,q^{k-2} + \sum_{k=2}^{n} k q^{k-2} \right] \\
&= pq \left[\sum_{k=1}^{n} \left(q^k\right)'' + \frac{1}{q} \sum_{k=1}^{n} \left(q^k\right)' \right] \\
&= pq \left[\left(\frac{1-q^n}{1-q} \right)'' + \frac{1}{q} \left(\frac{1-q^n}{1-q} \right)' \right]
\end{aligned}$$

Comme $0 < q < 1$, alors quand $n \to +\infty$, on a

$$\begin{aligned}
E(X^2) &= pq \left[\left(\frac{1}{1-q} \right)'' + \frac{1}{q} \left(\frac{1}{1-q} \right)' \right] \\
&= pq \left[\frac{2}{(1-q)^3} + \frac{1}{q} \frac{1}{(1-q)^2} \right] \\
&= \frac{2q}{p^2} + \frac{1}{p}.
\end{aligned}$$

En remplaçant dans (5.2.1), on trouve

$$\begin{aligned}
Var(x) &= \frac{2q+p-1}{p^2} \\
&= \frac{q}{p^2}.
\end{aligned}$$

3/ L'écart type de X

$$\delta\,(X) = \sqrt{Var(X)} = \frac{\sqrt{q}}{p}.$$

□

Exemple 5.2.3 *On lance un dé continuellement jusqu'a l'obtention d'un* 6. *Soit* X *le nombre de lancers nécéssaires.* X *suit la loi géométrique géo*(p) *où* $p = \frac{1}{6}$ *est la probabilité d'obtention* *d'un* 6. X *prend toutes les valeurs entières positives* $1, 2, 3, ...$*telles que pour tout* $k \in \mathbb{N}^*$, *on a*

$$P(X = k) = q^{k-1}p = \frac{1}{6}\left(\frac{5}{6}\right)^{k-1}.$$

En plus

1/ $E(X) = \frac{1}{p} = 6.$

2/ $Var(X) = \frac{q}{p^2} = 30.$

3/ $\delta(X) = \frac{\sqrt{q}}{p} = \sqrt{30}.$

5.2.4 Loi Hypergéométrique

Définition 5.2.4 *On tire sans remise* n *objets d'un ensemble de* N *objets dont* D *possèdent une caractéristique particulière et les autres* $(N - D)$ *ne la possèdent pas. Soit* X *le nombre d'objets de l'échantillon qui possèdent la caractéristique. Alors,* X *suit la loi hypergéométique de paramètre* $n,\ N,\ D$ *et on écrit* $X \sim H(N,\ D,\ n)$ *tel que*

$$\underset{k \in R_X}{P(X = K)} = \frac{C_D^k\ C_{N-D}^{n-k}}{C_N^n}$$

où

$$R_X = \{\max(0, n - N + D), ..., \min(n, D)\}.$$

Proposition 5.2.3 *Soit* X *une variable aléatoire discrète suit la loi hypergéométrique* $H(N,\ D,\ n)$.

1/ L'espérance mathématique de X est :

$$E(X) = n\frac{D}{N}.$$

2/ La variance de X est :

$$Var(x) = n\frac{D}{N}\left(1 - \frac{D}{N}\right)\left(\frac{N-n}{N-1}\right).$$

3/ L'ecart type de X est :

$$\delta(X) = \sqrt{n\frac{D}{N}\left(1 - \frac{D}{N}\right)\left(\frac{N-n}{N-1}\right)}.$$

Preuve. On distingue deux cas différents : i/ Si $\min(n, D) = n$, alors $R_X = \{0, ..., n\}$. ii/Si $\min(n, D) = D$, alors $R_X = \{0, ..., D\}$. On montre les propriétés 1/ et 2/ de la proposition (5.2.3) dans les deux cas de la même façon. On a

$$E(X) = \sum_{k=0}^{n} kP(X = k)$$

ce qui implique

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} k \frac{C_D^k \; C_{N-D}^{n-k}}{C_N^n} \\
&= \sum_{k=0}^{n} k \frac{\frac{D!}{k!(D-k)!} \frac{(N-D)!}{(n-k)!(N-D-n+k)!}}{\frac{N!}{n!(N-n)!}} \\
&= \sum_{k=0}^{n} \frac{D\,(D-1)!}{(k-1)!\,(D-k)!} \frac{(N-D)!}{(n-k)!\,(N-D-n+k)!} \frac{n\,(n-1)!\,(N-n)!}{N\,(N-1)\,...\,(N-n)!} \\
&= n\frac{D}{N}.
\end{aligned}$$

de manière similaire, on montre que

$$Var(x) = n\frac{D}{N}\left(1 - \frac{D}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Finalement

$$\delta\,(X) = \sqrt{n\frac{D}{N}\left(1 - \frac{D}{N}\right)\left(\frac{N-n}{N-1}\right)}.$$

□

Exemple 5.2.4 *Une boite contient 8 Composants parmi les quels 2 sont défectueux. Trois composants sont pris au hasard et sans remise de la boite. Soit X le nombre de composants défectueux dans l'échantillon. En effet $X = \{0,\ 1,\ 2\}$ définit une loi de probabilité hypergéométrique H de paramètres $N = 8,\ D = 2$ et $n = 3,$ tel que :*

$$P(X = 0) = \frac{C_2^0 C_6^3}{C_8^3} = \frac{20}{56},$$

$$P(X = 1) = \frac{C_2^1 C_6^2}{C_8^3} = \frac{30}{56},$$

$$P(X = 2) = \frac{C_2^2 C_6^1}{C_8^3} = \frac{6}{56}.$$

Il est clair que

$$\sum_{k=0}^{2} P(X = k) = 1.$$

5.2.5 Loi de Poisson

La loi de Poisson est une variable aléatoire discrète utilisée souvent dans l'étudie des phénomène rares dans certaines conditions. On cite par exemple, X : le nombre des personnes agés plus de 100 ans dans une population.

Définition 5.2.5 *Soit $\lambda \in \mathbb{R}^{*,+}$, On dit que la variable alétoire discrète X suit la loi de Poisson de paramètre λ , et on note $X \sim \boldsymbol{P}(\lambda)$, lorsque lensemble des valeurs prises par X est l'ensemble de tous les entiers positifs : $X(\Omega) = \mathbb{N}$. En plus, on a*

$$\forall k \in \mathbb{N} : P(X = k) = P_k = e^{-\lambda}\frac{\lambda^k}{k!},$$

vérifiant

$$\begin{aligned}\sum_{k\in\mathbb{N}} P_k &= \sum_{k=0}^{+\infty} e^{-\lambda}\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=0}^{+\infty}\frac{\lambda^k}{k!} \\ &= 1.\end{aligned}$$

Proposition 5.2.4 *Soit X une variable aléatoire discrète suit la loi de Poisson $\mathbf{P}(\lambda)$, $\lambda \in \mathbb{R}^{*,+}$.*

1/ L'espérance mathématique de X est :

$$E(X) = \lambda.$$

2/ La variance de X est :

$$Var(x) = \lambda.$$

3/ L'ecart type de X est :

$$\delta(X) = \sqrt{\boldsymbol{\lambda}}.$$

Preuve. Soient $\lambda \in \mathbb{R}^{*,+}$ et X une variable aléatoire discrète suit la loi de Poisson $\mathbf{P}(\lambda)$.

1/ L'espérance mathématique de X

$$\begin{aligned} E(X) &= \sum_{k=0}^{+\infty} kP_k \\ &= \sum_{k=0}^{+\infty} ke^{-\lambda}\frac{\lambda^k}{k!} \\ &= \lambda e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda. \end{aligned}$$

2/ La variance de X

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= \left[\sum_{k=0}^{+\infty} k^2 P_k\right] - \lambda^2 \\ &= \left[e^{-\lambda}\sum_{k=0}^{+\infty} k\frac{\lambda^k}{(k-1)!}\right] - \lambda^2 \\ &= e^{-\lambda}\left[\sum_{k=0}^{+\infty}(k-1)\frac{\lambda^k}{(k-1)!} + \sum_{k=0}^{+\infty}\frac{\lambda^k}{(k-1)!}\right] - \lambda^2 \\ &= e^{-\lambda}\left[\lambda^2\sum_{k=0}^{+\infty}\frac{\lambda^{k-2}}{(k-2)!} + \lambda\sum_{k=0}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!}\right] - \lambda^2 \\ &= \lambda. \end{aligned}$$

3/ L'écart type de X

$$\delta\,(X) = \sqrt{Var(X)} = \sqrt{\boldsymbol{\lambda}}.$$

$\square$

Exemple 5.2.5 *On considère la variable alétoire X : nombre de fautes de frappe par page du cours de Mathématiques. $X \sim \mathbf{P}\,(\lambda)$ telle que $\lambda = 0.1$. La probabilité d'avoir une erreur par page est :*

$$\begin{aligned} P_1 &= e^{-0.1}\frac{(0.1)}{1!} \\ &= 0,09. \end{aligned}$$

5.2.6 Approximation d'une loi binomiale par une loi du Poisson

Soit X une variable alétoire discrète suit la loi binomiale $X \sim B(n,\ p,\ q)$ avec n grand $(n \geq 30)$ et p petit $(p \leq 0.1)$. On peut approximer cette loi par la loi de Poisson de paramètre $\lambda = np$.

Exemple 5.2.6 *Dans une population une personne sur cent est un centenaire. On définit la variable alétoire discrète* X : *nombre des centenaires dans une population de 200 personnes.* X *suit la loi binomiale de paramètres* $n = 200,\ p = 0,01$ *et* $q = 1 - p = 0,99$. *Comme* $n > 30$ *et* $p < 0.1$, *alors on peut approximer la loi de* X *par la loi de Poisson* $\mathbf{P}(\lambda)$ *telle que* $\lambda = np = 2$.

5.3 Exercices

Exercice 1 : Un candidat se présente à un concours sous forme de QCM de 100 questions, à chaque question, sont proposés 4 réponses, dont une seule est correcte, l'examinateur fait le compte des réponses exactes données par le candidat.
Certains candidats répondent au hasard à chaque question. Soit la variable aléatoire X : «nombre de réponses exactes données par un candidat».

1/ Donner la loi de probabilité de X.

2/ Calculer son espérance et son écart type.

Exercice 2 : Dans une population une personne sur cent est un centenaire.

1/ Quelle est la probabilité de trouver au moins un centenaire parmi 100 personnes choisies au hasard.

2/ Même question parmi 200 personnes.

Exercice 3 : Un chercheur a analysé un lot complet de roches volcaniques. Il a compté le nombre de pierres précieuses dans chacune des roches de poids constant 1kg et a obtenu le tableau suivant :

$Nbr\ pp\ \ (xi)$	1	2	3	4	5	6	7	8
$Nbr\ rv\ \ (ni)$	63	19	17	32	5	2	1	1

1/ Pour une roche choisie au hasard parmi les 140 roches analysées, on désigne par X la variable aléatoire qui prend pour valeur le nombre de pierres précieuses trouvées. Calculer les probabilités $P(X = xi)$.

2/ Calculer $E(X)$ et $Var(X)$.

3/ En déduire la loi de probabilité de X et ses paramètres.

5.4 Corrigés

Exercice 1 : Comme X : «nombre de réponses exactes données par un candidat», alors

$$X = \{0,\ 1,\ 2,\ .....,100\}\,.$$

1/ La loi de la variable aléatoire X : il est clair que X suit la loi binomiale $\mathcal{B}(n,\ p,\ q)$ telle que

$$n = 100,\ p = \frac{1}{4},\ q = 1 - p = \frac{3}{4}.$$

2/ L'espérance et l'écart type de la variable X :

$$\begin{aligned} E(X) &= np \\ &= 100 \times \frac{1}{4} \\ &= 25. \end{aligned}$$

et

$$\begin{aligned} Var(X) &= npq \\ &= 100 \times \frac{1}{4} \times \frac{3}{4} \\ &= 18,75. \end{aligned}$$

Alors,

$$\begin{aligned} \sigma_X &= \sqrt{Var(X)} \\ &= \sqrt{18,75} \\ &= 4,33. \end{aligned}$$

Exercice 2 : Dans une population une personne sur cent est un centenaire. On définit la variable aléatoire discrète X : nombre des centenaires dans une population de 200 personnes. X suit la loi binomiale de paramètres $n = 100$, $p = 0.01$ et $q = 1 - p = 0.99$.

Comme $n > 30$ et $p < 0.1$, alors on peut approximer la loi de X par la loi de Poisson $\mathcal{P}(\lambda)$ telle que $\lambda = np = 1$ et

$$\underset{0 \leq k \leq n}{P(X = k)} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Probabilité d'avoir au moins un centenaire

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - e^{-1} \\ &= 0,63. \end{aligned}$$

De manière similaire si $n = 200$, on trouve $\lambda = 2$ et

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - e^{-2} \\ &= 0,86. \end{aligned}$$

Exercice 3 : On définit la variable aléatoire discrète X : nombre de pierres précieuses trouvées.

* x_i : nombre de pierres précieuses,

* n_i : nombre de roches volcaniques analysées.

La taille de l'echantillon $n = \sum_{i=1}^{8} n_i = 140$ et les probabilités sont calculées comme suit :

$$\begin{aligned} P(X = x_i) &= P_i \\ &= \frac{n_i}{n} \\ &= \frac{n_i}{140}. \end{aligned}$$

1/ Les probabilités $P(X = x_i)$

x_i	1	2	3	4	5	6	7	8	$Total$
$P(X = x_i)$	$\frac{63}{140}$	$\frac{19}{140}$	$\frac{17}{140}$	$\frac{32}{140}$	$\frac{5}{140}$	$\frac{2}{140}$	$\frac{1}{140}$	$\frac{1}{140}$	$\sum_{i=1}^{8} P(X = x_i) = 1$

2/

$$\begin{aligned} E(X) &= \overline{X} \\ &= \frac{1}{n}\sum_{i=1}^{8} n_i x_i \\ &= \frac{1}{140}(332) \\ &= 2,37. \end{aligned}$$

$$\begin{aligned}
\overline{X^2} &= \tfrac{1}{n}\sum_{i=1}^{8} n_i x_i^2 \\
&= \tfrac{1}{140}(1114) \\
&= 7,96.
\end{aligned}$$

$$\begin{aligned}
Var(X) &= \overline{X^2} - \left(\overline{X}\right)^2 \\
&= 7,96 - (2,37)^2 \\
&= 2,34.
\end{aligned}$$

3/ On remarque que $E\left(X\right) \simeq Var\left(X\right)$, alors X suit la loi de Poisson de paramètre $\lambda = E\left(X\right) = 2,4$.

Chapitre 6

Variables alétoires continues

6.1 Définitions et propriétés

Définition 6.1.1 *Soient* $(\Omega,\ \mathcal{F},P)$ *un espace probabilisé,* X *une variable aléatoire telle que*

$$\forall x\in\mathbb{R}:X^{-1}\left(\left]-\infty,\ x\right]\right)\in\mathcal{F}.$$

Il n'est pas possible de définir la loi de X *par la donnée des probabilités des évènemenets élémentaires quand* X *représente les valeurs d'un caractère quantitatif continu par exemple la taille, le poids...ect. Par contre, on peut déduire les probabilités que* X *prenne ses valeurs dans une partie de* $\mathbb{R}$ *à partie de* ***la fonction de répartition*** *définie par :*

$$F\left(x\right)=P\left(X\leq x\right)=\int\limits_{-\infty}^{x}f(t)dt,$$

où f *est* ***la fonction densité de probabilité*** *de la variable alétoire continue* X *vérifiant les propriétés suivantes :*

1/ La fonction densité f est une fonction positive et continue et

$$\int\limits_{-\infty}^{+\infty}f(x)dx=1.$$

2/ $\forall J=[a,b]\subset\mathbb{R},\ \int\limits_{a}^{b}f(x)dx=P\left(J\right).$

3/ La fonction de répartition F est une fonction continue et croissante telle que

i/ $\lim\limits_{x \to -\infty} F(x) = 0$ et $\lim\limits_{x \to +\infty} F(x) = 1.$

ii/ $\forall a,\ b \in \mathbb{R} :\ a < b \Rightarrow P\left(a \leq X \leq b\right) = F(b) - F(a).$

Remarque 6.1.1 *La fonction densité f est la dérivée de la fonction de répartition F de la variable alétoire continue X.*

Dans tout ce qui suit, on considère $(\Omega,\ \mathcal{F}, P)$ un espace probabilisé et X une variable aléatoire continue de Ω dans $\mathbb{R}$ de fonction densité f.

Définition 6.1.2 *L'**espérance mathématique** de X est :*

$$E(X) = \int\limits_{-\infty}^{+\infty} x f(x) dx.$$

Définition 6.1.3 ***La variance de** X est :*

$$\begin{aligned} Var(X) &= E\left[(X - E(X))^2\right] \\ &= E\left(X^2\right) - \left(E\left(X\right)\right)^2, \end{aligned}$$

où

$$E(X^2) = \int\limits_{-\infty}^{+\infty} x^2 f(x) dx.$$

Remarque 6.1.2 *Les propriétés de l'espérance mathématique et de la variance des propositions (5.1.1) et (5.1.2) d'une variable aléatoire discrète sont vérifiées en cas continu.*

6.2 Quelques variables alétoires continues

6.2.1 Loi uniforme

Soit X une variable aléatoire continue définie par sa fonction densité donnée par :

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

où $[a,b] \subset \mathbb{R}$ et vérifiant

$$\begin{aligned}\int_{-\infty}^{+\infty} f(x)dx &= \int_{-\infty}^{a} f(x)dx + \int_{a}^{b} f(x)dx + \int_{b}^{+\infty} f(x)dx \\ &= \int_{a}^{b} \frac{1}{b-a}dx \\ &= 1.\end{aligned}$$

X est appelée **"loi de probabilité uniforme"** notée par $X \sim U\left([a,b]\right)$. Sa fonction densité F est :

$$\begin{aligned}F(x) &= \int_{-\infty}^{x} f(t)dt \\ &= \int_{a}^{x} \frac{1}{b-a}dt \\ &= \frac{1}{b-a}\left(x-a\right).\end{aligned}$$

L'espérance mathématique de X est :

$$\begin{aligned}E\left(X\right) &= \int_{-\infty}^{+\infty} xf(x)dx \\ &= \int_{a}^{b} \frac{x}{b-a}dx \\ &= \left.\frac{x^2}{2\left(b-a\right)}\right]_a^b \\ &= \frac{b+a}{2}.\end{aligned}$$

La variance de X est :

$$Var\left(X\right) = E(X^2) - \left[E(X)\right]^2$$

telle que

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) dx \\ &= \int_a^b \frac{x^2}{b-a} dx \\ &= \left. \frac{x^3}{3(b-a)} \right]_a^b \\ &= \frac{b^2+ba+a^2}{3}. \end{aligned}$$

Par conséquant

$$Var(X) = \frac{(b-a)^2}{12}.$$

L'ecart type de X est

$$\delta(X) = \sqrt{VarX} = \frac{b-a}{2\sqrt{3}}.$$

6.2.2 Loi exponentielle

Soit λ un réel strictement positif. La fonction f définie par

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$

vérifiant

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx &= \int_{-\infty}^{0} f(x) dx + \int_{0}^{+\infty} f(x) dx \\ &= \int_0^{+\infty} \lambda \exp(-\lambda x)\, dx \\ &= -\exp(-\lambda x)]_0^{+\infty} \\ &= \lim_{x \to +\infty} [-\exp(-\lambda x)] + 1 \\ &= 1 \end{aligned}$$

est la fonction densité d'une loi de probabilité continue appelée **"loi exponentielle de paramètre λ"**. f est est continue et positive sur $[0, +\infty[$ et sa primitive F est la fonction de répartition de la loi de probabilité exponentielle. Alors, pour tout $x \geq 0$, on a

$$\begin{aligned}
F(x) &= P(X \leq x) \\
&= P(]-\infty, x]) \\
&= \int_{-\infty}^{x} f(x)dx \\
&= \int_{-\infty}^{x} \lambda \exp(-\lambda x)\, dx \\
&= 1 - \exp(-\lambda x)\,.
\end{aligned}$$

Généralement, les variables aléatoires décrivant une durée de vie suivent toutes une loi de probabilité exponentielle.

L'espérance mathématique de la loi exponentielle X est :

$$\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} x f(x)dx \\
&= \int_{0}^{+\infty} x\lambda \exp(-\lambda x)\, dx \\
&= -x \exp(-\lambda x) - \frac{1}{\lambda} \exp(-\lambda x)\Big]_{0}^{+\infty} \\
&= \frac{1}{\lambda}.
\end{aligned}$$

La variance de la loi exponentielle X est :

$$Var(X) = E(X^2) - [E(X)]^2$$

où

$$\begin{aligned}
E(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x)dx \\
&= \int_{0}^{+\infty} x^2 \lambda \exp(-\lambda x)\, dx \\
&= -x^2 \exp(-\lambda x)\Big]_0^{+\infty} + 2\int_{0}^{+\infty} x \exp(-\lambda x)\, dx \\
&= \frac{2}{\lambda^2}.
\end{aligned}$$

Par conséquant

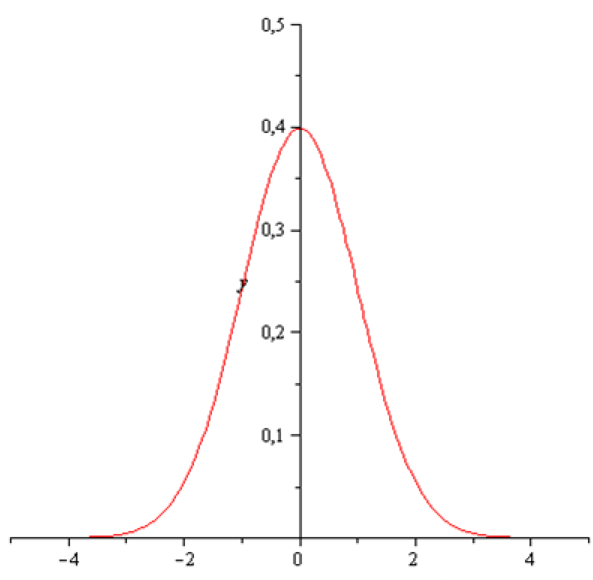
$$Var(X) = \frac{1}{\lambda^2}.$$

L'ecart type de la loi exponentielle X est :

$$\delta(X) = \sqrt{VarX} = \frac{1}{\lambda}.$$

6.2.3 Loi normale

C'est loi de probabilité continue la plus connue, elle est dite aussi loi de Laplace-Gauss et elle est caractérisée par sa courbe en cloche.



La courbe en cloche

Loi normale centrée réduite "standard"

La loi normale centrée réduite (ou standard) notée par $Z \sim N(0,\ 1)$ est définie à partir de sa fonction densité donnée par :

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right).$$

sa fonction de répartition est :

$$F(a) = \int_{-\infty}^{a} f(z)dz,\ \forall a \in \mathbb{R}.$$

Les probabilités d'évènements de la variable alétoire continue suivant la loi normale centrée réduite sont répèrtoriées dans la table suivante :

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

z	0.841	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
$\Phi(z)$	0.8000	0.9000	0.9500	0.9750	0.9800	0.9900	0.9950	0.9975	0.9990	0.9995

Table de la loi normale standard

Si $z = a.bc$ alors $F(z) = P(Z \leq z)$ est la valeur correspondante à l'initersection de la ligne $a..$ et la colonne $0.bc$. Par exemple pour trouver $F(1,62) = P(Z \leq 1,62)$, onlit la valeur qui correspond à l'intersection de la ligne $1,00$ et la colonne $0,62$ ce qui donne

$$P(Z \leq 1,62) = 0,9474.$$

Soient a et b deux réels positifs, la loi de probabilité normale centrée réduite vérifie les propriétés suivantes :

1/ $P(Z > a) = 1 - P(Z < a)$.

2/ $P(Z < -a) = P(Z > a)$.

3/ $P(a < Z < b) = F(b) - F(a)$.

L'espérance mathématique de loi de probabilité normale centrée réduite est :

$$E(x) = 0.$$

La variance de la loi de probabilité normale centrée réduite est :

$$Var(X) = 1.$$

L'ecart type de la loi de probabilité normale centrée réduite est :

$$\delta(X) = \sqrt{VarX} = 1.$$

Loi normale cas général

Soient m, $\delta \in \mathbb{R}$. La loi normale qu'on note par $X \sim N(m, \delta)$ est définie par sa fonction densité

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} \exp \frac{-(x-m)^2}{2\delta^2}$$

et sa fonction de répartition est :

$$F(a) = \int_{-\infty}^{a} f(x)dx, \ \forall a \in \mathbb{R}.$$

Son **espérance mathématique** est :

$$E(x) = m.$$

Sa **variance** est :

$$Var(X) = \delta^2 (X)$$

où $\delta (X)$ est son **ecart type**.

Passage de la loi normale à la la loi normale centrée réduite

Soit la variable de probabilité suivant la loi normale $X \sim N(m, \delta)$. en faisant le changement de variable $Z = \frac{X-m}{\delta}$, on obtient la nouvelle variable de probabilité $Z \sim N(0,\ 1)$ suivant la loi normale centrée réduite.

Exemple 6.2.1 *On veut étudier la capacité de mémoire d'adultes atteints d'une maladie neurologique. Pour cela, chaque individu lit* 30 *mots et doit ensuite les réciter. On définit la variable aléatoire* X *le nombre de mots retenus. Sachant que* $X \sim N(11,\ 2)$, *on cherche le quantile à* $97,5\%$ *pour la loi de* X, *cela revient à trouver* a *tel que*

$$\begin{aligned} P(X \leq a) &= \frac{97,5}{100} \\ &= 0,975. \end{aligned}$$

En posant $Z = \frac{X-m}{\delta}$, *on aura la nouvelle variable* $Z \sim N(0,\ 1)$. *Comme*

$$\begin{aligned} P(X \leq a) &= P\left(\frac{X-m}{\delta} \leq \frac{a-m}{\delta}\right) \\ &= F\left(\frac{a-11}{2}\right) \\ &= 0,975 \end{aligned}$$

alors, d'après la table de la loi normale centrée réduite

$$\frac{a-11}{2} = 1,96$$

ce qui implique

$$a = 14,92 \simeq 15 \ mots.$$

Approximation normale de la loi binomiale

Soit X une variable alétoire discrète qui suit la loi binomiale $X \sim B(n,\ p,\ q)$. Lorsque n est très grand, X comporte comme une variable normale de moyenne $m = np$ et de variance $Var(X) = npq$.

Remarque 6.2.1 *En pratique, on admet l'approximation normale de la loi binomiale dés que np et nq sont supérieurs ou égal à* 5.

Exemple 6.2.2 *Une population comporte en moyenne* 15 *personnes pesant moins de* 65 *kg sur* 40 *personnes. Soit X le nombre de personnes pesant plus de* 65 *kg dans un échantillon de taille* 500. *X est une variable alétoire discète prend toutes les valeurs de* 0 *à* 500 *et suit la loi binomiale $B(n,\ p,\ q)$ où $n = 500$, $p = \frac{25}{40} = 0,625$ et $q = \frac{15}{40} = 0,375$. Comme $np = 312,5 > 5$ et $nq = 187.5 > 5$ alors on peut approximer la loi de X par la loi normale de paramètres $m = np = 312,5$ et $\delta = \sqrt{npq} = 10,82$.*

6.3 Exercices

<u>Exercice 1 :</u> La température moyenne journalière en hiver sur les hauts plateaux est comprise entre $-4°$ et $16°$.

1/ Quelle est la probabilité d'avoir une moyenne de température inférieure à $12°$?

2/ Quelle est la probabilité d'avoir une moyenne de température supérieure à $10°$?

3/ Quelle est la probabilité d'avoir une moyenne de température inférieure à $10°$, sachant qu'elle est strictement positive.

4/ Quelle est la probabilité d'avoir une moyenne de température négative ?

<u>Exercice 2 :</u> La durée de vie d'un robot, exprimées en année, jusqu'à ce que survienne la première panne est une variable aléatoire qui suit une loi exponentielle de paramètre $\lambda > 0$. La probabilité pour qu'un robot survienne la première panne après 12 *ans* d'usage est $0,09$.

1/ Calculer λ.

2/ La probabilité qu'un robot tombe en panne pour la première fois est $0,3$, quelle est sa durée de vie ?

3/ On considère un lot de 10 robots fonctionnant de manière indépendante. Déterminer la probabilité que, dans ce lot, il y ait au moins un robot qui n'ait pas eu de panne au cours des deux premières années.

Exercice 3 : Une machine fabrique des biscuits destinées à être empilés dans des paquets de 16 pièces. L'épaisseur de chaque biscuit suit la loi normale de paramètres $m = 0,5cm$ et $\delta = 0,1$.

Soit X la variable aléatoire «épaisseur du paquet en cm». Calculez la probabilité pour que X soit compris entre $8,5cm$ et $9,5cm$.

Exercice 4 : Une population comporte en moyenne une personne mesurant plus de $1,85$ m sur 50 personnes. Soit X le nombre de personnes mesurant plus de $1,85$ m dans un échantillon de taille 200.

1/ Donner la loi de probabilité de X.

2/ Calculer son espérance mathématique et sa variance.

3/ Donner, en justifiant la réponse, une loi de probabilité continue permettant d'approcher la loi trouvée à la question 1.

4/ Calculer la probabilité qu'il y ait au moins une personne mesurant plus de $1.85m$.

Exercice 5 : Soit X une variable aléatoire qui suit une loi normale $\mathcal{N}$ (m, σ). Calculer ses paramètres m et δ sachant que $P(X < 0,32) = 0,5793$ et $P(X < 0,37) = 0,7580$.

6.4 Corrigés

Exercice 1 : La température moyenne journalière en hiver sur les hauts plateaux est comprise entre $-4°$ et $16°$.

On définit la variable aléatoire continue X : la moyenne de température en hiver. X suit la loi uniforme définie par sa fonction densité :

$$f(x) = \begin{cases} \frac{1}{20} & si \quad x \in [-4, 16] \\ 0 & sinon \end{cases}$$

1/ La probabilité d'avoir une moyenne de température inférieure à $12°$

$$\begin{aligned} P(X \leq 12) &= \int_{-4}^{12} \frac{1}{20} dt \\ &= 0,8. \end{aligned}$$

1. la probabilité d'avoir une moyenne de température supérieure à 10°

$$\begin{aligned} P(X>10) &= 1-P(X<10) \\ &= 1-\int_{-4}^{10}\tfrac{1}{20}dt \\ &= 1-\tfrac{14}{20} \\ &= 0,3. \end{aligned}$$

3/ La probabilité d'avoir une moyenne de température inférieure à 10°, sachant qu'elle est strictement positive

$$\begin{aligned} P(X<10/X>0) &= P(0\leq X\leq 10) \\ &= \int_{0}^{10}\tfrac{1}{20}dt \\ &= 0,5. \end{aligned}$$

4/ La probabilité d'avoir une moyenne de température négative

$$\begin{aligned} P(X\leq 0) &= P(X<0) \\ &= \int_{-4}^{0}\tfrac{1}{20}dt \\ &= 0,2. \end{aligned}$$

Exercice 2 : On définit la variable aléatoire continue X : la durée de vie d'un robot exprimées en année, jusqu'à ce que survienne la première panne.

X suit la loi exponentielle de paramètre $\lambda>0$, définie par sa fonction densité :

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & si \quad x\geqslant 0\\ 0 & si \quad x<0\end{cases}$$

1/ La probabilité pour qu'un robot survienne la première panne après 12 ans d'usage est $0,09$ c'est-à-dire :

$$P(X>12)=\int_{12}^{+\infty}\lambda e^{-\lambda x}dx=0,09.$$

Le paramètre λ

$$\begin{aligned} \int_{12}^{+\infty}\lambda e^{-\lambda x}dx = 0,09 \quad &\Leftrightarrow \quad e^{-12\lambda} = 0,09 \\ &\Leftrightarrow \quad -12\lambda = \ln(0,09) \\ &\Leftrightarrow \quad \lambda = \tfrac{-\ln(0,09)}{12} \\ &\Leftrightarrow \quad \lambda = 0,2. \end{aligned}$$

2/ La probabilité qu'un robot tombe en panne pour la première fois est $0,3$ c'est à dire :

$$P\left(X > x\right) = \int_0^x \lambda e^{-\lambda t} dt = 0,3.$$

Sa durée de vie x est

$$\begin{array}{rclcrcl} \int_0^x \lambda e^{-\lambda t} dt & = & 0,3 & \Longleftrightarrow & -e^{-0,2x}+1 & = & 0,3 \\ & & & \Longleftrightarrow & e^{-0,2x} & = & 0,7 \\ & & & \Longleftrightarrow & -0,2x & = & \ln\left(0,7\right) \\ & & & \Longleftrightarrow & x & = & \frac{-\ln(0,7)}{0,2} \\ & & & \Longleftrightarrow & x & = & 1,78\ ans. \end{array}$$

3/ On considère un lot de 10 robots fonctionnant de manière indépendante. On définit la variable aléatoire discrète Y : nombre de robots ayant une durée de vie superieur à 2. Alors, Y suit la loi binomiale de paramètres $n = 10$,

$$\begin{array}{rcl} p & = & P\left(X > 2\right) \\ & = & \int_2^{+\infty} \lambda e^{-\lambda t} dt = e^{-0,4} \end{array}$$

et

$$\begin{array}{rcl} q & = & 1 - p \\ & = & 1 - e^{-0,4}. \end{array}$$

Les probabilités sont calculèes par la formule suivante :

$$\underset{0 \leq k \leq 10}{P(Y = k)} = C_{10}^k p^k q^{10-k}.$$

La probabilité que dans ce lot, il y ait au moins un robot qui n'ait pas eu de panne au cours des deux premières années est :

$$\begin{array}{rcl} P\left(Y \geq 1\right) & = & 1 - P\left(Y < 1\right) \\ & = & 1 - P(Y = 0). \end{array}$$

Exercice 3 : On définit les variables aléatoires continues Y : épaisseur du biscuit et X : épaisseur d'un paquet du 16 biscuits en cm. Y suit la loi normale de paramètres

$$m = E\left(Y\right) = 0,5cm$$

et

$$\sigma_Y = \sqrt{Var\,(Y)} = 0,1.$$

On remarque que $X = 16Y$, alors

$$\begin{cases} E\,(X) & = & 16E\,(Y) & = & 16 \times 0,5 & = & 8cm \\ \sigma_X & = & 16\sigma_Y & = & 16 \times 0,1 & = & 1,6cm \end{cases}$$

Soit la variable centrée réduite Z définie par :

$$Z = \frac{X - E\,(X)}{\sigma_X} = \frac{X - 8}{1,6}$$

suivant la loi normal centrée réduite $(E\,(Z) = 0$ et $\sigma_Z = 1)\,.$

La probabilité pour que X soit compris entre $8,5cm$ et $9,5cm$ est :

$$\begin{array}{rcl} P\,(8,5 \leq X \leq 9,5) & = & P\left(\frac{8,5-8}{1,6} \leq Z \leq \frac{9,5-8}{1,6}\right) \\ & = & P\,(0,31 \leq Z \leq 0,93) \\ & = & P\,(0,93) - P\,(0,31)\,. \end{array}$$

D'après la table de la loi normal centré réduite on trouve que $P\,(0,93) = 0,8238$ et $P\,(0,31) = 0,6217$. Doù,

$$\begin{array}{rcl} P\,(8,5 \leq X \leq 9,5) & = & 0,8238 - 0,6217 \\ & = & 0,2021. \end{array}$$

<u>**Exercice 4 :**</u> On définit la variable aléatoire X : le nombre de personnes mesurant plus de $1m85$ dans un échantillon de taille 200.

1/ X suit la loi binomiale de paramètres $n = 200$, $p = \frac{1}{50}$ et $q = 1 - p = \frac{49}{50}$.

2/ L'espérance mathématique de X

$$\begin{array}{rcl} E(X) & = & np \\ & = & 200 \times \dfrac{1}{50} \\ & = & 4. \end{array}$$

Et

$$\begin{array}{rcl} Var(X) & = & npq \\ & = & 200 \times \dfrac{1}{50} \times \dfrac{49}{50} \\ & = & 3,92. \end{array}$$

3/ Comme $n = 200 >> 30$, alors on peut approximer la loi de X par la loi normal $\mathcal{N}\ (m,\ \sigma)$ de paramètres $m = 4$ et $\sigma = \sqrt{Var(X)} = 1,98$.

4/ La probabilité qu'il y ait au moins une personne mesurant plus de $1,85m$: $P(X \geq 1)$

$$P(X \geq 1) \quad = \quad 1 - P(X < 1)$$

En faisant le changement de variable

$$\begin{aligned} Z &= \frac{X-m}{\sigma} \\ &= \frac{X-4}{1,98} \end{aligned}$$

on obient

$$\begin{aligned} P(X \geq 1) &= 1 - P(Z \leq -1,52) \\ &= 1 - [1 - P(Z \leq 1,52)] \\ &= P(Z \leq 1,52). \end{aligned}$$

Comme la nouvelle variable Z suit la loi normale centrée réduite $\mathcal{N}\ (0,\ 1)$, alors de la table correspondante, on trouve

$$P(X \geq 1) = 0,9357.$$

<u>**Exercice 5 :**</u> Soit X une variable aléatoire qui suit une loi normale $\mathcal{N}\ (m,\ \sigma)$ de paramètres m et σ. La nouvelle variable

$$Z = \frac{X-m}{\sigma}$$

suit la loi normale centrée réduite $\mathcal{N}\ (0,\ 1)$.

On a

$$\left\{ \begin{array}{lcl} P\left(X < 0,32\right) &=& 0,5793 \\ P\left(X < 0,37\right) &=& 0,7580 \end{array} \right. \Rightarrow \left\{ \begin{array}{lcl} P\left(\frac{X-m}{\sigma} < \frac{0,32-m}{\sigma}\right) &=& 0,5793 \\ P\left(\frac{X-m}{\sigma} < \frac{0,37-m}{\sigma}\right) &=& 0,7580 \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{lcl} P\left(Z < \frac{0,32-m}{\sigma}\right) &=& 0,5793 \\ P\left(Z < \frac{0,37-m}{\sigma}\right) &=& 0,7580 \end{array} \right.$$

De la table de la loi normale centré réduite, on trouve :

$$\left\{ \begin{array}{lcl} \frac{0,32-m}{\sigma} &=& 0,2 \\ \frac{0,37-m}{\sigma} &=& 0,7 \end{array} \right. \Rightarrow \left\{ \begin{array}{lcl} 0,2\sigma + m &=& 0,32 \\ 0,7\sigma + m &=& 0,37 \end{array} \right.$$

Après la résolution de ce dernier système d'équations, on obtient

$$m \quad = \quad 0,3 \ \text{ et } \ \sigma \quad = \quad 0,1.$$

Bibliographie

[1] B. Courtebras, Mathématiser le hasard, Vuibert, 2008.

[2] J. Jacod, P. E. Protter, Probability Essentials, Springer, 2003.

[3] S. Méléard, Aléatoire, Introduction à la théorie et au calcul des probabilités, Éditions de l'École Polytechnique, 2010.

[4] T. Phan, J.P. Rowenczyk, Exercices et problèmes de statistique et probabilités, $2^{ème}$édition. Dunod, Paris 2012.

[5] G. Saporta, Probabilités, Analyse des données et statistiques, Paris, Éditions Technip, 2006.

[6] J. Tricot, Les topiques, vol. V, t. 1 à 8, VRIN, 1990.

[7] R.Veysseyre, Statistique et probabilités pour l'ingénieur, $2^{ème}$ édition. Dunod, 2006.