

UNIVERSITÉ PARIS-EST-CRÉTEIL
ESIAG

Statistiques descriptives et probabilités

Spécialité : L1

BÉATRICE DE TILIÈRE

TABLE DES MATIÈRES

I Statistique Descriptive	3
1 LES DONNÉES STATISTIQUES	5
1.1 Vocabulaire de base	5
1.2 Les types de variables	6
1.2.1 Variables qualitatives	6
1.2.2 Variables quantitatives	6
2 TABLEAUX ET REPRÉSENTATIONS GRAPHIQUES	9
2.1 Variables qualitatives	9
2.1.1 Effectifs et fréquences	9
2.1.2 Diagrammes à bandes	11
2.1.3 Diagrammes circulaires	12
2.2 Variables quantitatives discrètes	13
2.2.1 Effectif et fréquences	13
2.2.2 Diagramme en bâtons	14
2.2.3 Fonction de répartition empirique	14
2.3 Variables quantitatives continues	15
2.3.1 Regroupement en classes	15
2.3.2 Histogramme d'une variable continue	16
2.3.3 Fonction de répartition empirique	18
3 STATISTIQUE DESCRIPTIVE UNIVARIÉE	21
3.1 Paramètres de position	22
3.1.1 Le mode	22
3.1.2 La moyenne (arithmétique)	22
3.1.3 Médiane, quartiles, quantiles	24
3.1.4 Réflexions sur le mode, la moyenne, et la médiane	26
3.1.5 Moyennes géométrique, harmonique et quadratique	26
3.2 Paramètres de dispersion	27
3.2.1 Étendue et distance interquartile	27
3.2.2 Variance et écart-type	28
3.3 Box plot ou boîte à moustaches	30
3.4 Paramètres de forme	31
3.4.1 Variable centrée réduite	31
3.4.2 Moments d'une distribution	31
3.4.3 Coefficient d'asymétrie de Fisher	32
3.4.4 Coefficient d'aplatissement	33
3.5 Courbe de concentration de Lorenz et indice de Gini	34

3.5.1	Courbe de concentration de Lorenz	34
3.5.2	Indice de Gini	36
4	STATISTIQUE DESCRIPTIVE BI-VARIÉE	39
4.1	Distributions marginales et conditionnelles	39
4.2	Indépendance	42
4.2.1	Définition	42
4.2.2	Écart à l'indépendance	43
4.3	Nuage de points, résumés numériques	43
4.3.1	Nuage de points	43
4.3.2	Moyenne, variance, écart-type	44
4.3.3	Covariance	45
4.3.4	Coefficient de corrélation	46
4.3.5	Ajustement affine	48
4.3.6	Un exemple	50
II	Probabilités	53
5	MODÉLISATION DES PHÉNOMÈNES ALÉATOIRES	55
5.1	Introduction	55
5.2	L'espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$	56
5.2.1	Espace des états	56
5.2.2	Événements	57
5.2.3	Probabilité	58
6	CONSTRUCTION D'ESPACES PROBABILISÉS	61
6.1	Caractérisation d'une probabilité	61
6.2	Cas où l'univers est fini	61
6.2.1	Dénombrément, modèle d'urne	62
6.3	Espace des états infini dénombrable	66
7	CONDITIONNEMENT ET INDÉPENDANCE	67
7.1	Probabilité conditionnelle	67
7.1.1	Définition	67
7.1.2	Formule des probabilités totales et formule de Bayes	69
7.2	Indépendance des événements	71
8	VARIABLES ALÉATOIRES DISCRÈTES	73
8.1	Définitions et exemples	73
8.2	Fonction de répartition	76
8.2.1	Espérance	77
8.2.2	Variance, moments d'ordres supérieurs	78
8.2.3	Inégalité de Markov et de Bienaymé Tchebychev	80

Première partie

Statistique Descriptive

CHAPITRE 1

LES DONNÉES STATISTIQUES

La statistique descriptive permet de décrire et d'analyser un ensemble de données. Pour cela, on va tracer des graphiques, calculer des moyennes et d'autres indices, rechercher les scores extrêmes ou les distributions peu courantes.

Il ne faut pas confondre la statistique descriptive avec la *statistique mathématique* (ou inférentielle) qui a pour but de modéliser des comportements à partir d'observations.

La statistique descriptive procède d'une *étude statistique* qui est la succession des actions suivantes :

1. Définition de la population, des individus et des critères d'observation ;
2. Collecte et dépouillement des informations ;
3. Représentation(s) graphique(s) ;
4. Étude mathématique , *i.e.*, analyse quantifiée.

Remarque.

1. La statistique descriptive constitue les actions 3 et 4.
2. L'ensemble des informations dépouillées constitue la *série statistique*.
3. Si l'étude statistique porte sur un seul critère, on dit que la série statistique est *simple* (ou *univariée*). Si l'étude porte sur deux ou plusieurs critères, la série est dite respectivement *double* (ou *bivariée*) ou *multiple*.

1.1 VOCABULAIRE DE BASE

Définition 1.1. Un *individu* est un élément sur lequel porte l'étude statistique. La *population*, notée \mathcal{P} , est l'ensemble des individus. La *taille* de la population, notée N , est le nombre d'individus :

$$\text{card}(\mathcal{P}) = N.$$

Chaque individu de la population est observé à travers des caractéristiques, caractères ou indicateurs appelés *variables* (on les notera souvent X , Y , etc...).

Remarque. Par hypothèse, la population est un ensemble (fini) non vide, donc $N \in \mathbb{N}^*$. De plus, il s'avère souvent pratique, voire incontournable (anonymat, etc.), de désigner les individus par des nombres. On parle alors de l'individu u et $\mathcal{P} = \{1, \dots, N\}$. On désignera alors par $X(u)$ la valeur prise par la variable pour l'individu u .

Exemple 1.1.

- Si on considère les notes d'un groupe de TD à l'examen final, la population sera le groupe de TD, un individu sera un étudiant du groupe, et la variable sera la note à l'examen final.
- Si on veut étudier les accidents d'avions, la population sera l'ensemble des accidents d'avion, un individu sera un accident donné, et un exemple de variable pourrait être le modèle de l'avion.

Lorsque la population est trop importante, on peut être amené à faire porter l'étude statistique seulement sur une partie de la population, que l'on appelle *échantillon*. Ce cours ne traite pas des relations entre un échantillon et la population dont il est issu et *sauf mention explicite du contraire*, on considère que l'étude statistique porte sur la population complète.

1.2 LES TYPES DE VARIABLES

1.2.1 VARIABLES QUALITATIVES

Définition 1.2. Une variable est appelée *qualitative* lorsque les réponses possibles à la question posée, ou les valeurs prises par la variable, ne correspondent pas à une quantité mesurable par un nombre mais appartiennent à un groupe de catégories. On les appelle *modalités* de la variable.

Exemple 1.2. Le sexe, la couleur des yeux, la mention au baccalauréat, la fréquence d'une activité (jamais, rarement, parfois, souvent, très souvent).

On distingue :

- les variables qualitatives *nominales* : il n'y a pas de hiérarchie entre les différentes modalités ; exemple : sexe, couleur des yeux.
- les variables qualitatives *ordinaires* : les différentes modalités peuvent être ordonnées de manière naturelle ; exemple : la mention au baccalauréat, la fréquence d'une activité.

Remarque. Certaines variables nominales peuvent être désignées par un code numérique, qui n'a pas de valeur de quantité. Exemple : le code postal, le sexe (1=garçon, 2=fille).

1.2.2 VARIABLES QUANTITATIVES

Définition 1.3. Une variable est appelée *quantitative* lorsque les réponses possibles à la question posée, ou les valeurs prises par la variable, correspondent à des quantités mesurables et sont données sous forme de nombre.

On distingue :

- les variables quantitatives *discrettes* : elles prennent leurs valeurs dans un ensemble discret, le plus souvent fini ; exemple : le nombre d'enfants, la pointure du pied.
- les variables quantitatives *continues* : elles peuvent prendre toutes les valeurs d'un intervalle réel ; exemple : la taille des individus, le poids d'un individu.

Remarque. L'âge peut être vu et traité comme une variable quantitative discrète ou continue suivant la précision que l'on choisit et le nombre de valeurs qu'il prend au sein de la population. Il peut également exister des variables basées sur l'âge qui sont qualitatives. Si dans un sondage on pose la question "quelle est votre tranche d'âge parmi les possibilités suivantes : - de 25 ans, entre 25 et 40, entre 40 et 60 et + de 60 ans", on peut voir la variable "tranche d'âge" comme une variable qualitative ordinale.

CHAPITRE 1. LES DONNÉES STATISTIQUES

CHAPITRE 2

TABLEAUX ET REPRÉSENTATIONS GRAPHIQUES

Lorsque l'on est face à une série statistique dont les données sont écrites sous *forme brute*, on va dans un premier temps chercher à regrouper les informations dans un tableau faisant apparaître les *effectifs*, les *fréquences* ou *fréquences cumulées croissantes*. A l'aide de ces tableaux, on pourra tracer des graphiques dont la nature dépendra du type de variables.

2.1 VARIABLES QUALITATIVES

2.1.1 EFFECTIFS ET FRÉQUENCES

On considère une population de taille N pour laquelle on étudie une certaine variable X , dont les modalités sont m_1, \dots, m_p .

Définition 2.1. Soit $i \in \{1, \dots, p\}$.

1. Le nombre n_i d'individus pour lesquels la valeur prise par la variable est m_i est appelé *l'effectif* ou la *fréquence absolue* associé à la modalité m_i .
2. Le nombre $f_i := \frac{n_i}{N}$ est la *fréquence* ou *fréquence relative* associée à la modalité m_i .

On a alors,

Propriété 2.1.

$$\sum_{i=1}^p n_i = N, \quad \sum_{i=1}^p f_i = 1 \quad \text{et} \quad 0 \leq f_i \leq 1 \quad \forall i \in \{1 \dots p\}.$$

Démonstration. Les modalités possibles étant m_1, \dots, m_p , la première somme est évidente. On utilise ensuite la définition de la fréquence pour obtenir

$$\sum_{i=1}^p f_i = \sum_{i=1}^p \frac{n_i}{N} = \frac{1}{N} \sum_{i=1}^p n_i = \frac{N}{N} = 1.$$

L'encadrement de f_i vient du fait que $0 \leq n_i \leq N$. □

Définition 2.2. On se place dans le cas d'une variable qualitative *ordinale* où les modalités m_1, \dots, m_p sont ordonnées suivant l'ordre croissant naturel (ou hiérarchique ascendant) qui règne parmi ces modalités. Soit $i \in \{1, \dots, k\}$.

1. L'*effectif cumulé croissant* (respectivement *effectif cumulé décroissant*) jusqu'à la modalité m_i , noté ν_i (respectivement $\tilde{\nu}_i$), est la somme des effectifs des modalités m_1, \dots, m_i (respectivement m_{i+1}, \dots, m_p) :

$$\nu_i = \sum_{j=1}^i n_j \quad \text{et} \quad \tilde{\nu}_i = \sum_{j=i}^p n_j.$$

2. La *fréquence cumulée croissante* (respectivement *fréquence cumulée décroissante*) jusqu'à la modalité m_i , notée ϕ_i (respectivement $\tilde{\phi}_i$), est la somme des fréquences des modalités m_1, \dots, m_i (respectivement m_{i+1}, \dots, m_p) :

$$\phi_i = \sum_{j=1}^i f_j \quad \text{et} \quad \tilde{\phi}_i = \sum_{j=i}^p f_j.$$

A l'aide de ces quantités, on peut alors construire un tableau de distribution qui permet de résumer les données.

Variable X	m_1	m_2	\dots	m_p	Total
Fréquence absolue ou effectif	n_1	n_2	\dots	n_p	N
Fréquence relative	$f_1 = n_1/N$	$f_2 = n_2/N$	\dots	$f_p = n_p/N$	1
Fréquence cumulée croissante	$\phi_1 = f_1$	$\phi_2 = f_1 + f_2$	\dots	$\phi_p = 1$	X
Fréquence cumulée décroissante	$\tilde{\phi}_1 = 1$	$\tilde{\phi}_2 = f_2 + \dots + f_p$	\dots	$\tilde{\phi}_p = f_p$	X

Les deux dernières lignes de ce tableau n'ont de sens que si la variable est qualitative **ordinale**.

Exemple 2.1. On étudie la répartition des 9379079 français exerçant une pratique sportive en 2013. Les données sont les suivantes :

Discipline	Aviron	Basket-ball	Cyclisme	Football	Tennis	Autre
Nombre de licenciés	103084	536891	119247	2002398	1111316	5506143

La population est ici l'ensemble des personnes exerçant une activité sportive dans un club en France en 2013. La variable X est la discipline du sport choisi. Il s'agit bien d'une variable qualitative dont les modalités sont ici : aviron, basket-ball, cyclisme, football, tennis et autre.

Pour déterminer la fréquence d'une modalité, il suffit de diviser l'effectif de cette dernière par la taille de la population qui est ici de 9379079. On obtient ici pour l'aviron par exemple une fréquence égal à $\frac{103084}{9379079} = 0.0109908$. On arrondit ensuite le résultat à quatre chiffres après la virgule afin d'avoir une précision de deux chiffres après la virgule si on exprime le résultat en pourcentage. On obtient le tableau suivant :

X	Aviron	Basket-ball	Cyclisme	Football	Tennis	Autre	Total
Effectif	103084	536891	119247	2002398	1111316	5506143	9379079
Fréquence	0.0110	0.0572	0.0127	0.2135	0.1185	0.5871	1

La variable n'étant pas ordinaire, il n'y aurait pas de sens ici à considérer les effectifs cumulés croissants.

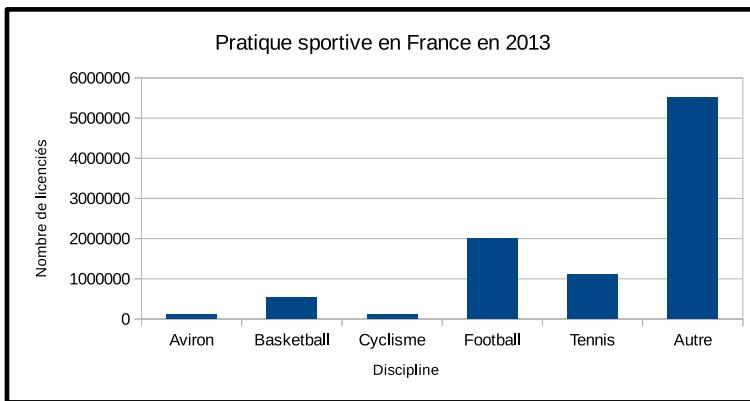
Donnons maintenant les deux représentations graphiques les plus courantes pour les variables qualitatives.

2.1.2 DIAGRAMMES À BANDES

Pour obtenir un diagramme à bandes, on trace un repère formé d'un axe horizontal non gradué et d'un axe vertical gradué.

- Sur l'axe horizontal, on trace des rectangles de même largeur, représentant les modalités, que l'on place à des distances régulières les uns des autres
- Les hauteurs des rectangles sont proportionnelles aux effectifs ou aux fréquences des modalités.

En reprenant les données de l'exemple 2.1, on obtient le diagramme suivant :



Remarque.

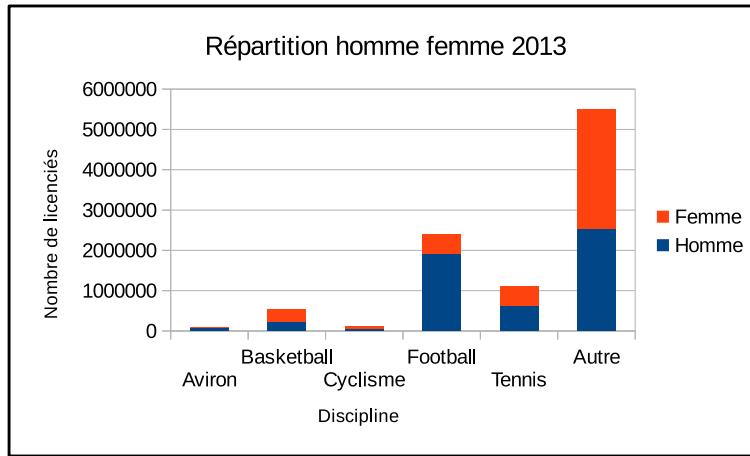
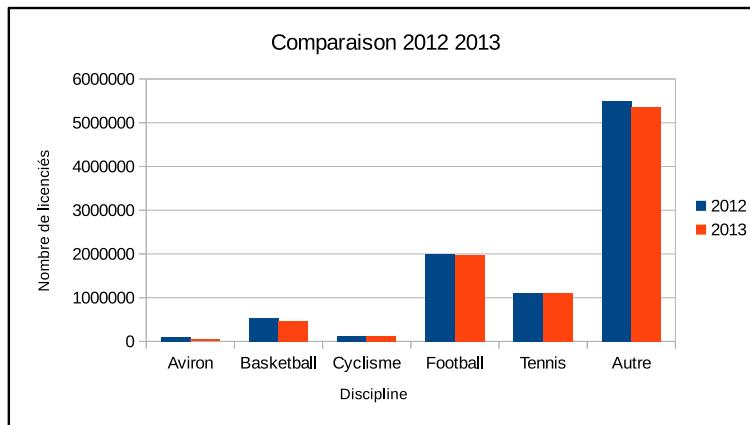
1. Étant donné que les modalités sont qualitatives, il n'y a pas d'ordre entre elles.
2. Ce type de graphique permet aussi de comparer les résultats sur différentes années ou différentes populations.

On possède également les données de la répartition des pratiques sportives en France en 2012 :

Année	Aviron	Basket-ball	Cyclisme	Football	Tennis	Autre
2013	103084	536891	119247	2002398	1111316	5506143
2012	43788	475465	115891	1973260	1103519	5369765

On peut aussi donner une représentation de sous-populations. Toujours avec le même exemple, on considère que la répartition entre hommes et femmes est la suivante :

Sexe	Aviron	Basket-ball	Cyclisme	Football	Tennis	Autre
Homme	80810	230532	65312	1912023	629311	2538924
Femme	22274	306359	53935	90375	482005	2967219



2.1.3 DIAGRAMMES CIRCULAIRES

Dans le cas d'un diagramme *circulaire* (ou *camembert*), les modalités sont représentées par un secteur angulaire d'un disque (ou d'un demi-disque), dont l'angle est proportionnel à l'effectif ou la fréquence. On effectue donc un produit en croix pour connaître l'angle de chaque secteur. Dans le cas d'un disque complet, on a ainsi les correspondances suivantes :

$$100\% \longleftrightarrow 360^\circ$$

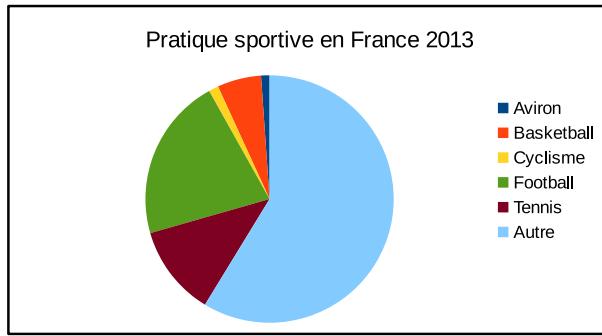
$$x\% \longleftrightarrow \frac{x \times 360}{100} = x \times 3.6^\circ,$$

et dans le cas d'un demi-disque, on a :

$$100\% \longleftrightarrow 180^\circ$$

$$x\% \longleftrightarrow \frac{x \times 180}{100} = x \times 1.8^\circ.$$

Avec l'exemple des pratiques sportives, on obtient :



2.2 VARIABLES QUANTITATIVES DISCRÈTES

2.2.1 EFFECTIF ET FRÉQUENCES

On considère une population de taille N sur laquelle on étudie une certaine variable X prenant un nombre fini de valeurs $x_1 < \dots < x_p$, que l'on suppose classées en ordre croissant.

De façon analogue à ce qui a été fait dans la section 2.1, pour $1 \leq i \leq p$, on note n_i l'*effectif* des individus pour lesquels la variable prend la valeur x_i . On note f_i la *fréquence relative* pour la valeur x_i , $\phi_i = f_1 + \dots + f_i$ la i -ème *fréquence cumulée croissante* et $\tilde{\phi}_i = f_i + \dots + f_p$ la i -ème *fréquence cumulée décroissante*. On résume habituellement les données comme dans le tableau-type suivant :

Valeurs prises par X	x_1	x_2	\dots	x_p	Total
Fréquence absolue ou effectif	n_1	n_2	\dots	n_p	N
Fréquence relative	$f_1 = n_1/N$	$f_2 = n_2/N$	\dots	$f_p = n_p/N$	1
Fréquence cumulée croissante	$\phi_1 = f_1$	$\phi_2 = f_1 + f_2$	\dots	$\phi_p = 1$	X
Fréquence cumulée décroissante	$\tilde{\phi}_1 = 1$	$\tilde{\phi}_2 = f_2 + \dots + f_p$	\dots	$\tilde{\phi}_p = f_p$	X

Exemple 2.2. Un responsable du service marketing d'une grande marque d'ordinateurs (portables et de bureau) fait une enquête dans une université pour savoir combien d'ordinateurs (portables ou de bureau) il y a dans le foyer de chaque étudiant.

Nombre d'ordinateur	0	1	2	3	4	5	Total
Nombre d'étudiants	300	876	1235	984	225	154	3774

La population est ici l'ensemble des étudiants de l'université, la variable est le nombre d'ordinateurs (il s'agit d'une variable quantitative discrète). On obtient le tableau suivant :

Nombre d'ordinateur	0	1	2	3	4	5	Total
Nombre d'étudiants	300	876	1235	984	225	154	3774
Fréquence	0,0795	0,2321	0,3272	0,2607	0,0596	0,0408	1
Fréquence cumulée croissante	0,0795	0,3116	0,6388	0,8996	0,9592	1	X
Fréquence cumulée décroissante	1	0,9205	0,6884	0,3612	0,1004	0,0408	X

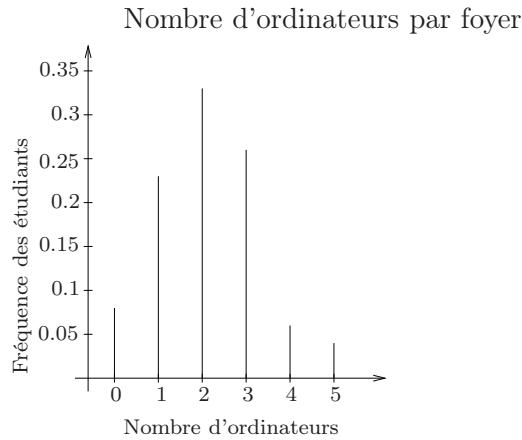
Nous donnons maintenant les représentations graphiques les plus courantes.

2.2.2 DIAGRAMME EN BÂTONS

Pour obtenir un diagramme en bâtons, on trace un repère formé de deux axes gradués orthogonaux.

- Sur l'axe des abscisses, on place les valeurs x_1, \dots, x_p prises par la variable.
- Sur l'axe des ordonnées, on place les effectifs ou les fréquences correspondant aux différentes valeurs.

Avec les données de l'exemple 2.2, on obtient :



Remarque. Le logiciel Excel ne permet pas vraiment de faire des diagrammes en bâtons. Il faut donc faire un histogramme et réduire au maximum la largeur des tubes.

Définition 2.3. La courbe joignant les sommets des bâtons est appelée *polygone des fréquences absolues*, si l'on a représenté les effectifs, ou *relatives*, si l'on a représenté les fréquences.

Avec les données de l'exemple 2.2, on obtient :

2.2.3 FONCTION DE RÉPARTITION EMPIRIQUE

La fonction de répartition empirique permet de décrire la série statistique de manière complète. Elle est définie de la manière suivante.

Définition 2.4. La *fonction de répartition empirique* de X est l'application $F_X : \mathbb{R} \rightarrow [0, 1]$ construite comme suit :

- $\forall x \in]-\infty, x_1[, F_X(x) = 0$;
- $\forall i \in \{1, \dots, p - 1\}, \forall x \in [x_i, x_{i+1}[, F_X(x) = \phi_i$;
- $\forall x \in [x_p, +\infty[, F_X(x) = 1$.

Remarque. La fonction F_X est une fonction càdlàg (continue à droite avec limite à gauche).

La fonction de répartition empirique de l'exemple 2.2 est donnée ci-dessous.

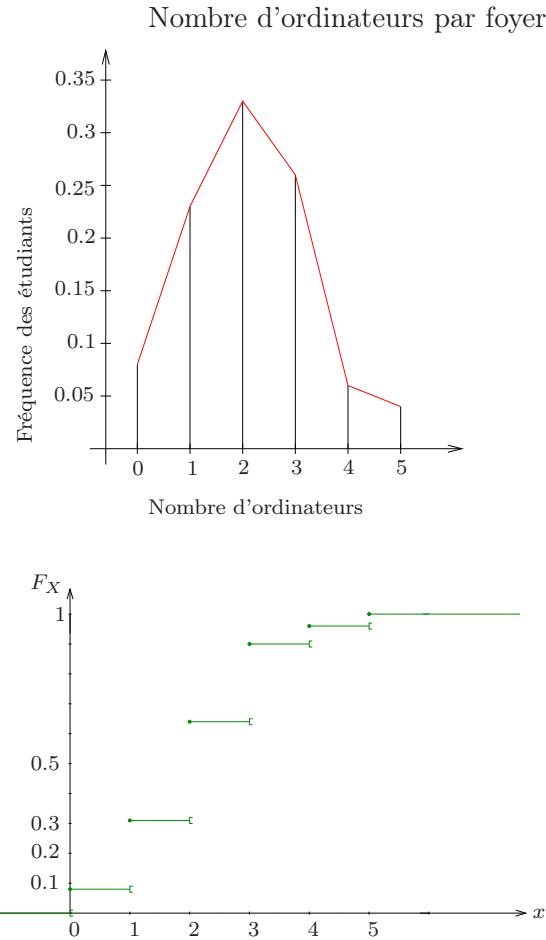


FIGURE 2.1 – Fonction de répartition empirique de l'exemple 2.2.

2.3 VARIABLES QUANTITATIVES CONTINUES

On considère une population de taille N . Les données brutes de la variable pour les individus sont, $X(1), \dots, X(N)$. Elles peuvent prendre n'importe quelle valeur dans un intervalle de \mathbb{R} et il est très rare d'avoir deux fois la même valeur pour deux individus différents. Il serait donc inutile de tracer un diagramme en bâtons comme dans le cas d'une variable discrète : il consisterait en un amoncellement illisible de bâtons de hauteur $1/N$. On choisit donc de faire un *regroupement en classes*.

2.3.1 REGROUPEMENT EN CLASSES

On considère une subdivision de l'intervalle où la variable prend ses valeurs :

$$[a_0; a_1[, [a_1; a_2[, \dots, [a_{p-1}; a_p[.$$

Noter qu'il est possible d'avoir des bornes infinies.

Comme précédemment, pour $1 \leq j \leq p$, on note n_j l'*effectif* associé à la classe $[a_{j-1}; a_j[, f_j = n_j/N$ la *fréquence relative* associée à cette classe, $\phi_j = f_1 + \dots + f_j$ la j -ième *fréquence*

cumulée croissante, et $\tilde{\phi}_j = f_j + \dots + f_p$ la j -ième fréquence cumulée décroissante. On définit de plus $l_j := a_j - a_{j-1}$ l'amplitude de la classe $[a_{j-1}; a_j[$ et $d_j := f_j/l_j$ la densité de proportion pour la classe $[a_{j-1}; a_j[$. On résume ensuite ces informations dans le tableau type suivant :

Variable X	$[a_0, a_1[$	$[a_1, a_2[$	\dots	$[a_{p-1}, a_p[$	Total
Fréquence absolue ou effectif	n_1	n_2	\dots	n_p	N
Fréquence relative	$f_1 = n_1/N$	$f_2 = n_2/N$	\dots	$f_p = n_p/N$	1
Fréquence cumulée croissante	$\phi_1 = f_1$	$\phi_2 = f_1 + f_2$	\dots	$\phi_p = 1$	X
Fréquence cumulée décroissante	$\tilde{\phi}_1 = 1$	$\tilde{\phi}_2 = f_2 + \dots + f_p$	\dots	$\tilde{\phi}_p = f_p$	X
Amplitude	$l_1 = a_1 - a_0$	$l_2 = a_2 - a_1$	\dots	$l_p = a_p - a_{p-1}$	X
Densité de proportion	$d_1 = f_1/l_1$	$d_2 = f_2/l_2$	\dots	$d_p = f_p/l_p$	X

Remarque.

1. La densité de proportion permet de comparer les effectifs dans chaque classe en tenant compte de la taille de ces classes (cf. la notion de densité de population en géographie).
2. Dans le cas de classes qui ont toutes la même longueur, il n'est pas nécessaire de calculer la densité de proportion, il est suffisant d'étudier les fréquences relatives ou absolues (qui sont directement proportionnelles à la densité de proportion).
3. Une variable quantitative discrète peut être vue comme une variable continue et on peut ainsi également effectuer un regroupement en classes pour une telle variable. Lorsque la variable prend un grand nombre de valeurs différentes, cela peut permettre une analyse plus claire des données.

Exemple 2.3. Un professeur obtient les moyennes de contrôle continu suivantes : 8,25 ; 14,25 ; 9,5 ; 14 ; 6,25 ; 11,75 ; 10 ; 10 ; 17 ; 14,75 ; 8 ; 6 ; 12,75 ; 10 ; 18,75 ; 11,5 ; 11,25 ; 19,25 ; 3 ; 13,5 ; 12,25 ; 13 ; 18,5 ; 15 ; 17 ; 10 ; 6 ; 12,25 ; 4,75 ; 16 ; 10,75 ; 9,75 ; 15,5 ; 12 ; 16,5 ; 15,25.

La population est ici le groupe de TD et la variable, que l'on notera X est la moyenne des étudiants. On obtient le tableau suivant :

X	$[0, 4[$	$[4, 8[$	$[8, 12[$	$[12, 16[$	$[16, 20]$	Total
Effectif	1	4	12	12	7	36
Fréquence relative	0,0278	0,1111	0,3333	0,3333	0,1944	1
Fréquence cumulée croissante	0,0278	0,1389	0,4722	0,8056	1	X
Fréquence cumulée décroissante	1,0000	0,9722	0,8611	0,5278	0,1944	X

Nous donnons maintenant les représentations graphiques les plus courantes.

2.3.2 HISTOGRAMME D'UNE VARIABLE CONTINUE

Pour obtenir un histogramme, on trace un repère formé de deux axes gradués orthogonaux.

- Sur l'axe des abscisses, on place les valeurs des bornes des classes,
- Sur l'axe des ordonnées, on place des pourcentages. Une classe $[a_{j-1}; a_j[$ est représentée par un rectangle de largeur l'amplitude l_j et de hauteur la densité de proportion d_j . Cela

revient en fait à ce que l'*aire* de ce rectangle soit la fréquence f_j . En effet, l'aire du rectangle représentant $[a_{j-1}; a_j[$ est :

$$l_j d_j = l_j \frac{f_j}{l_j} = f_j.$$

Voici deux histogrammes pour l'exemple 2.3, pour des largeurs de classes différentes. Dans les deux cas, l'aire totale est égale à 1.

X	$[0, 4[$	$[4, 8[$	$[8, 12[$	$[12, 16[$	$[16, 20]$	Total
Fréquence relative	0,0278	0,1111	0,3333	0,3333	0,1944	1
Densité de proportion	0,0069	0,0278	0,0833	0,0833	0,0486	0,25

X	$[0, 8[$	$[8, 12[$	$[12, 16[$	$[16, 20]$	Total
Fréquence relative	0,1389	0,3333	0,3333	0,1944	1
Densité de proportion	0,0174	0,0833	0,0833	0,0486	0,2500

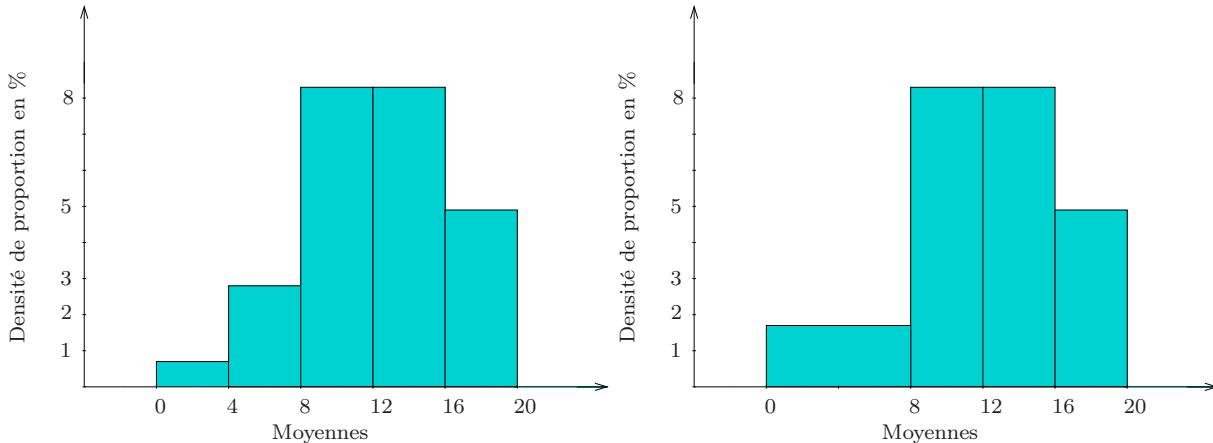
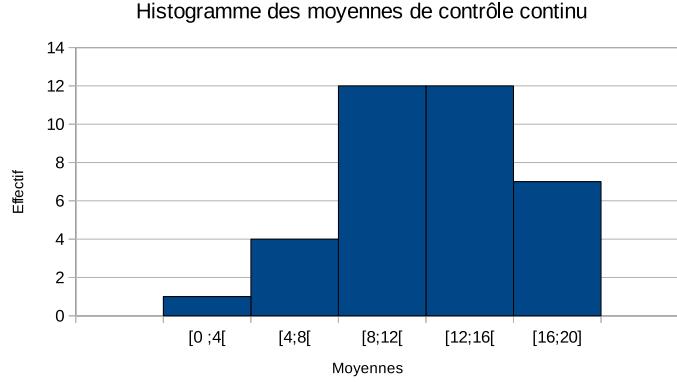


FIGURE 2.2 – Histogrammes de l'exemple 2.3 avec deux répartitions en classes différentes.

Remarque.

1. Il n'est pas possible d'avoir une classe non bornée dans cette représentation car alors l'amplitude est $+\infty$ et la densité de proportion 0. Si une classe est ouverte, il faut *introduire « artificiellement » des bornes finies*.
2. Ce qui compte dans l'histogramme est que l'aire d'une barre soit proportionnelle à l'effectif (ou la fréquence) de la classe. Ainsi, si l'amplitude des classes est constante, on peut mettre en ordonnée les effectifs ou les fréquences.
3. Avec un tableur, il ne semble pas possible de tracer un histogramme convenable lorsque l'amplitude n'est pas la même pour toutes les classes. De ce fait, lorsque nous ferons des représentations graphiques avec Excel, nous traiterons de cas où les amplitudes sont constantes.

Voici un histogramme fait avec Excel pour les effectifs des moyennes de contrôle continu, lorsque l'amplitude des classes est constante.



2.3.3 FONCTION DE RÉPARTITION EMPIRIQUE

On construit une fonction de répartition empirique de façon similaire à celle d'une variable discrète.

Définition 2.5. La *fonction de répartition empirique* de X est l'application $F_X : \mathbb{R} \rightarrow [0, 1]$ construite comme suit :

- $\forall x \in]-\infty, a_0[, F_X(x) = 0$;
- $\forall i \in \{1, \dots, p\}, \forall x \in [a_{i-1}, a_i[, F_X(x) = \phi_{i-1} + d_i(x - a_{i-1})$;
- $\forall x \in [a_p, +\infty[, F_X(x) = 1$.

La fonction de répartition empirique de l'exemple 2.3 est donnée ci-dessous.

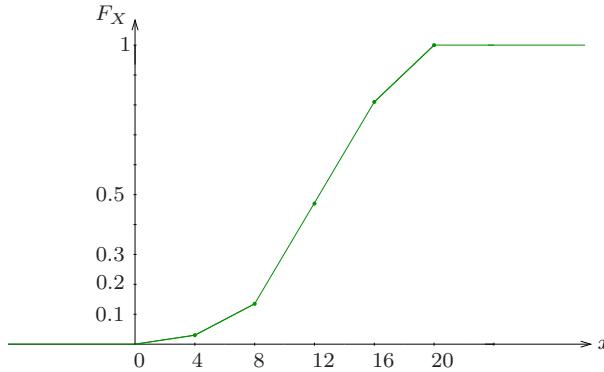


FIGURE 2.3 – Fonction de répartition empirique de l'exemple 2.3.

Remarque.

1. F_X est définie sur \mathbb{R} , affine par morceaux et croissante.
2. Pour tout $i \in \{1, \dots, p\}$, on a :

$$\phi_i = F_X(a_i) \quad \text{et} \quad d_i = \frac{F_X(a_i) - F_X(a_{i-1})}{a_i - a_{i-1}}. \quad (2.1)$$

Proposition 2.1. *L'application F_X est continue sur \mathbb{R} .*

Démonstration. L'application F_X est affine par morceaux donc continue par morceaux. Les seuls éventuels points de discontinuité de F_X sont donc les points a_0, \dots, a_p de la subdivision – qui sont également les bornes des classes.

En a_0 , on a :

$$\lim_{x \rightarrow a_0^-} F_X(x) = \lim_{x \rightarrow a_0^-} 0 = 0 \quad \text{et} \quad \lim_{x \rightarrow a_0^+} F_X(x) = \lim_{x \rightarrow a_0^+} (\phi_0 + d_1(x - a_0)) = \phi_0 = 0,$$

donc F_X est continue en a_0 .

De même, en a_p , on a :

$$\begin{aligned} \lim_{x \rightarrow a_p^-} F_X(x) &= \lim_{x \rightarrow a_p^-} (\phi_{p-1} + d_p(x - a_{p-1})) = \phi_{p-1} + d_p(a_p - a_{p-1}) = \phi_p = 1 \\ \text{et} \quad \lim_{x \rightarrow a_p^+} F_X(x) &= \lim_{x \rightarrow a_p^+} 1 = 1, \end{aligned}$$

donc F_X est continue en a_p .

Enfin, fixons $i \in \{1, \dots, p-1\}$. En a_i , on a :

$$\begin{aligned} \lim_{x \rightarrow a_i^-} F_X(x) &= \lim_{x \rightarrow a_i^-} (\phi_{i-1} + d_i(x - a_{i-1})) = \phi_{i-1} + d_i(a_i - a_{i-1}) = \phi_i \\ \text{et} \quad \lim_{x \rightarrow a_i^+} F_X(x) &= \lim_{x \rightarrow a_i^+} (\phi_i + d_{i+1}(x - a_i)) = \phi_i, \end{aligned}$$

donc F_X est continue en a_i . <

□

CHAPITRE 3

STATISTIQUE DESCRIPTIVE UNIVARIÉE

Considérons une population de taille $N \in \mathbb{N}^*$ sur laquelle on étudie *une* variable X . Ce chapitre n'a de sens que si la variable est *quantitative*, ce que nous supposons dans la suite.

Nous disposons donc d'*une série statistique univariée* à laquelle nous souhaitons associer des paramètres qui synthétisent (réduisent) l'information. Ils sont essentiellement de trois types :

- les paramètres de position,
- les paramètres de dispersion,
- les paramètres de forme.

Soit $\mathcal{P} = \{1, \dots, N\}$ la population. On suppose que les individus sont numérotés de sorte à ce que les valeurs de la variable soient croissantes :

$$X(1) \leq \dots \leq X(N).$$

Si X est discrète, on suppose qu'elle prend un nombre fini $p \in \mathbb{N}^*$ de valeurs distinctes, ordonnées en ordre croissant :

$$x_1 < \dots < x_p.$$

Si X est continue, on suppose que son image est $[\alpha, \beta]$ avec $\alpha < \beta$ des réels, et l'on procède à un regroupement en classes en considérant une subdivision (a_0, \dots, a_p) de $[\alpha, \beta]$, c'est à dire un p -uplet de réels $(a_i)_{i \in \{0, \dots, p\}}$ vérifiant $\alpha = a_0 < \dots < a_p = \beta$. Pour $i \in \{1, \dots, p\}$, on note $C_i = [a_{i-1}, a_i]$ le i -ième intervalle de la subdivision (ou i -ème classe) et $x_i = \frac{a_{i-1} + a_i}{2}$ le centre de l'intervalle C_i .

Ce regroupement en classes est *nécessaire* car on ne dispose que d'un nombre fini de données (la série statistique) pour décrire la variable. On ne peut donc qu'*approcher* une variable continue et tous les paramètres décrits par la suite sont des paramètres *approchés* pour les variables continues regroupées en classes.

Les exemples que nous donnons dans ce chapitre reprennent les données de l'exemple 2.2 pour le cas des variables discrètes. Pour rappel, nous avions obtenu le tableau suivant :

Nombre d'ordinateurs	0	1	2	3	4	5	Total
Nombre d'étudiants	300	876	1235	984	225	154	3774
Fréquence	0,0795	0,2321	0,3272	0,2607	0,0596	0,0408	1
Fréquence cumulée croissante	0,0795	0,3116	0,6388	0,8996	0,9592	1	X
Fréquence cumulée décroissante	1	0,9205	0,6884	0,3612	0,1004	0,0408	X

Dans le cas des variables continues, nous utilisons les données de l'exemple 2.3 :

Note	[0, 4[[4, 8[[8, 12[[12, 16[[16, 20]	Total
Effectif	1	4	12	12	7	36
Fréquence relative	0,0278	0,1111	0,3333	0,3333	0,1944	1
Fréquence cumulée croissante	0,0278	0,1389	0,4722	0,8056	1	pas de sens
Fréquence cumulée décroissante	1	0,9722	0,8611	0,5278	0,1944	pas de sens

3.1 PARAMÈTRES DE POSITION

3.1.1 LE MODE

Le mode contient l'information des valeurs où les données sont les plus « concentrées ».

Définition 3.1.

- Si la variable X est discrète, son *mode* est la ou les valeurs de la variable correspondant à la fréquence maximale.
- Si la variable X est continue, sa *classe modale*, est la ou les classes de densité de proportion maximale.

Exemple.

Ordinateurs. Le mode est égal à 2.

Notes. Il y a deux classes modales qui sont les classes [8, 12[et [12, 16[.

Remarque. Si X est discrète, une valeur de fréquence maximale est une valeur d'effectif maximal. Cette correspondance *n'a pas lieu* si X est continue car la densité de proportion tient compte de l'amplitude de la classe.

3.1.2 LA MOYENNE (ARITHMÉTIQUE)

La moyenne arithmétique représente le *barycentre* de la série statistique, autour de laquelle sont ensuite calculés les paramètres de dispersion.

Définition 3.2.

- Si X est discrète, la *moyenne arithmétique*, notée \bar{x} , est la somme des valeurs pondérée par les fréquences :

$$\bar{x} = \sum_{i=1}^p f_i x_i = \frac{1}{N} \sum_{i=1}^p n_i x_i.$$

À partir des données brutes, la moyenne arithmétique est le quotient de la somme des valeurs associées à chaque individu par la taille de la population :

$$\bar{x} = \frac{1}{N} \sum_{u=1}^N X(u).$$

- Si X est continue, a priori, on ne peut faire qu'un encadrement de la moyenne arithmétique. Cependant, afin de pouvoir calculer un nombre, on appellera *moyenne arithmétique* \bar{x} , l'approximation de la moyenne obtenue en prenant comme valeurs les centres des classes :

$$\bar{x} = \sum_{i=1}^p f_i x_i = \frac{1}{N} \sum_{i=1}^p n_i x_i.$$

Remarque. La moyenne arithmétique a une unité : celle des valeurs de X .

Exemple.

Ordinateur. Pour la moyenne arithmétique, on obtient :

$$\bar{x} = \frac{1}{3774} (300 \times 0 + 876 \times 1 + 1235 \times 2 + 984 \times 3 + 225 \times 4 + 154 \times 5) = 2.11,$$

ou de manière équivalente :

$$\bar{x} = (0.08 \times 0 + 0.23 \times 1 + 0.33 \times 2 + 0.26 \times 3 + 0.06 \times 4 + 0.04 \times 5) = 2.11.$$

Note. La variable étant continue, il faut considérer les centres des classes pour obtenir :

$$\bar{x} = \frac{1}{36} (2 \times 1 + 6 \times 4 + 10 \times 12 + 14 \times 12 + 18 \times 7) = 12.22,$$

ou de manière équivalente :

$$\bar{x} = (2 \times 0.0278 + 6 \times 0.1111 + 10 \times 0.3333 + 14 \times 0.3333 + 18 \times 0.1944) = 12.22,$$

La moyenne possède une propriété intéressante vis-à-vis des changements de variable affines. Précisément :

Proposition 3.1. Soient $(a, b) \in \mathbb{R}^* \times \mathbb{R}$ et Y la variable définie par $Y = aX + b$. La moyenne \bar{y} de Y satisfait à l'égalité suivante :

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

Démonstration. Si X est discrète (respectivement continue), alors Y est discrète (respectivement continue) et ses valeurs (respectivement centres des classes) y_1, \dots, y_p vérifient :

$$\forall i \in \{1, \dots, p\}, y_i = ax_i + b.$$

Donc la fréquence de la valeur (respectivement classe de centre) y_i est exactement celle de la valeur (respectivement classe de centre) x_i . Ainsi :

$$\bar{y} = \sum_{i=1}^p f_i y_i = \sum_{i=1}^p f_i (ax_i + b) = a \left(\sum_{i=1}^p f_i x_i \right) + b = a\bar{x} + b,$$

comme souhaité. □

3.1.3 MÉDIANE, QUARTILES, QUANTILES

Définition 3.3. Supposons que la variable X soit discrète. La *médiane* m est un nombre tel qu’au moins la moitié de l’effectif de la série soit inférieur ou égal à m et au moins la moitié de l’effectif de la série soit supérieur ou égal à m .

- si N est impair, $N = 2P + 1$, la série classée est :

$$X(1) \leq \cdots \leq X(P) \leq X(P+1) \leq X(P+2) \leq \cdots \leq X(2P+1).$$

La médiane est donc unique et égale à $X(P+1)$.

- si N est pair, $N = 2P$, la série classée est :

$$X(1) \leq \cdots \leq X(P) \leq X(P+1) \leq \cdots \leq X(2P).$$

Tout nombre de l’intervalle $[X(P), X(P+1)]$ convient. Il existe différentes conventions, telles que $\frac{X(P)+X(P+1)}{2}$ ou $X(P)$.

Dans le contexte de ce cours, nous fixons la médiane comme étant égale à $X(P)$.

Avec cette convention, la médiane dans le cas où N est pair ou impair satisfait à la définition suivante :

$$m = \min \left\{ x_i \mid i \in \{1, \dots, p\}, \phi_i \geq \frac{1}{2} \right\}.$$

Remarque. La médiane possède une unité : celle des valeurs de X .

Définition 3.4. Supposons que la variable X soit continue. La *classe médiane* est telle que au moins la moitié de l’effectif appartienne aux classes “inférieures ou égales” à cette classe et au moins la moitié appartienne aux classes “supérieures ou égales” à cette classe. De manière analogue au cas discret, il est plus aisé d’avoir une définition qui donne un nombre que l’on puisse calculer. On introduit donc la définition suivante.

La *médiane* d’une variable continue X , notée m , est la *plus petite* solution de l’équation,

$$F_X(x) = \frac{1}{2}, \tag{3.1}$$

où F_X est la fonction de répartition empirique de X . Autrement dit,

$$m = \min \left\{ x \in [\alpha, \beta] \mid F_X(x) = \frac{1}{2} \right\}.$$

Remarque. Comme X est continue, d’après la proposition 2.1, sa fonction de répartition empirique F_X est continue. Donc F_X vérifie le théorème des valeurs intermédiaires, ce qui montre que l’équation (3.1) admet toujours au moins une solution. Comme l’intervalle $[\alpha, \beta]$ est compact, la solution est unique.

Propriété 3.1. La médiane d’une variable continue X satisfait à l’égalité suivante :

$$m = a_{i_0-1} + \frac{a_{i_0} - a_{i_0-1}}{\phi_{i_0} - \phi_{i_0-1}} \left(\frac{1}{2} - \phi_{i_0-1} \right),$$

$$\text{où } i_0 = \min \left\{ i \in \{1, \dots, p\} \mid F_X(a_i) \geq \frac{1}{2} \right\}.$$

Démonstration. La fonction de répartition empirique F_X de X étant croissante, il existe un unique $i_0 \in \{1, \dots, p\}$ tel que :

$$F_X(a_{i_0-1}) < \frac{1}{2} \leq F_X(a_{i_0}), \quad \text{avec} \quad i_0 = \min \left\{ i \in \{1, \dots, p\} \mid F_X(a_i) \geq \frac{1}{2} \right\}.$$

De plus, on sait que $d_{i_0} \neq 0$ car $F_X(a_{i_0-1}) \neq F_X(a_{i_0})$. On résout alors l'équation (3.1) en utilisant l'expression de F_X sur $[a_{i_0-1}, a_{i_0}]$:

$$\frac{1}{2} = F_X(m) = \phi_{i_0-1} + d_{i_0}(m - a_{i_0-1}) = F_X(a_{i_0-1}) + \frac{F_X(a_{i_0}) - F_X(a_{i_0-1})}{a_{i_0} - a_{i_0-1}}(m - a_{i_0-1}),$$

en utilisant les expressions de l'effectif cumulé croissant et de la densité de proportion. On obtient ainsi :

$$m = a_{i_0-1} + \frac{a_{i_0} - a_{i_0-1}}{F_X(a_{i_0}) - F_X(a_{i_0-1})} \left(\frac{1}{2} - F_X(a_{i_0-1}) \right) = a_{i_0-1} + \frac{a_{i_0} - a_{i_0-1}}{\phi_{i_0} - \phi_{i_0-1}} \left(\frac{1}{2} - \phi_{i_0-1} \right).$$

□

On peut généraliser le principe de définition de la médiane, ce qui donne lieu à la notion de quantile.

Définition 3.5. Soit $r \in]0, 1]$.

- Si la variable X est discrète, le *quantile d'ordre r* , noté Q_r , est l'unique valeur x_{i_0} , $i_0 \in \{1, \dots, p\}$, telle que $\phi_{i_0-1} < r \leq \phi_{i_0}$:

$$Q_r = \min\{x_i \mid i \in \{1, \dots, p\} \text{ et } \phi_i \geq r\}.$$

- Si la variable X est continue, le *quantile d'ordre r* , noté Q_r , est la plus petite solution de l'équation $F_X(x) = r$, où F_X est la fonction de répartition empirique de X :

$$Q_r = \min\{x \in [\alpha, \beta] \mid F_X(x) = r\}.$$

Remarque.

1. Comme la médiane, les quantiles possèdent une unité : celle des valeurs de X .
2. En prenant $r = 1/2$, on retrouve la définition de la médiane, donc $m = Q_{1/2}$.

Définition 3.6.

- Pour tout $n \in \{1, \dots, 99\}$, le n -ième *centile* est $Q_{\frac{n}{100}}$, le quantile d'ordre $n/100$.
- Pour tout $n \in \{1, \dots, 9\}$, le n -ième *décile* est $Q_{\frac{n}{10}}$, le quantile d'ordre $n/10$. Les déciles sont aussi notés d_1, \dots, d_9 .
- Pour $n \in \{1, 2, 3\}$, le n -ième *quartile*, est $Q_{\frac{n}{4}}$, le quantile d'ordre $n/4$. Les quartiles sont aussi notés q_1, q_2, q_3 .

Remarque. La médiane est également le 50-ième centile et le second quartile. De même, les premier et troisième quartiles sont respectivement les 25-ième et 75-ième centiles.

Propriété 3.2. Le quantile d'ordre r d'une variable continue X satisfait à l'égalité suivante :

$$Q_r = a_{i_0-1} + \frac{a_{i_0} - a_{i_0-1}}{F_X(a_{i_0}) - F_X(a_{i_0-1})} (r - F_X(a_{i_0-1})) = a_{i_0-1} + \frac{a_{i_0} - a_{i_0-1}}{\phi_{i_0} - \phi_{i_0-1}} (r - \phi_{i_0-1}), \quad (3.2)$$

où $i_0 = \min\{i \in \{1, \dots, p\} \mid F_X(a_i) \geq r\}$.

Remarque. La médiane est plus robuste que la moyenne : une ou plusieurs données erronées ne font pratiquement pas, voire pas du tout, changer la médiane, alors qu'elles peuvent affecter considérablement la moyenne.

Exemple. Pour le calcul des quantiles, on regarde les fréquences cumulées croissantes.

Ordinateurs. On obtient,

$$m = 2, \ q_1 = 1, \ q_3 = 3, \ d_1 = 1, \ d_9 = 4, \ Q_{\frac{1}{100}} = 0, \ Q_{\frac{99}{100}} = 5.$$

Notes. On utilise la formule (3.2) pour obtenir :

$$\begin{aligned} m &= 12 + \frac{4}{0.8056 - 0.4722} \left(\frac{1}{2} - 0.4722 \right) = 12.34 \\ q_1 &= 8 + \frac{4}{0.4722 - 0.1389} \left(\frac{1}{4} - 0.1389 \right) = 9.33 \\ q_3 &= 12 + \frac{4}{0.8056 - 0.4722} \left(\frac{3}{4} - 0.4722 \right) = 15.34 \\ d_1 &= 4 + \frac{4}{0.1389 - 0.0278} \left(\frac{1}{10} - 0.0278 \right) = 6.60 \\ d_9 &= 16 + \frac{4}{1 - 0.8056} \left(\frac{9}{10} - 0.8056 \right) = 17.94. \end{aligned}$$

3.1.4 RÉFLEXIONS SUR LE MODE, LA MOYENNE, ET LA MÉDIANE

Les trois paramètres de *tendance centrale* que sont le mode, la moyenne (arithmétique) et la médiane possèdent un certain nombre de propriétés en commun. Elles sont définies de façon *objective* : à partir des mêmes données, des personnes différentes aboutissent aux mêmes résultats numériques. Elles s'expriment dans la même unité que celle de la valeur de la variable. Elles peuvent être interprétées de façon facile et immédiate. Elles sont simples à calculer.

La moyenne (arithmétique) possède une propriété supplémentaire importante : sa valeur dépend de toutes les observations retenues (si on fait varier une des valeurs observées, la moyenne sera toujours modifiée au contraire du mode et de la médiane qui peuvent rester inchangées). C'est essentiellement pour cette raison que la moyenne, qui est plus longue à calculer que le mode ou la médiane, est considérée comme le principal paramètre de tendance centrale.

3.1.5 MOYENNES GÉOMÉTRIQUE, HARMONIQUE ET QUADRATIQUE

Il existe des types de moyennes autres que la moyenne arithmétique. Leur utilité sera vue en exercice. Comme précédemment, si la variable est continue, on prendra comme valeurs de la série les centres des classes. Strictement parlant, il s'agit dans ce cas d'approximation des moyennes.

Définition 3.7.

1. La *moyenne géométrique* d'une variable X est :

$$G = \left(x_1^{n_1} \times x_2^{n_2} \times \cdots \times x_p^{n_p} \right)^{1/N}.$$

2. La *moyenne harmonique* d'une variable X est : que :

$$\frac{1}{H} = \frac{1}{N} \sum_{i=1}^p n_i \frac{1}{x_i}.$$

3. La *moyenne quadratique* d'une variable X est :

$$Q = \sqrt{\frac{1}{N} \sum_{i=1}^p n_i \times x_i^2}.$$

Remarque. On a :

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

Propriété 3.3. *On a l'encadrement $H \leq G \leq \bar{x}$.*

3.2 PARAMÈTRES DE DISPERSION

3.2.1 ÉTENDUE ET DISTANCE INTERQUARTILE

Rappelons qu'avec nos conventions, la plus petite valeur de la série est $X(1)$ et la plus grande est $X(N)$.

Définition 3.8. L'*étendue* de la série, notée δ_e , est la longueur du segment $[X(1), X(N)]$:

$$\delta_e = X(N) - X(1).$$

Remarque.

1. L'étendue possède une unité : celle des valeurs de X .
2. L'étendue *n'est pas très informative* car elle ne tient pas compte de la répartition des données dans le segment $[X(1), X(N)]$.

Définition 3.9. L'*intervalle inter-quartile* est le segment $[q_1, q_3]$. Il contient la moitié « centrale » des données de la série. La *distance inter-quartile*, notée δ_q , est la longueur de l'intervalle inter-quartile :

$$\delta_q = q_3 - q_1.$$

Remarque.

1. La distance inter-quartile possède une unité : celle des valeurs de X .
2. La distance inter-quartile est plus précise que l'étendue car elle ne tient compte que de des données « proches » de la médiane en un certain sens.

3.2.2 VARIANCE ET ÉCART-TYPE

Définition 3.10. Si X est discrète (respectivement continue), sa *variance*, notée $\text{Var}(X)$, est la moyenne des écarts quadratiques des valeurs (respectivement des centres des classes) de X à sa moyenne :

$$\text{Var}(X) = \sum_{i=1}^p f_i(x_i - \bar{x})^2 = \frac{1}{N} \sum_{i=1}^p n_i(x_i - \bar{x})^2.$$

Remarque.

1. La variance *n'a pas* la même unité que les valeurs de X , mais celle de leur carré.
2. Lorsque la variable est discrète, avec les données brutes, la variance de X est le quotient de la somme des distances quadratiques à la moyenne des valeurs associées à chaque individu par la taille de la population :

$$\text{Var}(X) = \frac{1}{N} \sum_{u=1}^N (X(u) - \bar{x})^2.$$

Proposition 3.2. La variance de X s'exprime également comme :

$$\text{Var}(X) = \left(\sum_{i=1}^p f_i x_i^2 \right) - \bar{x}^2 = \frac{1}{N} \left(\sum_{i=1}^p n_i x_i^2 \right) - \bar{x}^2.$$

Démonstration. Il s'agit simplement d'utiliser une identité remarquable :

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^p f_i(x_i - \bar{x})^2 = \sum_{i=1}^p f_i(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^p f_i x_i^2 - 2\bar{x} \sum_{i=1}^p f_i x_i + \bar{x}^2 \sum_{i=1}^p f_i \\ &= \sum_{i=1}^p f_i x_i^2 - 2\bar{x}(\bar{x}) + \bar{x}^2(1), \end{aligned}$$

par application de la définition de la moyenne \bar{x} et de la propriété 2.1. Le résultat suit. \square

Exemple 3.1.

Ordinateurs. On utilise la proposition précédente :

$$\begin{aligned} \text{Var}(X) &= \frac{1}{3774} (300 \times 0^2 + 876 \times 1^2 + 1235 \times 2^2 + 984 \times 3^2 + 225 \times 4^2 + 154 \times 5^2) - 2.11^2 \\ &= 1.40 \end{aligned}$$

ou de manière équivalente :

$$\begin{aligned} \text{Var}(X) &= (0.08 \times 0^2 + 0.23 \times 1^2 + 0.33 \times 2^2 + 0.26 \times 3^2 + 0.06 \times 4^2 + 0.04 \times 5^2) - 2.11^2 \\ &= 1.40. \end{aligned}$$

Notes. La variable étant continue, il faut considérer les centres des classes pour obtenir :

$$\text{Var}(X) = \frac{1}{36} \left(2^2 \times 1 + 6^2 \times 4 + 10^2 \times 12 + 14^2 \times 12 + 18^2 \times 7 \right) - 12.22^2 = 16.39,$$

ou de manière équivalente :

$$\begin{aligned} \text{Var}(X) &= \left(2^2 \times 0.0278 + 6^2 \times 0.1111 + 10^2 \times 0.3333 + 14^2 \times 0.3333 + 18^2 \times 0.1944 \right) - 12.22^2 \\ &= 16.39, \end{aligned}$$

Comme somme de carrés, la variance est un nombre positif. C'est un bon indicateur de dispersion de données car la variance s'annule si et seulement si toutes les observations effectuées sont :

- identiques si X est discrète ;
- dans la même classe si X est continue.

Précisément, on a le résultat suivant :

Théorème 3.1. *La variance de X s'annule si et seulement si :*

$$\exists i_0 \in \{1, \dots, p\}, \forall i \neq i_0, n_i = 0. \quad (3.3)$$

Auquel cas, un tel i_0 est unique, donc toutes les observations sont identiques (respectivement dans la même classe) si X est discrète (respectivement continue).

Démonstration. Supposons l'assertion (3.3) vérifiée. Alors $N = n_1 + \dots + n_p = n_{i_0}$ et la moyenne de X est :

$$\bar{x} = \frac{1}{N} \sum_{i=1}^p n_i x_i = \frac{1}{n_{i_0}} n_{i_0} x_{i_0} = x_{i_0},$$

et la variance de X vaut :

$$\text{Var}(X) = \frac{1}{N} \sum_{i=1}^p n_i (x_i - \bar{x})^2 = \frac{1}{n_{i_0}} n_{i_0} (x_{i_0} - x_{i_0})^2 = 0.$$

Réiproquement, supposons $\text{Var}(X) = 0$,

$$\begin{aligned} \text{Var}(X) = 0 &\iff \frac{1}{N} \sum_{i=1}^p n_i (x_i - \bar{x})^2 = 0 \iff \sum_{i=1}^p n_i (x_i - \bar{x})^2 = 0. \\ &\iff \forall i \in \{1, \dots, p\}, n_i (x_i - \bar{x})^2 = 0, \end{aligned}$$

car une somme de termes positifs n'est nulle que si chacun des termes la composant est nul. Ainsi, on obtient :

$$\forall i \in \{1, \dots, p\}, (n_i = 0 \text{ ou } x_i = \bar{x}).$$

Comme $N \neq 0$, il existe $i_0 \in \{1, \dots, p\}$ tel que $n_{i_0} \neq 0$, ce qui entraîne $\bar{x} = x_{i_0}$. Soit alors $i \neq i_0$. On a $x_i \neq x_{i_0}$, donc $x_i \neq \bar{x}$, ce qui entraîne $n_i = 0$ et conclut la preuve. \square

Définition 3.11. L'*écart-type* de X , noté σ_x , est la racine carrée de la variance : $\sigma_x = \sqrt{\text{Var}(X)}$.

Remarque. L'écart-type est également un bon indicateur de dispersion car il vérifie la propriété du théorème 3.1. Mais de plus, l'écart-type a la même unité que les valeurs de X .

Exemple. Grâce aux calculs de l'exemple 3.1, on obtient directement $\sigma_x = 1.18$ pour le nombre d'ordinateur et $\sigma_x = 4.05$ pour les notes.

De même que pour la moyenne à la proposition 3.1, on peut étudier le comportement de la variance et de l'écart-type par changement de variable affine.

Théorème 3.2. *Soient $(a, b) \in \mathbb{R}^* \times \mathbb{R}$ et Y la variable définie par $Y = aX + b$. Alors :*

$$\text{Var}(Y) = a^2 \text{Var}(X) \quad \text{et} \quad \sigma_y = |a|\sigma_x.$$

Démonstration. Si la variance se comporte comme annoncé, on obtient :

$$\sigma_y = \sqrt{\text{Var}(Y)} = \sqrt{a^2 \text{Var}(X)} = |a|\sqrt{\text{Var}(X)} = |a|\sigma_x.$$

Reste donc à montrer l'égalité sur les variances. Pour ce faire, on écrit :

$$\text{Var}(Y) = \sum_{i=1}^p f_i(y_i - \bar{y})^2,$$

avec f_i fréquence de la valeur (respectivement classe de centre) x_i si X est discrète (respectivement continue) et $y_i = ax_i + b$ (voir proposition 3.1). Donc :

$$\text{Var}(Y) = \sum_{i=1}^p f_i((ax_i + b) - (a\bar{x} + b))^2 = a^2 \sum_{i=1}^p f_i(x_i - \bar{x})^2 = a^2 \text{Var}(X),$$

comme escompté. □

3.3 BOX PLOT OU BOÎTE À MOUSTACHES

La boîte à moustaches, ou “box plot” en anglais, est une représentation graphique mettant en exergue certains des paramètres de position et de dispersion précédents. Elle permet de visualiser sommairement la répartition des données de la série statistique et les données “extrêmes”. Elle est constituée :

- d'un diagramme en boîte : un rectangle de bornes q_1, q_3 coupé au niveau de la médiane m et éventuellement également au niveau de la moyenne \bar{x} (alors signifié en pointillés) ;
- de moustaches : deux segments joignant q_1 à $X(1)$ pour l'un et q_3 à $X(N)$ pour l'autre.

La figure 3.1 représente l'allure générale d'une telle représentation.

Remarque. Il existe des variantes sur les bornes des moustaches. Plutôt que de les faire aller jusqu'à $X(1), X(N)$, on peut les faire aller jusqu'aux premier et neuvième décile, ou aux cinquième et 95-ième centiles, ou aux premier et 99-ième centiles, etc. On appelle alors outliers les données à l'extérieur des moustaches.

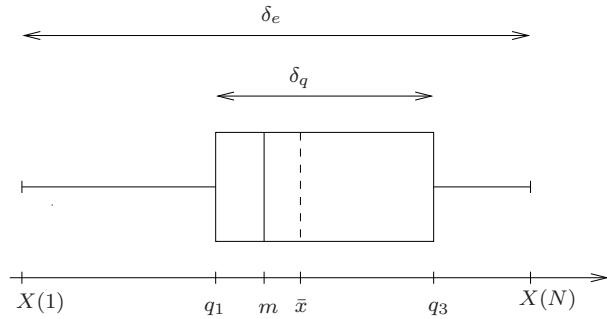


FIGURE 3.1 – Allure d'une boîte à moustaches.

3.4 PARAMÈTRES DE FORME

3.4.1 VARIABLE CENTRÉE RÉDUITE

Définition 3.12. Une variable X est dite *centrée* si elle est de moyenne arithmétique nulle. Elle est dite *réduite* si elle est de variance égale à 1.

Propriété 3.4. Soit X une variable, et Y la variable définie par,

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

Alors, la variable Y est centrée et réduite. Elle est appelée la variable centrée-réduite associée à la variable X .

Démonstration. Grâce aux formules de changements de variables affines pour la moyenne et la variance, on a :

$$\bar{y} = \frac{\bar{x} - \bar{x}}{\sigma_x} = 0 \quad \text{et} \quad \text{Var}(Y) = \frac{1}{\text{Var}(X)} \text{Var}(X) = 1.$$

□

Quand on transforme une variable en la variable centrée réduite associée, on retire à cette variable toute l'information concernant son échelle ou unité, et sa localisation. Il ne reste plus que des informations sur la forme de la distribution. Cette transformation permet de comparer plusieurs variables sur le plan de la forme, même si ce sont des variables exprimées dans des échelles différentes ou qui ont des moyennes complètement différentes.

3.4.2 MOMENTS D'UNE DISTRIBUTION

Définition 3.13.

1. On appelle *moment à l'origine* d'ordre $r \in \mathbb{N}$ le paramètre

$$m'_r = \sum_{i=1}^p f_i \cdot x_i^r = \frac{1}{N} \sum_{i=1}^p n_i \cdot x_i^r.$$

2. On appelle *moment centré* d'ordre $r \in \mathbb{N}$ le paramètre

$$m_r = \sum_{i=1}^p f_i(x_i - \bar{x})^r = \frac{1}{N} \sum_{i=1}^p n_i(x_i - \bar{x})^r.$$

Remarque.

1. Pour une variable discrète, avec les données brutes, on a

$$m'_r = \frac{1}{N} \sum_{u=1}^N X(u)^r \quad \text{et} \quad m_r = \frac{1}{N} \sum_{u=1}^N (X(u) - \bar{x})^r.$$

2. La moyenne est le moment à l'origine d'ordre 1, et la variance le moment centré d'ordre 2, qui s'exprime également en fonction des moments à l'origine d'ordre 1 et 2 :

$$m_2 = \text{Var}(X) = \frac{1}{N} \sum_{u=1}^N X(u)^2 - \bar{x}^2 = m'_2 - (m'_1)^2.$$

3.4.3 COEFFICIENT D'ASYMÉTRIE DE FISHER

Une distribution est parfaitement *symétrique*, si les valeurs qu'elle prend sont également dispersées de part et d'autre de la moyenne. Dans ce cas, son mode, sa moyenne et sa médiane sont confondues, ses quartiles q_1 et q_3 , ses déciles d_1 et d_9, \dots , sont équidistants de la valeur centrale, et son histogramme admet un axe de symétrie (symétrie par rapport à la valeur de la moyenne).

Ce n'est pas le cas pour les distributions *asymétriques* ou *obliques*. On peut mesurer l'asymétrie d'une distribution au moyen du moment centré d'ordre 3 :

Définition 3.14. Le *coefficient d'asymétrie de Fisher* est la quantité suivante :

$$\gamma_1 = \frac{m_3}{\sigma_x^3} = \frac{\sum_{i=1}^p f_i(x_i - \bar{x}_i)^3}{\sigma_x^3} = \sum_{i=1}^p f_i \left(\frac{x_i - \bar{x}_i}{\sigma_x} \right)^3.$$

Il s'agit du moment à l'origine d'ordre 3 des données centrées-réduites.

Ce coefficient peut prendre des valeurs positives, négatives ou nulles :

- si la distribution est symétrique, il est nul,
- si la distribution est allongée à gauche (les grandes valeurs sont plus fréquentes que les petites), il est négatif,
- si la distribution est allongée à droite (les petites valeurs sont plus fréquentes que les grandes), il est positif.

Exemple.

Ordinateurs. En rappelant que $\bar{x} = 2.11$, $\sigma_x = 1.18$ et en utilisant les données, on obtient :

$$\begin{aligned} \gamma_1 &= \frac{1}{1.18^3} \frac{1}{3774} \left(300 \times (0 - 2.11)^3 + 876 \times (1 - 2.11)^3 + 1235 \times (2 - 2.11)^3 \right. \\ &\quad \left. + 984 \times (3 - 2.11)^3 + 225 \times (4 - 2.11)^3 + 154 \times (5 - 2.11)^3 \right) = 0.30. \end{aligned}$$

Notes. En rappelant que $\bar{x} = 12.22$, $\sigma_x = 4.05$, et en utilisant les données (la variable étant continue, il faut considérer les centres des classes), on obtient :

$$\gamma_1 = \frac{1}{4.05^3} \frac{1}{36} \left((2 - 12.22)^3 \times 1 + (6 - 12.22)^3 \times 4 + (10 - 12.22)^3 \times 12 + (14 - 12.22)^3 \times 12 + (18 - 12.22)^3 \times 7 \right) = -0.31.$$

3.4.4 COEFFICIENT D'APLATISSEMENT

L'aplatissement (ou kurtosis en anglais) est mesuré à l'aide du moment centré d'ordre 4 :

Définition 3.15.

1. Le coefficient d'aplatissement de Pearson est la quantité suivante :

$$\beta_2 = \frac{m_4}{\sigma_x^4} = \frac{\sum_{i=1}^p f_i (x_i - \bar{x})^4}{\sigma_x^4} = \sum_{i=1}^p f_i \left(\frac{x_i - \bar{x}_i}{\sigma_x} \right)^4.$$

Il s'agit du moment à l'origine d'ordre 4 des données centrées-réduites.

2. Le coefficient d'aplatissement de Fisher ou coefficient de Yule est la quantité suivante :

$$\gamma_2 = \beta_2 - 3 = \frac{m_4}{\sigma_x^4} - 3.$$

Une distribution est dite :

- *mésokurtique* si $\gamma_2 \simeq 0$ (la densité de la loi normale, ou courbe en cloche, est mésokurtique).
- *leptokurtique* si $\gamma_2 > 0$: son histogramme est plus pointu et possède des queues plus longues (distribution moins aplatie que la distribution normale).
- *platykurtique* si $\gamma_2 < 0$: son histogramme est plus arrondi et possède des queues plus courtes (distribution plus aplatie que la distribution normale).

Exemple.

Ordinateurs. En rappelant que $\bar{x} = 2.11$, $\sigma_x = 1.18$ et en utilisant les données, on obtient :

$$\gamma_2 = \frac{1}{1.18^4} \frac{1}{3774} \left(300 \times (0 - 2.11)^4 + 876 \times (1 - 2.11)^4 + 1235 \times (2 - 2.11)^4 + 984 \times (3 - 2.11)^4 + 225 \times (4 - 2.11)^4 + 154 \times (5 - 2.11)^4 \right) - 3 = -0.11.$$

Notes. En rappelant que $\bar{x} = 12.22$, $\sigma_x = 4.05$, et en utilisant les données (la variable étant continue, il faut considérer les centres des classes), on obtient :

$$\gamma_2 = \frac{1}{4.05^4} \frac{1}{36} \left((2 - 12.22)^4 \times 1 + (6 - 12.22)^4 \times 4 + (10 - 12.22)^4 \times 12 + (14 - 12.22)^4 \times 12 + (18 - 12.22)^4 \times 7 \right) - 3 = -0.40.$$

3.5 COURBE DE CONCENTRATION DE LORENZ ET INDICE DE GINI

On dispose d'observations d'une variable décrivant une quantité correspondant à une richesse, par exemple le chiffre d'affaires, le patrimoine, les salaires. Il est donc naturel de supposer les données positives, ce qui implique en particulier que $\bar{x} > 0$.

On se pose alors des questions sur la répartition égalitaire ou non de ces richesses : les salaires sont-ils répartis de manière égalitaire ? Le chiffre d'affaire des entreprises d'un secteur est-il concentré entre les mains d'une poignée d'entreprises au détriment de toutes les autres ? Quelle fortune représentent les 3% de familles les plus riches de France ?

3.5.1 COURBE DE CONCENTRATION DE LORENZ

Définition 3.16.

1. On appelle *masses cumulées* de la variable X les quantités définies, pour $1 \leq i \leq p$, par :

$$\sum_{j=1}^i n_j x_j.$$

2. On appelle *masses cumulées relatives* de la variable X les quantités définies, pour $1 \leq i \leq p$, par :

$$M_i = \frac{\sum_{j=1}^i n_j x_j}{\sum_{j=1}^p n_j x_j} = \frac{\sum_{j=1}^i n_j x_j}{N \bar{x}} = \frac{\sum_{j=1}^i f_j x_j}{\bar{x}}.$$

3. On appelle *masse totale* le dénominateur, $\sum_{j=1}^p n_j x_j = N \bar{x}$.

4. On appelle *courbe de concentration de Lorenz* la ligne polygonale qui joint les points O et P_1, \dots, P_p définis par leurs coordonnées :

$$O = (0, 0), \quad P_1 = (\phi_1, M_1), \quad P_2 = (\phi_2, M_2), \dots, \quad P_p = (\phi_p, M_p) = (1, 1), \quad (3.4)$$

où on rappelle que les $(\phi_i)_{i \in \{1, \dots, p\}}$ sont les fréquences cumulées croissantes.

Propriété 3.5. Pour tout $i \in \{1, \dots, p\}$ on a, $0 \leq M_i \leq 1$. De plus, la courbe de concentration de Lorenz est toujours située sous la droite d'équation $y = x$.

Démonstration. La première partie est une conséquence du fait que les données soient supposées positives. Pour démontrer la seconde partie, il suffit de voir que, pour tout $i \in \{1, \dots, p\}$, $M_i \leq \phi_i$. On rappelle que,

$$M_i = \frac{\sum_{j=1}^i n_j x_j}{N \bar{x}} \quad \text{et} \quad \phi_i = \sum_{j=1}^i \frac{n_j}{N}.$$

Ainsi, on a les équivalences suivantes :

$$M_i \leq \phi_i \iff \frac{\sum_{j=1}^i n_j x_j}{\bar{x}} \leq \sum_{j=1}^i n_j \iff \sum_{j=1}^i n_j x_j \leq \sum_{j=1}^i n_j \bar{x} \iff 0 \leq \sum_{j=1}^i n_j (\bar{x} - x_j) \quad (*).$$

Soit i_0 tel que $x_{i_0} \leq \bar{x} \leq x_{i_0+1}$.

- si $i \leq i_0$, alors $\forall j \leq i$, $x_j \leq x_{i_0} \leq \bar{x}$ et donc $\bar{x} - x_j \geq 0$, et la condition (*) est vérifiée.
- si $i > i_0$, la condition (*) à démontrer peut se ré-écrire en la condition équivalente,

$$\sum_{j=i_0+1}^i n_j(x_j - \bar{x}) \leq \sum_{j=1}^{i_0} n_j(\bar{x} - x_j) \quad (**).$$

Or,

$$\begin{aligned} \sum_{j=i_0+1}^i n_j(x_j - \bar{x}) &\leq \sum_{j=i_0+1}^p n_j(x_j - \bar{x}), \text{ car les termes que l'on rajoute sont positifs} \\ &= \sum_{j=1}^p n_j(x_j - \bar{x}) - \sum_{j=1}^{i_0} n_j(x_j - \bar{x}) \\ &= \sum_{j=1}^p n_j(x_j - \bar{x}) + \sum_{j=1}^{i_0} n_j(\bar{x} - x_j) \\ &= \sum_{j=1}^{i_0} n_j(\bar{x} - x_j), \end{aligned}$$

car $\sum_{j=1}^p n_j(x_j - \bar{x}) = N\bar{x} - (\sum_{j=1}^p n_j)\bar{x} = N\bar{x} - N\bar{x} = 0$. Ceci achève la preuve de la condition (**) et donc du résultat. □

Remarque. La diagonale représente la répartition égalitaire des richesses : 20% de la population détient 20% des revenus, 60% en détient 60% des revenus, etc. Plus la courbe est éloignée de la diagonale, plus la répartition des richesses est inégalitaire.

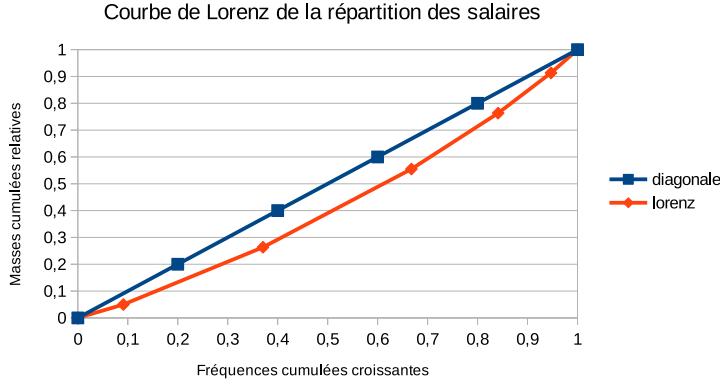
Exemple 3.2. On connaît la répartition des salaires mensuels dans une entreprise :

Salaires	[600 ; 900[[900 ; 1200[[1200 ; 1500[[1500 ; 1800[[1800 ; 2100[[2100 ; 2400[
Effectif	31	95	101	59	36	18

Afin de tracer la courbe de Lorenz, il faut calculer les fréquences cumulées ainsi que les masses cumulées relatives. On obtient le tableau suivant :

Salaires	[600 ; 900[[900 ; 1200[[1200 ; 1500[[1500 ; 1800[[1800 ; 2100[[2100 ; 2400[Total
Effectif	31	95	101	59	36	18	340
Centre de classe	750	1050	1350	1650	1950	2250	X
Fréquence	0,0912	0,2794	0,2971	0,1735	0,1059	0,0529	1
Fréquence cumulée croissante	0,0912	0,3706	0,6676	0,8412	0,9471	1	X
Masse cumulée relative	0,0497	0,2631	0,5549	0,7631	0,9133	1	X

On obtient la courbe suivante :



3.5.2 INDICE DE GINI

Définition 3.17. On appelle *indice de Gini*, noté I_G , le réel : $I_G = 2S$ où S est la surface de la partie du carré unité délimitée par la courbe de Lorenz et la diagonale d'équation $y = x$.

Remarque. On a toujours $0 \leq I_G \leq 1$. Plus I_G est élevé, plus la série est inégalitaire, car la courbe de Lorenz est d'autant plus éloignée de la diagonale. Cependant, I_G donne moins d'information que la courbe étant donné qu'il représente l'aire sous la courbe et non pas la courbe elle-même.

Proposition 3.3. L'indice de Gini est égal à :

$$I_G = 1 - \sum_{i=1}^p (M_{i-1} + M_i) \times f_i.$$

Démonstration. Pour calculer l'indice de Gini, on utilise la *méthode des trapèzes*. On observe que la surface S est égal à l'aire du triangle dont les sommets sont $(0,0)$, $(0,1)$, $(1,1)$, soit $\frac{1}{2}$, à laquelle il faut retrancher l'aire A comprise entre la courbe de Lorenz et l'axe des abscisses. Cette aire A n'est autre que la somme des aires des trapèzes dont les sommets sont $(\phi_{i-1}, 0)$, (ϕ_{i-1}, M_{i-1}) , (ϕ_i, M_i) , $(\phi_i, 0)$ pour $i \in \{1, \dots, p\}$ (avec la convention $\phi_0 = M_0 = 0$). On rappelle que l'aire d'un trapèze se calcule de la façon suivante :

$$\frac{\text{petite base} + \text{grande base}}{2} \times \text{hauteur},$$

ce qui donne :

$$\begin{aligned} I_G &= 2S = 2(0.5 - A) = 1 - 2A = 1 - 2 \sum_{i=1}^p \frac{M_{i-1} + M_i}{2} \times (\phi_i - \phi_{i-1}) \\ &= 1 - \sum_{i=1}^p (M_{i-1} + M_i) \times (\phi_i - \phi_{i-1}) = 1 - \sum_{i=1}^p (M_{i-1} + M_i) \times f_i. \end{aligned}$$

□

Exemple. Avec les données de l'exemple 3.2, on obtient :

$$\begin{aligned} I_G = 1 - & \left((0 + 0.05) \times 0.09 + (0.05 + 0.26) \times 0.28 + (0.26 + 0.55) \times 0.30 + (0.55 + 0.76) \times 0.17 \right. \\ & \left. + (0.76 + 0.91) \times 0.11 + (1 + 0.91) \times 0.05 \right) = 0.16. \end{aligned}$$

CHAPITRE 3. STATISTIQUE DESCRIPTIVE UNIVARIÉE

CHAPITRE 4

STATISTIQUE DESCRIPTIVE BI-VARIÉE

Considérons une population de taille $N \in \mathbb{N}^*$ sur laquelle on étudie *deux* variables X et Y . Ces variables peuvent être soit quantitatives discrètes ou continues, soit qualitatives.

Soit $\mathcal{P} = \{1, \dots, N\}$ la population. Nous disposons donc d'une *série statistique double* :

$$\{(X(1), Y(1)), \dots, (X(N), Y(N))\}.$$

Si X est qualitative, on suppose qu'elle prend p modalités différentes ; si elle est quantitative discrète, elle prend p valeurs différentes ; si elle est quantitative continue, on a un regroupement en p classes. Nous avons l'analogue pour la variable Y en remplaçant p par q .

On note alors x_1, \dots, x_p (respectivement y_1, \dots, y_q) les modalités/valeurs/centres des classes de la variable X (respectivement Y). De plus, cette notation respecte l'ordre naturel dans le cas d'une variable quantitative ou qualitative ordinaire.

Étudier les relations entre X et Y revient, mathématiquement, à étudier les propriétés du *couple* (X, Y) . On définit dans ce chapitre essentiellement deux types de quantités, celles dites :

- *marginales* qui ne dépendent que d'*un seul* critère mais pas des deux ;
- *conditionnelles* qui renseignent sur un critère *en fonction* des valeurs ou modalités de l'autre.

Remarque.

- Désormais, l'indice i servira exclusivement à indexer des données relatives au premier membre du couple, soit X , et l'indice j exclusivement à indexer des données relatives au second, soit Y .
- Comme déjà expliqué dans le chapitre précédent, travailler avec une variable continue dont les valeurs ne sont pas regroupées en classes revient à considérer cette variable comme discrète.

4.1 DISTRIBUTIONS MARGINALES ET CONDITIONNELLES

Il est naturel, dans un premier temps, de regarder le couple (X, Y) comme *un seul* critère d'observation. On a alors les mêmes notions qu'au chapitre 1 :

Définition 4.1.

- Soient $i \in \{1, \dots, p\}$ et $j \in \{1, \dots, q\}$. L'*effectif* du couple (x_i, y_j) , noté n_{ij} , est le nombre d'individus associés à la fois à x_i pour X et à y_j pour Y et sa *fréquence*, noté f_{ij} , est le quotient de son effectif par la taille de la population :

$$n_{ij} = |\{u \in \mathcal{P} | X(u) = x_i \text{ et } Y(u) = y_j\}| \quad \text{et} \quad f_{ij} = \frac{n_{ij}}{N}.$$

- Soit $i \in \{1, \dots, p\}$ (respectivement $j \in \{1, \dots, q\}$). L'*effectif marginal* de x_i (respectivement y_j), noté $n_{i\cdot}$ (respectivement $n_{\cdot j}$), est l'effectif de x_i (respectivement y_j) pour la série simple de X (respectivement Y) :

$$n_{i\cdot} = \sum_{j=1}^q n_{ij} \quad \text{et} \quad n_{\cdot j} = \sum_{i=1}^p n_{ij}.$$

- La *fréquence marginale* de x_i (respectivement y_j), notée $f_{i\cdot}$ (respectivement $f_{\cdot j}$), est alors le quotient de l'effectif marginal de x_i (respectivement y_j) par la taille de la population :

$$f_{i\cdot} = \frac{n_{i\cdot}}{N} = \sum_{j=1}^q f_{ij} \quad \text{et} \quad f_{\cdot j} = \frac{n_{\cdot j}}{N} = \sum_{i=1}^p f_{ij}.$$

Remarque. Les effectifs/fréquences marginaux sont également les effectifs/fréquences pour les séries statistiques simples de X ou de Y .

Propriété 4.1.

$$\begin{aligned} \sum_{i=1}^p \sum_{j=1}^q n_{ij} &= \sum_{i,j} n_{ij} = N, & \sum_{i,j} f_{ij} &= 1, \\ \sum_{i=1}^p n_{i\cdot} &= \sum_{j=1}^q n_{\cdot j} = N & \text{et} & \sum_{i=1}^p f_{i\cdot} = \sum_{j=1}^q f_{\cdot j} = 1. \end{aligned}$$

Démonstration. Les deux premières égalités sont évidentes. Par symétrie entre X et Y , il suffit de démontrer le résultat pour les sommes des $n_{i\cdot}$ et $f_{i\cdot}$. De plus, en divisant par N , on déduit du résultat sur la somme des $n_{i\cdot}$ celui sur la somme des $f_{i\cdot}$. On calcule donc :

$$\sum_{i=1}^p n_{i\cdot} = \sum_{i=1}^p \left(\sum_{j=1}^q n_{ij} \right) = \sum_{i,j} n_{ij} = N,$$

comme attendu. □

Définition 4.2. Le *tableau de contingence des effectifs* du couple de variables (X, Y) est un tableau dans lequel les valeurs/classes/modalités de X sont en lignes, celles de Y en colonnes et, pour tout $i \in \{1, \dots, p\}$ et $j \in \{1, \dots, q\}$, l'effectif n_{ij} se situe à l'intersection de la ligne i et de la colonne j . On rajoute une colonne à droite qui contient les effectifs marginaux des x_i (obtenus en faisant la somme des effectifs sur les colonnes) et une ligne en bas qui contient les effectifs marginaux des y_j (obtenus en faisant la somme des effectifs sur les lignes). La case en bas à droite contient l'effectif total N .

On peut aussi faire le *tableau de contingence des fréquences* avec les fréquences/fréquences marginales plutôt que les effectifs/effectifs marginaux.

$X \backslash Y$	y_1	\cdots	y_j	\cdots	y_q	Total
x_1	n_{11}	\cdots	n_{1j}	\cdots	n_{1q}	$n_{1\cdot}$
\vdots	\vdots		\vdots			\vdots
x_i	n_{i1}	\cdots	n_{ij}	\cdots	n_{iq}	$n_{i\cdot}$
\vdots	\vdots		\vdots			\vdots
x_p	n_{p1}	\cdots	n_{pj}	\cdots	n_{pq}	$n_{p\cdot}$
Total	$n_{\cdot 1}$	\cdots	$n_{\cdot j}$	\cdots	$n_{\cdot q}$	N

 Tableau de contingence des effectifs du couple de variables (X, Y)

$X \backslash Y$	y_1	\cdots	y_j	\cdots	y_q	Total
x_1	f_{11}	\cdots	f_{1j}	\cdots	f_{1q}	$f_{1\cdot}$
\vdots	\vdots		\vdots			\vdots
x_i	f_{i1}	\cdots	f_{ij}	\cdots	f_{iq}	$f_{i\cdot}$
\vdots	\vdots		\vdots			\vdots
x_p	f_{p1}	\cdots	f_{pj}	\cdots	f_{pq}	$f_{p\cdot}$
Total	$f_{\cdot 1}$	\cdots	$f_{\cdot j}$	\cdots	$f_{\cdot q}$	1

 Tableau de contingence des fréquences du couple (X, Y) .

Remarque. Un tableau de contingence se lit comme une matrice : le premier indice est celui de la ligne et le second celui de la colonne.

En extrayant une colonne (respectivement une ligne) du tableau de contingence, on obtient une *distribution conditionnelle* de X (respectivement de Y). Précisément, l'extraction de la colonne j (respectivement de la ligne i) permet de répondre à la question « Sachant que Y vaut y_j (respectivement X vaut x_i), comment se comporte X (respectivement Y) ? » Ainsi, une distribution conditionnelle est l'étude d'un critère lorsque l'autre reste fixe – à une valeur ou modalité connue. On note $X|Y=y_j$ (respectivement $Y|X=x_i$) la *distribution conditionnelle de X sachant $Y = y_j$* (respectivement de Y sachant $X = x_i$).

Définition 4.3. Soient $i \in \{1, \dots, p\}$ et $j \in \{1, \dots, q\}$. La *fréquence conditionnelle* de x_i sachant y_j (respectivement y_j sachant x_i), notée $f_{i|Y=y_j}$ ou $f_{i|j}$ (respectivement $f_{j|X=x_i}$ ou $f_{j|i}$), est la fréquence d'occurrence de x_i dans la colonne j (respectivement y_j dans la ligne i) du tableau de contingence :

$$f_{i|Y=y_j} = \frac{n_{ij}}{n_{\cdot j}} \quad \text{et} \quad f_{j|X=x_i} = \frac{n_{ij}}{n_{i\cdot}}.$$

Propriété 4.2. Les fréquences conditionnelles vérifient les égalités suivantes :

$$\forall j \in \{1, \dots, q\}, \sum_{i=1}^p f_{i|Y=y_j} = 1 \quad \text{et} \quad \forall i \in \{1, \dots, p\}, \sum_{j=1}^q f_{j|X=x_i} = 1.$$

Démonstration. De nouveau par symétrie entre X et Y , il suffit de montrer, par exemple, la

première égalité. Soit donc $j \in \{1, \dots, q\}$. On écrit :

$$\sum_{i=1}^p f_{i|Y=y_j} = \sum_{i=1}^p \frac{n_{ij}}{n_{\cdot j}} = \frac{1}{n_{\cdot j}} \sum_{i=1}^p n_{ij} = \frac{1}{n_{\cdot j}} n_{\cdot j} = 1.$$

□

Définition 4.4. Le tableau des *profils-colonne* représente les distributions conditionnelles de $X|Y$. On ajoute une ligne en bas avec les totaux, égaux à 1 d'après la propriété ci-dessus. Le tableau des *profils-ligne* représente les distributions conditionnelles $Y|X$. On ajoute une colonne à droite avec les totaux, égaux à 1 d'après la propriété ci-dessus.

$\begin{array}{c} Y \\ \diagdown \\ X Y \end{array}$	y_1	\cdots	y_q
x_1	$f_{1 Y=y_1}$	\cdots	$f_{1 Y=y_q}$
\vdots	\vdots		\vdots
x_p	$f_{p Y=y_1}$	\cdots	$f_{p Y=y_q}$
Total	1	\cdots	1

Tableau des profils-colonne

$\begin{array}{c} Y X \\ \diagdown \\ X \end{array}$	y_1	\cdots	y_q	Total
x_1	$f_{1 X=x_1}$	\cdots	$f_{q X=x_1}$	1
\vdots	\vdots			\vdots
x_p	$f_{1 X=x_p}$	\cdots	$f_{q X=x_p}$	1

Tableau des profils-ligne

Remarque. Un calcul simple donne pour tout $i \in \{1, \dots, p\}$ et $j \in \{1, \dots, q\}$:

$$f_{ij} = f_{i|Y=y_j} f_{\cdot j} \quad \text{et} \quad f_{ij} = f_{j|X=x_i} f_{i\cdot}.$$

4.2 INDÉPENDANCE

Une question centrale dans les relations entre variables est celle de l'indépendance.

4.2.1 DÉFINITION

Définition 4.5. Les variables X et Y sont dites *indépendantes* lorsque :

$$\forall (i, j) \in \{1, \dots, p\} \times \{1, \dots, q\}, f_{ij} = f_{i\cdot} f_{\cdot j}. \quad (4.1)$$

Proposition 4.1. Soient deux variables X, Y . Les assertions suivantes sont équivalentes :

- (i) X et Y sont indépendantes.
- (ii) $\forall (i, j) \in \{1, \dots, p\} \times \{1, \dots, q\}, n_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{N}.$
- (iii) $\forall (i, j) \in \{1, \dots, p\} \times \{1, \dots, q\}, f_{i|Y=y_j} = f_{i\cdot}.$
- (iv) $\forall (i, j) \in \{1, \dots, p\} \times \{1, \dots, q\}, f_{j|X=x_i} = f_{\cdot j}.$

Démonstration. Il suffit de modifier l'écriture de l'égalité (4.1) définissant l'indépendance. On considère donc $i \in \{1, \dots, p\}$ et $j \in \{1, \dots, q\}$. Pour l'équivalence entre les points (i) et (ii) on écrit :

$$f_{ij} = f_{i\cdot} f_{\cdot j} \iff \frac{n_{ij}}{N} = \frac{n_{i\cdot}}{N} \frac{n_{\cdot j}}{N} \iff n_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{N}.$$

Pour l'équivalence entre les points (ii) et (iii) on écrit :

$$n_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{N} \iff \frac{n_{ij}}{n_{\cdot j}} = \frac{n_{i\cdot}}{N} \iff f_{i|Y=y_j} = f_{i\cdot}.$$

Et de façon similaire :

$$n_{ij} = \frac{n_{i \cdot} n_{\cdot j}}{N} \iff \frac{n_{ij}}{n_{i \cdot}} = \frac{n_{\cdot j}}{N} \iff f_{j|X=x_i} = f_{\cdot j},$$

on obtient l'équivalence entre les points (ii) et (iv). \square

4.2.2 ÉCART À L'INDÉPENDANCE

Étant donné les variables X et Y , on s'intéresse à savoir si elles sont indépendantes ou non ; et à quantifier l'écart à l'indépendance. Pour cela, on peut calculer la quantité suivante.

Définition 4.6. Le *chi-carré observé* est :

$$\chi^2 = N \sum_{i,j} \frac{(f_{ij} - f_{i \cdot} f_{\cdot j})^2}{f_{i \cdot} f_{\cdot j}} = \sum_{i,j} \left(\frac{n_{i \cdot} n_{\cdot j}}{N} \right)^{-1} \left(n_{ij} - \frac{n_{i \cdot} n_{\cdot j}}{N} \right)^2.$$

Propriété 4.3. Le χ^2 est toujours positif ou nul et,

$$\chi^2 = 0 \iff \text{les variables } X \text{ et } Y \text{ sont indépendantes.}$$

Il se trouve que le χ^2 observé croît avec la taille de la population. Pour cette raison, il est utile de le renormaliser de manière adéquate.

Définition 4.7. Le coefficient V de Cramer du couple de variables (X, Y) est :

$$V = \sqrt{\frac{\chi^2}{N \min(p-1, q-1)}}.$$

Propriété 4.4. Le coefficient V de Cramer vérifie :

$$0 \leq V \leq 1,$$

et $V = 0$ si et seulement si les variables sont indépendantes.

Ainsi le V de Cramer permet de mesurer l'écart à l'indépendance et sa taille ne dépend pas de la taille de la population.

4.3 NUAGE DE POINTS, RÉSUMÉS NUMÉRIQUES

Cette section n'a de sens que si les variables X et Y sont quantitatives, discrètes ou continues et regroupées en classes, ce que nous supposons dans la suite.

4.3.1 NUAGE DE POINTS

Rappelons que nous avons à disposition la statistique double,

$$\{(X(1), Y(1)), \dots, (X(N), Y(N))\}.$$

Cette dernière peut être représentée par un nuage de N points dans \mathbb{R}^2 , noté $\mathcal{N}(\mathcal{P})$:

$$\mathcal{N}(\mathcal{P}) = \{(X(u), Y(u)) \in \mathbb{R}^2 : u \in \{1, \dots, N\}\}.$$

4.3.2 MOYENNE, VARIANCE, ÉCART-TYPE

En considérant les variables X et Y indépendamment l'une de l'autre, on peut définir tous les paramètres du chapitre 3, que l'on qualifie alors de *marginal*.

LES MOYENNES MARGINALES

En reprenant les définitions du chapitre 3, nous avons :

$$\bar{x} = \sum_{i=1}^p f_i x_i = \sum_{i=1}^p \left(\sum_{j=1}^q f_{ij} \right) x_i = \sum_{i,j} f_{ij} x_i \quad \text{et} \quad \bar{y} = \sum_{j=1}^q f_j y_j = \sum_{j=1}^q \left(\sum_{i=1}^p f_{ij} \right) y_j = \sum_{i,j} f_{ij} y_j.$$

Si les variables X et Y sont discrètes, à partir des données brutes, on a :

$$\bar{x} = \frac{1}{N} \sum_{u=1}^N X(u), \quad \text{et} \quad \bar{y} = \frac{1}{N} \sum_{u=1}^N Y(u).$$

Définition 4.8. Le point $G = (\bar{x}, \bar{y})$ est appelé le *barycentre* ou *centre de gravité* du nuage de points.

La proposition suivante montre que la moyenne est linéaire.

Proposition 4.2. Soient $a, b \in \mathbb{R}$. La moyenne \bar{z} de la variable $Z = aX + bY$ s'écrit :

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

Démonstration. La variable $Z = aX + bY$ prend comme valeurs $z_{ij} = ax_i + by_j$, pour $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$, avec f_{ij} comme fréquence pour la valeur z_{ij} . Ainsi, sa moyenne \bar{z} est :

$$\bar{z} = \sum_{i,j} f_{ij} z_{ij} = \sum_{i,j} f_{ij} (ax_i + by_j) = a \sum_{i,j} f_{ij} x_i + b \sum_{i,j} f_{ij} y_j = a\bar{x} + b\bar{y}. \quad \square$$

LES VARIANCES ET ÉCARTS-TYPES MARGINaux

D'après le chapitre 3, nous avons :

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^p f_i (x_i - \bar{x})^2 = \sum_{i=1}^p \left(\sum_{j=1}^q f_{ij} \right) (x_i - \bar{x})^2 = \sum_{i,j} f_{ij} (x_i - \bar{x})^2 \\ \text{Var}(Y) &= \sum_{j=1}^q f_j (y_j - \bar{y})^2 = \sum_{j=1}^q \left(\sum_{i=1}^p f_{ij} \right) (y_j - \bar{y})^2 = \sum_{i,j} f_{ij} (y_j - \bar{y})^2. \end{aligned}$$

Les écarts-types σ_x, σ_y sont les racines des variances $\text{Var}(X), \text{Var}(Y)$.

Si les variables X et Y sont discrètes, à partir des données brutes, on a :

$$\text{Var}(X) = \frac{1}{N} \sum_{u=1}^N (X(u) - \bar{x})^2, \quad \text{et} \quad \text{Var}(Y) = \frac{1}{N} \sum_{u=1}^N (Y(u) - \bar{y})^2.$$

Attention, à la différence de la moyenne, la variance n'est pas linéaire. Afin de calculer la variance d'une combinaison linéaire de variables, nous devons d'abord introduire la covariance.

4.3.3 COVARIANCE

Soient X, Y deux variables quantitatives.

Définition 4.9. La covariance de X et Y , notée $\text{Cov}(X, Y)$, est la moyenne des produits des écarts des valeurs ou centres des classes de X et Y à leur moyenne respective :

$$\text{Cov}(X, Y) = \sum_{i,j} f_{ij}(x_i - \bar{x})(y_j - \bar{y}) = \frac{1}{N} \sum_{i,j} n_{ij}(x_i - \bar{x})(y_j - \bar{y}).$$

Remarque.

1. La covariance possède une unité : celle du produit des unités de X et Y .
2. La covariance est symétrique : $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
3. En faisant $Y = X$, on obtient $\text{Cov}(X, X) = \text{Var}(X)$.
4. Si les variables X et Y sont discrètes, à partir des données brutes on a :

$$\text{Cov}(X, Y) = \frac{1}{N} \sum_{u=1}^N (X(u) - \bar{x})(Y(u) - \bar{y}).$$

Proposition 4.3. La covariance de X et Y vérifie les propriétés suivantes :

1. La covariance de X et Y s'exprime également comme :

$$\text{Cov}(X, Y) = \left(\sum_{i,j} f_{ij} x_i y_j \right) - \bar{x} \bar{y} = \left(\frac{1}{N} \sum_{i,j} n_{ij} x_i y_j \right) - \bar{x} \bar{y}.$$

Si les variables X et Y sont discrètes, à partir des données brutes on a :

$$\text{Cov}(X, Y) = \left(\frac{1}{N} \sum_{u=1}^N X(u) Y(u) \right) - \bar{x} \bar{y}.$$

2. Soient $(a, b), (c, d) \in \mathbb{R}^* \times \mathbb{R}$. On a la formule de changement de variables suivante :

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y).$$

3. Soient $a, b \in \mathbb{R}$. La variance de la variable $aX + bY$ s'écrit :

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + 2ab \text{Cov}(X, Y) + b^2 \text{Var}(Y).$$

Démonstration. Pour le point 1, on développe la formule de la covariance :

$$\begin{aligned} \text{Cov}(X, Y) &= \sum_{i,j} f_{ij}(x_i - \bar{x})(y_j - \bar{y}) = \sum_{i,j} f_{ij}(x_i y_j - x_i \bar{y} - y_j \bar{x} + \bar{x} \bar{y}) \\ &= \left(\sum_{i,j} f_{ij} x_i y_j \right) - \bar{y} \left(\sum_{i,j} f_{ij} x_i \right) - \bar{x} \left(\sum_{i,j} f_{ij} y_j \right) + \bar{x} \bar{y} \left(\sum_{i,j} f_{ij} \right), \end{aligned}$$

et en utilisant :

$$\bar{x} = \sum_{i,j} f_{ij} x_i, \quad \bar{y} = \sum_{i,j} f_{ij} y_j \quad \text{et} \quad \sum_{i,j} f_{ij} = 1,$$

on obtient le résultat.

Pour le point 2, on reprend la définition de la covariance sachant que les moyennes de $aX + b$ et $cY + d$ sont respectivement $a\bar{x} + b$ et $c\bar{y} + d$:

$$\begin{aligned}\text{Cov}(aX + b, cY + d) &= \sum_{i,j} f_{ij} ((ax_i + b) - (a\bar{x} + b)) ((cy_j + d) - (c\bar{y} + d)) \\ &= ac \sum_{i,j} f_{ij} (x_i - \bar{x})(y_j - \bar{y}) = ac \text{Cov}(X, Y).\end{aligned}$$

Pour le point 3, posons $Z = aX + bY$. Alors, d'après la proposition 4.2, $\bar{z} = a\bar{x} + b\bar{y}$. Ainsi,

$$\begin{aligned}\text{Var}(aX + bY) &= \text{Var } Z = \sum_{i,j} f_{ij} (z_{ij} - \bar{z})^2 \\ &= \sum_{i,j} f_{ij} ((ax_i + by_j) - (a\bar{x} + b\bar{y}))^2 \\ &= \sum_{i,j} f_{ij} (a(x_i - \bar{x}) + b(y_j - \bar{y}))^2 \\ &= a^2 \sum_{i,j} f_{ij} (x_i - \bar{x})^2 + 2ab \sum_{i,j} f_{ij} (x_i - \bar{x})(y_j - \bar{y}) + b^2 \sum_{i,j} f_{ij} (y_j - \bar{y})^2 \\ &= a^2 \text{Var}(X) + 2ab \text{Cov}(X, Y) + b^2 \text{Var}(Y),\end{aligned} \quad \square$$

La covariance permet de quantifier la dépendance de X et Y dans le sens suivant :

Théorème 4.1. *Si les variables X et Y sont indépendantes, alors $\text{Cov}(X, Y) = 0$.*

Démonstration. On reprend l'écriture de la covariance du point 1. de la proposition 4.3, que l'on combine avec les expressions des moyennes marginales :

$$\begin{aligned}\text{Cov}(X, Y) &= \left(\sum_{i,j} f_{ij} x_i y_j \right) - \bar{x} \bar{y} = \left(\sum_{i,j} f_{ij} x_i y_j \right) - \left(\sum_{i=1}^p f_{i\cdot} x_i \right) \left(\sum_{j=1}^q f_{\cdot j} y_j \right) \\ &= \left(\sum_{i,j} f_{ij} x_i y_j \right) - \left(\sum_{i,j} f_{i\cdot} f_{\cdot j} x_i y_j \right) = \sum_{i,j} (f_{ij} - f_{i\cdot} f_{\cdot j}) x_i y_j.\end{aligned}$$

Or, par définition, si X et Y sont indépendantes, pour tout $i \in \{1, \dots, p\}$, $j \in \{1, \dots, q\}$, on a $f_{ij} = f_{i\cdot} f_{\cdot j}$; d'où $\text{Cov}(X, Y) = 0$. \square

Corollaire 4.2. *Si les variables X et Y sont indépendantes, alors :*

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

4.3.4 COEFFICIENT DE CORRÉLATION

Théorème 4.3 (Inégalité de Cauchy-Schwarz). *Soient X, Y des variables discrètes ou continue regroupées en classes. On a l'inégalité suivante :*

$$[\text{Cov}(X, Y)]^2 \leq \text{Var}(X) \text{Var}(Y) \quad \text{ou de façon équivalente} \quad |\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y. \quad (4.2)$$

De plus, si les variables X et Y ne sont pas constantes, alors l'égalité dans l'équation (4.2) a lieu si et seulement si :

$$Y = \epsilon \frac{\sigma_Y}{\sigma_X} X + b \quad \text{avec } \epsilon \in \{-1, 1\} \text{ et } b \in \mathbb{R}.$$

Démonstration. On considère la variable $\lambda X + Y$ où $\lambda \in \mathbb{R}$. On a donc d'après le point 3. de la proposition 4.3 :

$$\text{Var}(\lambda X + Y) = \lambda^2 \text{Var}(X) + \lambda(2 \text{Cov}(X, Y)) + \text{Var}(Y),$$

que l'on voit comme un trinôme du second degré en λ de discriminant :

$$[2 \text{Cov}(X, Y)]^2 - 4 \text{Var}(X) \text{Var}(Y) = 4[\text{Cov}(X, Y)]^2 - 4 \text{Var}(X) \text{Var}(Y).$$

De plus, comme $\text{Var}(\lambda X + Y) \geq 0$, on sait que ce discriminant est négatif, d'où :

$$\begin{aligned} [\text{Cov}(X, Y)]^2 - \text{Var}(X) \text{Var}(Y) \leq 0 &\iff [\text{Cov}(X, Y)]^2 \leq \text{Var}(X) \text{Var}(Y) \\ &\iff |\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y, \end{aligned}$$

ce qui montre la première partie. De plus, l'égalité dans l'équation (4.2) signifie d'une part que $\text{Cov}(X, Y) = \epsilon \sigma_X \sigma_Y$ pour un certain $\epsilon \in \{-1, 1\}$ et d'autre part que le discriminant est nul, donc que le polynôme admet une (unique, car X et Y ne sont pas constantes) racine :

$$\lambda_0 = -\frac{2 \text{Cov}(X, Y)}{2 \text{Var}(X)} = -\epsilon \frac{\sigma_Y}{\sigma_X}.$$

Mais alors, la variable $\lambda_0 X + Y$ est de variance nulle, donc constante de valeur notée $b \in \mathbb{R}$, d'après le théorème 3.1. \square

Définition 4.10. Lorsque les variables X et Y ne sont pas constantes, le *coefficient de corrélation* de X et Y , noté $r(X, Y)$, est le quotient de la covariance de X et Y par le produit des écarts-type de X et Y :

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

Corollaire 4.4 (du théorème 4.3). *Le coefficient de corrélation de X et Y vérifie :*

$$-1 \leq r(X, Y) \leq 1.$$

De plus, on a $|r(X, Y)| = 1$ si et seulement si X et Y vérifient une relation affine de la forme $Y = aX + b$, avec $b \in \mathbb{R}$ et :

$$r(X, Y) = 1 \iff a = \frac{\sigma_Y}{\sigma_X} > 0 \quad \text{et} \quad r(X, Y) = -1 \iff a = -\frac{\sigma_Y}{\sigma_X} < 0.$$

Remarque. Lorsque le coefficient de corrélation est « proche » de 1 en valeur absolue – prenons $|r(X, Y)| \geq 0,9$ pour fixer les idées –, cela signifie qu'il y a une *liaison affine* importante entre les variables X et Y ; cette liaison étant exacte lorsque $|r(X, Y)| = 1$. En revanche, lorsque le coefficient de corrélation est proche de 0, alors il n'existe pas de relation affine entre X et Y , ce qui ne signifie pas qu'un autre type de liaison ne peut avoir lieu.

4.3.5 AJUSTEMENT AFFINE

On suppose ici qu'il y a une liaison affine importante entre X et Y . Le principe d'un *ajustement affine*, aussi appelé *régression linéaire*, est de déterminer la « meilleure » droite pour exprimer la liaison affine entre X et Y , c'est-à-dire la droite qui approxime au mieux le nuage de points $\mathcal{N}(\mathcal{P})$.

Remarque. Dans les cas où la liaison entre X et Y est « indirectement » affine, c'est-à-dire de la forme $Y = af(X) + b$, on détermine dans un premier temps la fonction f . On étudie ensuite la variable $Z = f(X)$ et l'on effectue une régression linéaire sur le couple (Z, Y) . C'est une *régression linéaire après changement de variable*.

MÉTHODE DES MOINDRES CARRÉS

Pour tous $a, b \in \mathbb{R}$, on considère la variable $\Delta_{(a,b)} = Y - (aX + b)$. Cette variable contient l'information sur l'erreur commise si l'on modélise Y à partir de X par le changement de variable affine $aX + b$. On calcule la moyenne et la variance de $\Delta_{(a,b)}$:

$$\begin{aligned}\bar{\Delta}_{(a,b)} &= \bar{y} - (a\bar{x} + b) \quad \text{et} \\ \text{Var}(\Delta_{(a,b)}) &= \text{Var}(Y - (aX + b)) \\ &= \text{Var}(Y) - 2 \text{Cov}(Y, aX + b) + \text{Var}(aX + b) \\ &= \text{Var}(Y) - 2a \text{Cov}(X, Y) + a^2 \text{Var}(X).\end{aligned}$$

La *méthode des moindres carrés* consiste à déterminer $a, b \in \mathbb{R}$ de telle sorte que $\Delta_{(a,b)}$ soit le plus proche de la variable nulle, ce qui revient à minimiser $|\bar{\Delta}_{(a,b)}|$ et $\text{Var}(\Delta_{(a,b)})$.

Proposition 4.4 (Méthode des moindres carrés). *Si X n'est pas constante, la droite de régression par la méthode des moindres carrés de Y en fonction de X est la droite d'équation $y = \hat{a}x + \hat{b}$ dans le repère du nuage de points de (X, Y) avec :*

$$\hat{a} = r(X, Y) \frac{\sigma_Y}{\sigma_X} \quad \text{et} \quad \hat{b} = \bar{y} - r(X, Y) \frac{\sigma_Y}{\sigma_X} \bar{x}.$$

Démonstration. Comme $\text{Var}(\Delta_{(a,b)})$ est un trinôme du second degré en a de discriminant négatif (car la variance est toujours positive), il existe une unique valeur \hat{a} de a minimisant $\text{Var}(\Delta_{(a,b)})$:

$$\hat{a} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = r(X, Y) \frac{\sigma_Y}{\sigma_X}.$$

On peut alors fixer b pour annuler la moyenne :

$$\hat{b} = \bar{y} - \hat{a}\bar{x} = \bar{y} - r(X, Y) \frac{\sigma_Y}{\sigma_X} \bar{x}. \quad \square$$

Remarque.

1. La droite de régression obtenue par la méthode des moindres carrés passe par le barycentre (\bar{x}, \bar{y}) du nuage de points $\mathcal{N}(\mathcal{P})$.

2. En prenant $\bar{\Delta}_{(a,b)} = 0$, on obtient :

$$\text{Var}(\Delta_{(a,b)}) = \frac{1}{N} \sum_{u=1}^N (\Delta_{(a,b)}(u))^2 = \frac{1}{N} \sum_{u=1}^N (Y(u) - (aX(u) + b))^2,$$

qui est $\frac{1}{N}$ fois la somme des distances au carré entre les points $(X(u), Y(u))$ du nuage et les points $(X(u), aX(u) + b)$ de la droite. Ceci explique le nom de méthode des moindres carrés.

MÉTHODE DE MAYER

Cette méthode d'ajustement est plus simple à mettre en place. Elle consiste, après avoir rangé les couples (x_i, y_i) selon l'ordre croissant des x_i , à fractionner la série statistique en deux séries dont les effectifs sont égaux (pour un nombre pair de points) ou différents de 1 (pour un nombre impair de points). On calcule pour chacune de ces deux parties les coordonnées de leurs points moyens $G_1 = (\bar{x}_1, \bar{y}_1)$ et $G_2 = (\bar{x}_2, \bar{y}_2)$. On trace enfin la droite (G_1G_2) .

Pour trouver l'équation de la droite obtenue, on utilise le fait qu'elle passe par les points G_1 et G_2 afin d'obtenir :

$$\bar{y}_1 = a\bar{x}_1 + b \quad \text{et} \quad \bar{y}_2 = a\bar{x}_2 + b.$$

Ce qui donne

$$a = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2} \quad \text{et} \quad b = \bar{y}_1 - \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2} \bar{x}_1 = \frac{\bar{x}_1 \bar{y}_2 - \bar{y}_1 \bar{x}_2}{\bar{x}_1 - \bar{x}_2}.$$

Proposition 4.5 (Méthode de Mayer). *Après avoir classé la série statistique (X, Y) par ordre croissant des x_i , on la divise en deux sous-séries (X_1, Y_1) et (X_2, Y_2) de même effectif (ou différents de 1). On désigne par $\bar{x}_1, \bar{y}_1, \bar{x}_2$ et \bar{y}_2 les moyennes de X_1, Y_1, X_2 et Y_2 respectivement. La droite de régression par la méthode de Mayer de Y en fonction de X est la droite d'équation $y = \tilde{a}x + \tilde{b}$ avec :*

$$\tilde{a} = \frac{\bar{y}_1 - \bar{y}_2}{\bar{x}_1 - \bar{x}_2} \quad \text{et} \quad \tilde{b} = \frac{\bar{x}_1 \bar{y}_2 - \bar{y}_1 \bar{x}_2}{\bar{x}_1 - \bar{x}_2}.$$

PRÉVISIONS

A l'aide de l'équation $y = ax + b$ de la droite obtenu par la méthode des moindres carrés ou par la méthode de Mayer, il est possible d'estimer la valeur que prendrait la variable X (respectivement Y) pour une valeur de Y donnée (respectivement X).

Pour une valeur x_0 de la variable X qui ne fait pas partie des observations, on peut faire une prédition de la valeur correspondante de Y en calculant l'ordonnée du point d'abscisse x_0 en prenant $y_0 = ax_0 + b$.

Pour une valeur y_0 de la variable Y qui ne fait pas partie des observations, on peut faire une prédition de la valeur correspondante de X en calculant l'antécédent du point d'ordonnée y_0 en prenant $x_0 = \frac{y_0 - b}{a}$.

4.3.6 UN EXEMPLE

On s'intéresse au lien éventuel entre le nombre d'heures d'ensoleillement et la température du mois de juillet dans plusieurs villes de France.

Ville	Température moyenne en juillet (X)	Ensoleillement, en heures par an (Y)
Ajaccio	22,2	2726
Bordeaux	20,8	1992
Clermont-Ferrand	19,7	1898
Brest	16,6	1492
Lille	17,9	1617
Lyon	21,3	2010
Millau	19,3	2121
Nice	23,1	2668
Paris	20	1630
Strasbourg	19,5	1633
Toulouse	21,6	2010
Fort-de-France	27,5	2437
Papeete	25	2685

On obtient directement :

$$\bar{x} = 21.1, \quad \bar{y} = 2071, \quad \text{Var}(X) = 7.8, \quad \text{Var}(Y) = 174629, \quad \sigma_X = 2.8, \quad \sigma_Y = 418.$$

On calcule ensuite la covariance :

$$\text{Cov}(X, Y) = \left(22.2 \times 2726 + 20.8 \times 1992 + 19,7 \times 1898 + \dots + 25 \times 2685 \right) - 21.1 * 2071 = 925.3,$$

d'où le coefficient de corrélation suivant :

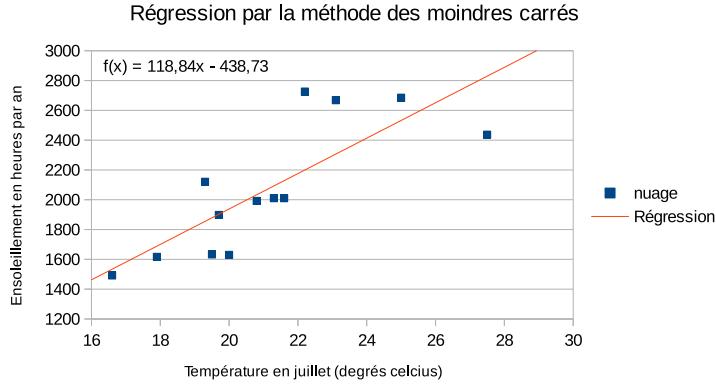
$$r(X, Y) = \frac{925.3}{2.8 * 418} = 0.79.$$

Bien que le coefficient de corrélation ne soit pas supérieur à 0.9, il en est proche et on va donc effectuer un ajustement affine.

Commençons par utiliser la méthode des moindres carrés : les coefficients de la droite sont donnés par :

$$\hat{a} = 0.79 * \frac{418}{2.8} = 118.84 \quad \text{et} \quad \hat{b} = \bar{y} - \hat{a}\bar{x} = 2071 - 118.84 * 21.1 = -438.73.$$

On obtient le graphique suivant :



Pour la méthode de Mayer, on commence par classer les données par ordre croissant en X :

$$(16.6; 1492), (17.9; 1617), (19.3; 2121), \dots, (25; 2685), (27.5; 2437).$$

On découpe ensuite la série en deux sous-séries de même taille (on a une taille impaire ici, on choisit donc (arbitrairement) de prendre un individu de plus dans la première sous-série). On obtient ainsi :

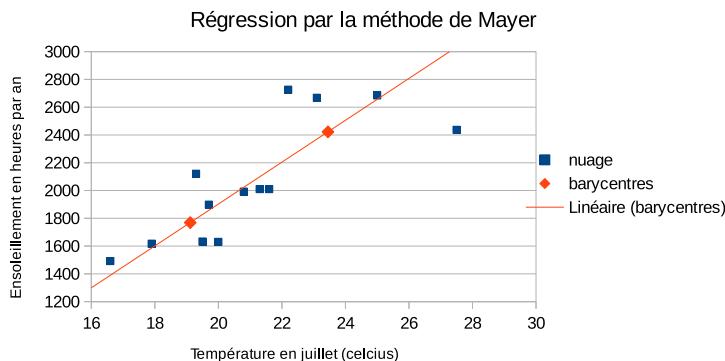
$$S_1 = \{(16.6; 1492), (17.9; 1617), (19.3; 2121), (19.5; 1633), (19.7; 1898), (20; 1630), (20.8, 1992)\},$$

$$S_2 = \{(21.3; 2010), (21.6; 2010), (22.2; 2726), (23.1, 2668), (25; 2685), (27.5; 2437)\}.$$

Pour tracer la droite de régression en utilisant la méthode de Mayer, il nous reste à déterminer les moyennes de chaque sous-série afin d'obtenir les coordonnées des deux points moyens :

$$G_1 = (\bar{x}_1, \bar{y}_1) = (19.1; 1769) \quad \text{et} \quad G_2 = (23.5; 2423).$$

On obtient le graphique suivant :



CHAPITRE 4. STATISTIQUE DESCRIPTIVE BI-VARIÉE

Deuxième partie

Probabilités

CHAPITRE 5

MODÉLISATION DES PHÉNOMÈNES ALÉATOIRES

5.1 INTRODUCTION

Une *expérience* (ou *phénomène*) aléatoire consiste en une expérience pour laquelle toutes les issues possibles sont connues, mais où interviennent de nombreux facteurs, dont nous ne connaissons ou maîtrisons qu'une petite partie. Dans ce cas, l'issue n'est pas prévisible avec certitude. La *théorie des probabilités* consiste en l'étude de ces expériences aléatoires.

Citons quelques exemples : le résultat d'un jeu de hasard (pile ou face, jet de dé, roulette etc.) ; durée de vie d'un atome radioactif, d'un individu, d'une ampoule ; les instants de passage d'un bus à un arrêt donné ; la promenade d'un ivrogne dans la rue ; la trajectoire d'une poussière à la surface de l'eau etc.

Les applications de la théorie des probabilités sont nombreuses : base de la statistique, outil puissant en finance, dans les assurances, théorie des jeux. Elle permet également de modéliser de nombreux phénomènes complexes en biologie, médecine, sciences humaines, climatologie. Elle s'est aussi révélée utile dans de nombreux domaines des mathématiques pures. Mais surtout, elle a acquis une place importante au sein des mathématiques en tant que discipline à part entière, de part son intérêt intrinsèque.

Historiquement, les jeux des hasards sont présents en Égypte, en Grèce et à Rome dès l'Antiquité. Il est cependant intéressant de constater qu'un traitement systématique n'est apparu qu'au XVI^e siècle dans le livre *Liber de Ludo Alea* de Gerolamo Cardano (1501-1576). La véritable étincelle se trouve dans la correspondance entre Blaise Pascal (1623-1662) et Pierre de Fermat (~1605-1665), au sujet de problèmes posés par le chevalier de Méré. Encouragé par Pascal, Christian Huygens (1629-1695) publie *De ratiocinis in ludo aleae* (raisonnements sur les jeux de dés) en 1657. Ce livre est le premier ouvrage important sur les probabilités. Il y définit la notion d'espérance et y développe plusieurs problèmes de partages de gains lors de jeux ou de tirages dans des urnes. Deux ouvrages fondateurs sont également à noter : *Ars Conjectandi* de Jacques Bernoulli (1654-1705) qui définit la notion de variable aléatoire et donne la première version de la loi des grands nombres, et *The Doctrine of Chance* d'Abraham de Moivre (1668-1754) qui généralise l'usage de la combinatoire. On mentionnera également Pierre-Simon de Laplace (1749-1827), Leonhard Euler (1707-1783) et Johann Carl Friedrich Gauss (1777-1855).

La théorie des probabilités classique ne prend réellement son essor qu'avec les notions de mesure et d'ensembles mesurables qu'Émile Borel (1871-1956) introduit en 1897. Cette notion de mesure

est complétée par Henri Léon Lebesgue (1875-1941) et sa théorie de l'intégration. La première version moderne du théorème central limite est donnée par Alexandre Liapounov en 1901 et la première preuve du théorème moderne est due à Paul Lévy en 1910. Il faudra attendre 1933 pour que la théorie des probabilités sorte d'un ensemble de méthodes et d'exemples divers et devienne une véritable théorie, axiomatisée par Andreï Nikolaïevitch Kolmogorov (1903-1987).

5.2 L'ESPACE PROBABILISÉ $(\Omega, \mathcal{A}, \mathbb{P})$

Le but de la théorie des probabilités est de fournir un modèle mathématique pour décrire les expériences aléatoires. Sous sa forme moderne, la formulation de cette théorie contient trois ingrédients : l'*espace des états*, les *événements*, et la *loi de probabilité* ou simplement la *probabilité*. Dans toute la suite, nous considérons une expérience aléatoire que nous cherchons à modéliser.

5.2.1 ESPACE DES ÉTATS

Définition. L'*espace des états* appelé aussi *univers*, noté Ω , est l'ensemble des résultats possibles de l'expérience.

Exemple 5.1. Voici quelques exemples de choix d'univers.

1. Lancer d'une pièce de monnaie. $\Omega = \{P, F\}$.
2. Deux lancers successifs d'une même pièce de monnaie. $\Omega = \{PP, PF, FP, FF\}$.
3. Lancer d'un dé. $\Omega = \{1, 2, 3, 4, 5, 6\}$.
4. Deux lancers successifs d'un même dé, et on s'intéresse à la somme des nombres obtenus.
Dans ce cas, il y a trois choix raisonnables :

$$\begin{aligned}\Omega_1 &= \{(i, j) : i \in \{1, \dots, 6\}, j \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^2, \\ \Omega_2 &= \{2, 3, 4, \dots, 12\}, \\ \Omega_3 &= \{\{i, j\} : i \in \{1, \dots, 6\}, j \in \{1, \dots, 6\}, i \leq j\}.\end{aligned}$$

5. Lancer d'un même dé indéfiniment.

$$\Omega = \{(u_n)_{n \geq 1} : \forall n \in \mathbb{N}^*, u_n \in \{1, \dots, 6\}\} = \{1, \dots, 6\}^{\mathbb{N}^*}.$$

6. Durée de vie d'un individu. $\Omega = \{x \in \mathbb{R}^+ : 0 \leq x \leq 120\}$.
7. Promenade d'un ivrogne dans une rue (un pas en avant, un pas en arrière).

$$\Omega = \{(u_n)_{n \geq 1} : \forall n \in \mathbb{N}^*, u_n \in \{-1, 1\}\} = \{-1, 1\}^{\mathbb{N}^*}.$$

8. Trajectoire d'une poussière à la surface de l'eau pendant un intervalle de temps $[0, T]$.
 $\Omega = C([0, T]; \mathbb{R}^2)$.

5.2.2 ÉVÉNEMENTS

Définition 5.1. Un *événement* est une propriété dont on peut dire si elle est réalisée ou non, une fois l'issue de l'expérience connue. Un événement correspond un sous-ensemble A de l'univers Ω .

Un singleton, c'est-à-dire un événement réduit à un seul élément de Ω , est appelé un *événement élémentaire*, sinon on parle d'*événement composite*.

Exemple 5.2. On reprend la numérotation de l'exemple 5.1. Voici quelques exemples d'événements écrits d'abord en mots, puis en tant que sous-ensembles de l'espace des états Ω .

2. “Le premier jet donne pile” est le sous-ensemble $\{PP, PF\}$ de Ω .
4. “La somme des résultats obtenus est égale à 4” est le sous-ensemble $\{(1, 3), (2, 2), (3, 1)\}$ de Ω_1 , au sous-ensemble $\{4\}$ de Ω_2 , et au sous-ensemble $\{1, 3\}, \{2, 2\}\}$ de Ω_3 .
5. “Le premier 1 est obtenu au N -ième lancer” est le sous-ensemble

$$\{(u_n)_{n \geq 1} \in \Omega : u_1 \geq 2, \dots, u_{N-1} \geq 2, u_N = 1\}.$$

6. “L'individu atteint au moins 50 ans” est le sous-ensemble :

$$\{x \in \mathbb{R}^+ : 50 \leq x \leq 120\}.$$

7. “L'ivrogne avance au N -ième pas” est le sous-ensemble :

$$\{(u_n)_{n \geq 1} \in \Omega : u_N = 1\}.$$

Remarque. Les événements, qui sont par définition des sous-ensembles de l'univers, sont en général décrits à l'aide de phrases dans un premier temps. En effet, on commence par se poser une question liée à une expérience aléatoire, puis on introduit un modèle probabiliste pour y répondre. Par exemple, on cherche la probabilité que la somme de deux dés lancés au hasard soit égale à 4 ; l'événement considéré est alors “la somme des dés est égale à 4”.

Une fois fixé le choix de l'univers, un événement correspond à un *unique* sous-ensemble de ce dernier. Comme il n'y a pas forcément unicité du modèle et qu'alors les événements peuvent s'écrire en termes de sous-ensembles sous des formes différentes, la phrase qui décrit un événement permet de se comprendre, quel que soit le modèle choisi, voir par exemple les exemples 5.1 et 5.2 numéro 4. Remarquons aussi que, étant donné un sous-ensemble d'un univers, il est souvent possible de le décrire par différentes phrases, qui représentent toutes le même événement. Par exemple l'événement $\{PF, FF\}$ de l'exemple 5.2 numéro 2 peut se traduire par “le premier jet donne pile” ou “le premier jet ne donne pas face”.

Puisque les événements sont des sous-ensembles, on peut effectuer les opérations habituelles, avec la correspondance suivante entre les terminologies ensembliste et probabiliste.

Notation	Terminologie ensembliste	Terminologie probabiliste
Ω	ensemble entier	espace des états, événement certain
ω	élément de Ω	événement élémentaire
A	sous-ensemble de Ω	événement
$\omega \in A$	ω appartient à A	A est réalisé si ω est le résultat de l'expérience
$A \subset B$	A est inclus dans B	si A est réalisé alors B aussi
$A \cup B$	réunion de A et B	l'événement “ A ou B ” (ou non exclusif!)
$A \cap B$	intersection de A et B	l'événement “ A et B ”
A^c	complémentaire de A	l'événement contraire de A
\emptyset	ensemble vide	événement impossible
$A \cap B = \emptyset$	A et B sont disjoints	A et B sont incompatibles

Exemple. Deux lancers successifs d'une même pièce de monnaie. Soient $A = \{PP\}$, $B = \{PF\}$, $C = \{FP, FF\}$. Alors,

- $A \cup B = \{PP, PF\} = C^c$, est l'événement “le premier jet donne pile” ;
- $A \cap B = \emptyset$, est l'événement impossible, A et B sont incompatibles.

Propriété 5.1. *Les opérations sur les événements satisfont aux règles suivantes. Pour tout événements A , B , C , on a*

- *commutativité* : $A \cup B = B \cup A$;
- *associativité* : $(A \cup B) \cup C = A \cup (B \cup C)$;
- *distributivité* : $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$;
- *lois de De Morgan* : $(A \cup B)^c = A^c \cap B^c$, et $(A \cap B)^c = (A^c \cup B^c)$.

Un événement est donc un sous-ensemble des parties de Ω , $\mathcal{P}(\Omega)$. Comme ensemble des événements il semblerait donc naturel de prendre tout $\mathcal{P}(\Omega)$. C'est ce que l'on va faire lorsque Ω est fini ou dénombrable. Cependant, il existe des exemples d'univers où il est impossible d'associer à chaque événement une probabilité de façon cohérente. Dans ces cas-là, il est donc nécessaire de se restreindre à un sous-ensemble strict de $\mathcal{P}(\Omega)$ contenant les événements “intéressants”. Ils devront satisfaire à quelques propriétés naturelles : ils doivent former une tribu. Nous ne donnons pas cette définition car elle est hors-programme.

Finalement, nous notons \mathcal{A} l'ensemble des événements. Si Ω est fini ou dénombrable, $\mathcal{A} = \mathcal{P}(\Omega)$; sinon \mathcal{A} peut-être strictement inclus dans $\mathcal{P}(\Omega)$.

5.2.3 PROBABILITÉ

Nous souhaitons maintenant associer à chacun des événements une *probabilité*, qui mesure la vraisemblance que l'on accorde a priori à l'événement avant la réalisation de l'expérience. C'est une des données du modèle, que l'on peut comprendre intuitivement de différentes manières, en voici deux.

Approche utilisant les symétries. On considère un dé non-pipé. Il est alors naturel de supposer que chacune des issues possibles ait la même probabilité égale à 1/6. Il faut cependant être prudent avec cette approche. En effet, supposons que nous souhaitions déterminer la probabilité du sexe

d'un nouveau né. Il n'y a aucune raison de penser qu'il y a plus de chances d'avoir un garçon ou une fille, de sorte qu'il est naturel d'associer une probabilité 1/2 à chacun des événements élémentaires. Cependant, les statistiques montrent que la proportion de garçons nouvellement né est de 51,2% (INED, France métropolitaine).

Approche fréquentiste. On suppose qu'une expérience d'univers Ω est exécutée plusieurs fois sous les mêmes conditions. Pour chaque événement A de Ω , on définit $n_N(A)$ comme le nombre de fois où l'événement A survient lors des N premières répétitions de l'expérience. Alors la *probabilité de l'événement* A , notée $\mathbb{P}(A)$, est définie comme la limite, dans un sens à préciser, du quotient $n_N(A)/N$.

Cela veut dire que $\mathbb{P}(A)$ est définie comme la limite du pourcentage du nombre de fois où A survient par rapport au nombre total des répétitions. C'est donc la fréquence limite de A . Bien que cette définition soit intuitivement commode, elle présente un sérieux inconvénient. En effet, il faut justifier de l'existence de la limite, ce qui est difficile a priori.

Il est plus raisonnable d'admettre que les probabilités satisfont à un ensemble d'axiomes simples et intuitivement acceptables, pour ensuite démontrer qu'une telle fréquence limite existe dans un certain sens (il s'agit de la *loi des grands nombres*).

Définition. Étant donnés un espace d'états Ω et \mathcal{A} l'ensemble des événements, une *probabilité* \mathbb{P} sur (Ω, \mathcal{A}) , est une application de \mathcal{A} dans $[0, 1]$, possédant les propriétés suivantes.

1. L'événement certain est de probabilité 1 : $\mathbb{P}(\Omega) = 1$.
2. *Axiome de σ -additivité* : pour toute famille dénombrable $(A_n)_{n \geq 0}$ d'événements de \mathcal{A} , deux-à-deux disjoints, on a

$$\mathbb{P}\left(\bigcup_{n \geq 0} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n).$$

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est alors appelé un *espace probabilisé*.

On a les conséquences immédiates suivantes.

Proposition 5.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et soient deux événements $A \in \mathcal{A}$, $B \in \mathcal{A}$.

1. Additivité. Si A et B sont disjoints, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. En particulier, $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, et $\mathbb{P}(\emptyset) = 0$.
2. Si $A \subset B$, alors : $\mathbb{P}(A) \leq \mathbb{P}(B)$.
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
4. Plus généralement, on a la formule de Poincaré : Soit $(A_n)_{n=1}^N$ une famille d'événements de \mathcal{A} , alors :

$$\mathbb{P}\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N (-1)^{n-1} \sum_{\substack{J \subset \{1, \dots, N\} \\ \text{card}(J) = n}} \mathbb{P}\left(\bigcap_{k \in J} A_k\right).$$

CHAPITRE 5. MODÉLISATION DES PHÉNOMÈNES ALÉATOIRES

CHAPITRE 6

CONSTRUCTION D'ESPACES PROBABILISÉS

On suppose dans la suite que l'univers Ω est fini ou dénombrable. L'ensemble des événements est donc $\mathcal{P}(\Omega)$.

6.1 CARACTÉRISATION D'UNE PROBABILITÉ

Avec la définition de probabilité du chapitre précédent, il est n'est pas si aisés de définir concrètement une probabilité. Dans le cas où Ω est fini ou dénombrable, on dispose de la caractérisation suivante.

Proposition 6.1. *Une probabilité sur $(\Omega, \mathcal{P}(\Omega))$ est caractérisée par sa valeur sur les singletons $\{\omega\}$, pour tout $\omega \in \Omega$. Réciproquement, à toute famille $(p_\omega)_{\omega \in \Omega}$ telle que :*

1. *pour tout $\omega \in \Omega$, $0 \leq p_\omega \leq 1$,*
2. $\sum_{\omega \in \Omega} p_\omega = 1$,

on peut associer une unique probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$ définie par : $\mathbb{P}(\{\omega\}) = p_\omega$. On étend ensuite \mathbb{P} à $\mathcal{P}(\Omega)$ par σ -additivité : pour tout $A \in \mathcal{P}(\Omega)$,

$$\mathbb{P}(A) = \sum_{\omega \in A} p_\omega.$$

6.2 CAS OÙ L'UNIVERS EST FINI

Un exemple particulier et très classique de choix de probabilité nous amène à toutes les questions de dénombrements.

Définition. Une probabilité \mathbb{P} sur (Ω, \mathcal{A}) est dite *uniforme*, si $\mathbb{P}(\{\omega\})$ ne dépend pas de $\omega \in \Omega$. On dit alors que l'on est en situation d'*équiprobabilité*.

Corollaire 6.1. *Dans ce cas, pour tout $\omega \in \Omega$, $\mathbb{P}(\{\omega\}) = \frac{1}{\text{card}(\Omega)}$, et, pour tout événement $A \in \mathcal{P}(\Omega)$, on a :*

$$\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}.$$

Exemple. Reprenons la question 4 des exemples 5.1 et 5.2 du Chapitre 5 et calculons la probabilité de l'événement A “la somme des dés est égale à 4”.

Supposons que l'on ait choisi l'espace Ω_1 . Alors, on est en situation d'équiprobabilité et la probabilité \mathbb{P}_1 sur Ω_1 est uniforme, de sorte que pour tout $(i, j) \in \{1, \dots, 6\}^2$:

$$\mathbb{P}_1[\{(i, j)\}] = \frac{1}{36}.$$

Ainsi, $\mathbb{P}_1(A) = \mathbb{P}_1[\{(1, 3), (2, 2), (3, 1)\}] = \frac{\text{card}(A)}{\text{card}(\Omega_1)} = \frac{3}{36}$.

Supposons maintenant que l'on ait choisi l'espace Ω_2 . Alors, on n'est plus en situation d'équiprobabilité. Au vu des conditions de l'expérience, on définit \mathbb{P}_2 ainsi :

$$\begin{aligned}\mathbb{P}_2(\{2\}) &= \mathbb{P}_2(\{12\}) = \frac{1}{36}, & \mathbb{P}_2(\{3\}) &= \mathbb{P}_2(\{11\}) = \frac{1}{18}, & \mathbb{P}_2(\{4\}) &= \mathbb{P}_2(\{10\}) = \frac{1}{12}, \\ \mathbb{P}_2(\{5\}) &= \mathbb{P}_2(\{9\}) = \frac{1}{9}, & \mathbb{P}_2(\{6\}) &= \mathbb{P}_2(\{8\}) = \frac{5}{36}, & \mathbb{P}_2(\{7\}) &= \frac{1}{6}.\end{aligned}$$

Ainsi, $\mathbb{P}_2(A) = \mathbb{P}_2(\{4\}) = \frac{1}{12}$ mais $\frac{\text{card}(A)}{\text{card}(\Omega_2)} = \frac{1}{11}$, d'où $\mathbb{P}_2(A) \neq \frac{\text{card}(A)}{\text{card}(\Omega_2)}$.

Remarquer que si l'on avait choisi l'univers Ω_3 , on ne serait pas non plus en situation d'équiprobabilité.

Cet exemple montre qu'il est très important de spécifier le choix d'univers et de probabilité. Bien que les résultats finaux ne changent pas, les raisonnements pour y arriver sont différents et doivent être explicités.

Lorsque l'espace des états est fini, les calculs de probabilités se ramènent essentiellement à des problèmes de dénombrement, sujet de la section suivante.

6.2.1 DÉNOMBREMENT, MODÈLE D'URNE

Définition. Une *population* de taille N est un ensemble $\mathcal{S} = \{s_1, \dots, s_N\}$ de N éléments. Les éléments de \mathcal{S} sont les *individus* de la population \mathcal{S} . La *taille* N est le nombre d'éléments de \mathcal{S} .

TIRAGES ORDONNÉS

Un *échantillon* de taille r est un r -uplet $(s_{i_1}, \dots, s_{i_r})$ d'éléments de \mathcal{S} . Deux procédures sont possibles.

- *Le tirage avec remise.* Dans ce cas, chaque élément de l'ensemble peut être choisi à plusieurs reprises. On parle alors d'*échantillon de taille r avec répétitions*. Soit Ω_1 l'ensemble de ces échantillons, alors $\Omega_1 = \{s_1, \dots, s_N\}^r$, et :

$$\text{card}(\Omega_1) = N^r.$$

- *Le tirage sans remise.* Dans ce cas, chaque élément de l'ensemble peut être choisi au plus une fois. On parle alors d'*échantillon de taille r sans répétition*, ou d'*arrangement des éléments de \mathcal{S} pris r à r* . Naturellement, on impose les conditions supplémentaires $r \leq N$, et $\forall j \neq k, i_j \neq i_k$. Soit Ω_2 l'ensemble de ces échantillons, on a alors :

$$\text{card}(\Omega_2) = N(N - 1) \cdots (N - r + 1) = \frac{N!}{(N - r)!}$$

Ce nombre a deux notations usuelles : A_N^r ou $(N)_r$ (*symbole de Pochhammer*).

Exemple. On considère une population de taille N et un échantillon aléatoire de taille r avec répétition. On choisit alors comme univers Ω_1 que l'on munit de la probabilité uniforme, notée \mathbb{P} . On s'intéresse à l'événement A “aucun individu n'a été choisi plus d'une fois” qui est la même chose que “tous les individus sont distincts”. Alors on a, $A = \Omega_2$ et :

$$\mathbb{P}(A) = \frac{\text{card}(\Omega_2)}{\text{card}(\Omega_1)} = \frac{A_N^r}{N^r}.$$

Donnons quelques applications de ce résultat.

1. On jette un dé six fois de suite. Alors la probabilité d'obtenir six nombres distincts est $\frac{6!}{6^6} \sim 0,015$.
2. Supposons que dans une ville, il y ait sept accidents par semaine. Alors la probabilité d'avoir exactement un accident chaque jour de la semaine est $\frac{7!}{7^7} \sim 0,00612$.

TIRAGES NON ORDONNÉS

Une *sous-population* de taille r est un sous-ensemble $\{s_{i_1}, \dots, s_{i_r}\}$ d'éléments de \mathcal{S} . De manière similaire aux tirages ordonnés, deux procédures sont possibles.

- *Le tirage sans remise.* On parle alors de *sous-population de taille r sans répétition*, ou de *combinaison de r éléments*. On impose à nouveau les conditions supplémentaires, $r \leq N$, et $\forall j \neq k, i_j \neq i_k$. Soit Ω_3 l'ensemble de ces populations, on a alors :

$$\text{card}(\Omega_3) = \frac{N!}{(N-r)!r!}.$$

Ce nombre, appelé *coefficient binomial*, a deux notations usuelles : C_N^r (notation qui n'est plus tellement en vigueur) ou $\binom{N}{r}$.

Démonstration. Chacun des sous-ensembles à r éléments fournit $r!$ échantillons de taille r sans répétition, de sorte que $\text{card}(\Omega_2) = r! \text{card}(\Omega_3)$. \square

Exemple 6.1.

1. On appelle main de poker l'ensemble des 5 cartes que chacun des quatre joueurs reçoit lors de la distribution d'un jeu qui en contient 32. Alors il existe $\binom{32}{5}$ mains différentes. Soit A l'événement “les hauteurs des 5 cartes sont différentes”, calculons $\text{card}(A)$. On peut choisir ces hauteurs de $\binom{8}{5}$ manières différentes. Il faut ensuite choisir la couleur (trèfle, carreau, cœur, pique) de chacune des hauteurs. Ainsi :

$$\text{card}(A) = \binom{8}{5} 4^5.$$

Étant donné que toutes les mains sont supposées équiprobables, la probabilité d'obtenir une main dont les 5 cartes ont une hauteur différente est :

$$\mathbb{P}(A) = \frac{\binom{8}{5} 4^5}{\binom{32}{5}}.$$

2. Une urne contient N_b boules blanches et N_n boules noires. Posons $N = N_b + N_n$. On tire r boules avec remise dans l'urne, il y a alors N^r tirages possibles. Soit A_k l'événement “on a tiré exactement k boules blanches”, calculons $\text{card}(A_k)$. L'événement A_k est réalisé lorsque l'issu est constituée de k boules blanches et $r - k$ boules noires. Il y a $\binom{r}{k}$ façons de choisir la position des boules blanches, la position des boules noires est ensuite fixée. Pour chacune des positions de boule blanche, il y a ensuite N_b choix de boules blanches possibles, et pour chacune des positions de boule noire, il y a N_n choix possibles, ainsi :

$$\text{card}(A_k) = \binom{r}{k} N_b^k N_n^{r-k}.$$

Étant donné que tous les tirages sont supposés équiprobables, la probabilité d'obtenir exactement k boules blanches lors d'un tirage de r boules avec remise est :

$$\mathbb{P}(A_k) = \frac{\binom{r}{k} N_b^k N_n^{r-k}}{N^r} = \binom{r}{k} \left(\frac{N_b}{N}\right)^k \left(\frac{N_n}{N}\right)^{r-k}.$$

Ceci est un exemple de la *loi binomiale*, que nous reverrons plus tard.

3. Soit \mathcal{S} une population de taille N (*ex.* des étudiants), que l'on range en deux catégories a et b incompatibles (*ex.* filles et garçons), de tailles respectives N_a et $N_b = N - N_a$. On choisit “au hasard” une sous-population de taille r sans répétition, il y a alors $\binom{N}{r}$ choix possibles. Soit A_k l'événement “on a choisi exactement k individus de la catégorie a ”, calculons $\text{card}(A_k)$. L'événement est réalisé lorsque l'issu est constituée de k individus de la catégorie a et $r - k$ de la catégorie b . Il y a $\binom{N_a}{k}$ façons de choisir les k individus de la catégorie a et pour chacune il y a $\binom{N-N_a}{r-k}$ façons de choisir les individus restants dans la catégorie b , ainsi :

$$\text{card}(A_k) = \binom{N_a}{k} \binom{N - N_a}{r - k}.$$

Remarquer que pour que ceci ait un sens, il faut que $0 \leq k \leq \min\{r, N_a\}$. Étant donné que tous les tirages sont supposés équiprobables, la probabilité d'obtenir k individus de la catégorie a lors de ce tirage est :

$$\mathbb{P}(A_k) = \frac{\binom{N_a}{k} \binom{N_b}{r-k}}{\binom{N}{r}}.$$

Ceci est un exemple de la *loi hypergéométrique*.

Remarque. Supposons que $N_a = N_a(N)$ soit une fonction de N et que le nombre total de boules tende vers l'infini, de sorte que la proportion $\frac{N_a}{N}$ tende vers p (et donc que $\frac{N_b}{N}$ tende vers $1 - p$), avec $0 < p < 1$. Ainsi, N_a et N_b tendent vers $+\infty$ avec N . Fixons $r \geq 0$ et k compris entre 0 et r . Alors, pour N assez grand, on a $N_a \geq k$, $N_b \geq r - k$ et $\mathbb{P}(A_k)$

peut s'écrire :

$$\begin{aligned}
 \mathbb{P}(A_k) &= \frac{N_a(N_a - 1) \dots (N_a - k + 1)}{k!} \frac{N_b(N_b - 1) \dots (N_b - r + k + 1)}{(r - k)!} \frac{r!}{N(N - 1) \dots (N - r + 1)} \\
 &= \frac{r!}{k!(r - k)!} \frac{N_a(N_a - 1) \dots (N_a - k + 1) N_b(N_b - 1) \dots (N_b - r + k + 1)}{N(N - 1) \dots (N - r + 1)} \\
 &= \frac{r!}{k!(r - k)!} \frac{N^k \frac{N_a}{N} \left(\frac{N_a}{N} - \frac{1}{N}\right) \dots \left(\frac{N_a}{N} - \frac{k-1}{N}\right) N^{r-k} \frac{N_b}{N} \left(\frac{N_b}{N} - \frac{1}{N}\right) \dots \left(\frac{N_b}{N} - \frac{r-k-1}{N}\right)}{N^r 1\left(1 - \frac{1}{N}\right) \dots \left(1 - \frac{r-1}{N}\right)} \\
 &= \binom{r}{k} \frac{\frac{N_a}{N} \left(\frac{N_a}{N} - \frac{1}{N}\right) \dots \left(\frac{N_a}{N} - \frac{k-1}{N}\right) \frac{N_b}{N} \left(\frac{N_b}{N} - \frac{1}{N}\right) \dots \left(\frac{N_b}{N} - \frac{r-k-1}{N}\right)}{1\left(1 - \frac{1}{N}\right) \dots \left(1 - \frac{r-1}{N}\right)} \\
 &\xrightarrow[N \rightarrow +\infty]{} \frac{\binom{r}{k} p^k (1-p)^{r-k}}{1}.
 \end{aligned}$$

Ainsi, $\mathbb{P}(A_k)$ tend vers $\binom{r}{k} p^k (1-p)^{r-k}$. On a donc obtenu la loi binomiale comme limite de lois hypergéométriques. Ce résultat est intuitif, car lorsque le nombre de boules est très grand, que le tirage s'effectue avec ou sans remise ne change pas grand chose : on a peu de chance de tirer deux fois la même boule.

- *Partitionnement*

Soient r_1, \dots, r_k des entiers positifs (éventuellement nuls) tels que, $r_1 + \dots + r_k = N$. Le nombre de façons de répartir N objets dans k familles de sorte que la i -ième famille contienne r_i éléments est égal à :

$$\frac{N!}{r_1! \dots r_k!}.$$

Ce nombre se note $\binom{N}{r_1 \dots r_k}$ et s'appelle *coefficient multinomial*.

Démonstration. Pour remplir la première famille, il faut choisir r_1 objets parmi N , ce qui peut se faire de $\binom{N}{r_1}$ façons. Pour remplir la seconde famille, il faut choisir r_2 objets parmi $N - r_1$, soit $\binom{N-r_1}{r_2}$. En continuant ainsi, on obtient que le nombre de telles répartitions est de :

$$\binom{N}{r_1} \binom{N - r_1}{r_2} \dots \binom{N - r_1 - \dots - r_{k-1}}{r_k} = \frac{N!}{r_1! \dots r_k!}.$$

□

Exemple. Le nombre d'anagrammes du mot CHERCHER est $\frac{8!}{2!2!2!2!}$.

- *Le tirage avec remise.* On parle alors de *sous-population de taille r avec répétitions*. Soit Ω_4 l'ensemble de ces populations, on a alors :

$$\text{card}(\Omega_4) = \binom{N+r-1}{N-1} = \binom{N+r-1}{r}.$$

Démonstration. Ce problème revient à placer r boules indistinguables dans N urnes. En effet, le nombre de boules dans la i -ième urne compte le nombre de répétitions de l'individu i lors du tirage. Représentons les r boules par r étoiles alignées, avec une cloison à chacune des extrémités. Par exemple, lorsque $r = 7$,

| * * * * * * *

Répartir les r boules dans N urnes revient à rajouter $N - 1$ cloisons formant les N urnes. Par exemple, lorsque $r = 7$, $N = 3$,

$$| * * | * * * * * |,$$

représente le tirage : $s_1, s_1, s_3, s_3, s_3, s_3, s_3$. Ainsi, ce problème revient à placer $N - 1$ cloisons sur $N + r - 1$ positions, les positions restantes étant occupées par des *.

□

Exemple. Soient $r \in \mathbb{N}^*$ et $n \in \mathbb{N}^*$. On cherche à compter le nombre de suites d'entiers naturels r_1, \dots, r_n , telles que :

$$r_1 + \dots + r_n = r.$$

Ce problème revient à placer r boules indistinguables dans n urnes, où le nombre de boules dans la i -ième urne représente r_i . Ainsi, le nombre de ces suites est $\binom{n+r-1}{n-1}$. Par exemple, si $r = 10$, $n = 3$,

$$| * * | * * * * * | * * * |$$

représente la partition $(2, 5, 3)$ de 10. Remarquer que ces suites sont naturellement ordonnées de sorte que l'on distingue $(2, 5, 3)$ de $(5, 3, 2)$.

6.3 ESPACE DES ÉTATS INFINI DÉNOMBRABLE

Lorsque Ω est infini dénombrable, on procède de la même manière que dans le cas fini : on prend $\mathcal{A} = \mathcal{P}(\Omega)$, et on associe à chaque événement élémentaire $\omega \in \Omega$, sa probabilité, $\mathbb{P}(\{\omega\}) = p_\omega \in [0, 1]$, avec :

$$\sum_{\omega \in \Omega} p_\omega = 1.$$

Exemple. On jette une pièce de monnaie jusqu'à l'obtention du premier pile. On peut choisir $\Omega = \mathbb{N}^* \cup \{\infty\}$ où, pour tout $k \in \mathbb{N}^*$, $\{k\}$ représente l'événement “le premier pile est obtenu au k -ième jet”, et $\{\infty\}$ représente l'événement “pile ne sort jamais”. Si la pièce est équilibrée, on aura, pour tout $k \in \mathbb{N}^*$:

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

Comme \mathbb{N}^* et $\{\infty\}$ sont disjoints, $1 = \mathbb{P}[\Omega] = \mathbb{P}[\{\infty\}] + \mathbb{P}[\bigcup_{k \in \mathbb{N}^*} \{k\}]$. Ainsi, la probabilité que pile ne sorte jamais est donnée par :

$$\begin{aligned} \mathbb{P}(\{\infty\}) &= 1 - \mathbb{P}\left[\bigcup_{k \in \mathbb{N}^*} \{k\}\right] \\ &= 1 - \sum_{k=1}^{+\infty} \mathbb{P}(\{k\}), \text{ car les événements sont disjoints} \\ &= 1 - \sum_{k=1}^{+\infty} \frac{1}{2^k} = 0. \end{aligned}$$

La probabilité que le premier pile sorte après un nombre pair de lancers est :

$$\mathbb{P}(\{2, 4, 6, \dots\}) = \sum_{k=1}^{+\infty} \mathbb{P}(\{2k\}) = \sum_{k=1}^{+\infty} \frac{1}{2^{2k}} = \frac{1}{3}.$$

CHAPITRE 7

CONDITIONNEMENT ET INDÉPENDANCE

7.1 PROBABILITÉ CONDITIONNELLE

7.1.1 DÉFINITION

Motivons la définition de probabilité conditionnelle sur un exemple.

Soit Ω une population partitionnée en $\Omega = S \cup S^c$, où S représente l'ensemble des individus fumeurs. Soit F l'ensemble des femmes, de sorte que $F \cup F^c$ représente une autre partition de Ω . On suppose que l'on choisit un individu "au hasard", de sorte que l'on munit Ω de la probabilité uniforme, notée \mathbb{P} . Ainsi, la probabilité que l'individu choisi soit fumeur est :

$$\mathbb{P}(S) = \frac{\text{card}(S)}{\text{card}(\Omega)}.$$

Si maintenant on choisit un individu avec l'information supplémentaire qu'il s'agit d'une femme, tout se passe comme si l'univers considéré est F , et que l'on a partitionné F (et non plus Ω) en $S \cap F$ et $S^c \cap F$. Ainsi la probabilité que l'individu choisi soit fumeur, étant donné l'information que c'est une femme, est égale à :

$$\frac{\text{card}(S \cap F)}{\text{card}(F)} = \frac{\text{card}(S \cap F)/\text{card}(\Omega)}{\text{card}(F)/\text{card}(\Omega)},$$

quantité encore égale, avec les notations précédentes, à

$$\frac{\mathbb{P}(S \cap F)}{\mathbb{P}(F)}.$$

Définition. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et B un événement tel que $\mathbb{P}(B) > 0$. Pour tout $A \in \mathcal{A}$, on définit la *probabilité conditionnelle de A sachant B* , notée $\mathbb{P}_B(A)$, par :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{7.1}$$

Lemme 7.1. *Sous les hypothèses de la définition ci-dessus, l'application \mathbb{P}_B est une probabilité sur (Ω, \mathcal{A}) , ainsi $(\Omega, \mathcal{A}, \mathbb{P}_B)$ est un espace probabilisé.*

Démonstration. Il faut montrer que \mathbb{P}_B satisfait aux trois axiomes caractérisant une probabilité.

- Par hypothèse, $\mathbb{P}(B) > 0$. De plus, pour tout $A \in \mathcal{A}$, on a $A \cap B \subset B$, d'où $0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B)$, et donc :

$$0 \leq \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \leq 1.$$

- Comme $\Omega \cap B = B$, on a $\mathbb{P}_B(\Omega) = 1$.
- Soit $(A_n)_{n \geq 1}$ une famille dénombrable d'événements deux à deux disjoints. Alors on a l'égalité $(\bigcup_{n \geq 1} A_n) \cap B = \bigcup_{n \geq 1} (A_n \cap B)$. Comme les événements $(A_n)_{n \geq 1}$ sont deux à deux disjoints, il en est de même pour les événements $(A_n \cap B)_{n \geq 1}$. Ainsi la σ -additivité de \mathbb{P} implique $\mathbb{P}\left(\bigcup_{n \geq 1} (A_n \cap B)\right) = \sum_{n \geq 1} \mathbb{P}(A_n \cap B)$, et on conclut :

$$\mathbb{P}_B\left(\bigcup_{n \geq 1} A_n\right) = \frac{\mathbb{P}\left(\left(\bigcup_{n \geq 1} A_n\right) \cap B\right)}{\mathbb{P}(B)} = \sum_{n \geq 1} \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \sum_{n \geq 1} \mathbb{P}_B(A_n).$$

□

Remarque.

- On utilise aussi la notation $\mathbb{P}(A|B)$. La notation $\mathbb{P}_B(A)$ met en valeur le fait que \mathbb{P}_B soit une probabilité sur (Ω, \mathcal{A}) .
- Comme $(\Omega, \mathcal{A}, \mathbb{P}_B)$ est un espace probabilisé, \mathbb{P}_B satisfait à toutes les propriétés de la proposition 5.1.
- Si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, on peut réécrire l'équation (7.1) sous la forme :

$$\mathbb{P}(A \cap B) = \mathbb{P}_A(B) \mathbb{P}(A) = \mathbb{P}_B(A) \mathbb{P}(B).$$

- De manière plus générale, si $(A_k)_{1 \leq k \leq n+1}$ sont des événements qui satisfont à $\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) > 0$, alors :

$$\mathbb{P}\left(\bigcap_{k=1}^{n+1} A_k\right) = \mathbb{P}(A_1) \mathbb{P}_{A_1}(A_2) \mathbb{P}_{A_1 \cap A_2}(A_3) \cdots \mathbb{P}_{A_1 \cap \cdots \cap A_n}(A_{n+1}).$$

Exemple 7.1. On considère deux lancers successifs d'un même dé. Sachant que le premier jet donne 3, on souhaite calculer la probabilité que la somme soit strictement supérieure à 6.

Supposons que l'on ait choisi l'univers $\Omega = \{(i, j) : i \in \{1, \dots, 6\}, j \in \{1, \dots, 6\}\}$, que l'on munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. On est alors en situation d'équiprobabilité, et on munit Ω de la probabilité uniforme, notée \mathbb{P} . Ainsi, pour tout événement A de \mathcal{A} , $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{1}{36}$.

Soit B l'événement “le premier jet donne 3”; B est le sous-ensemble $\{(3, j) : j \in \{1, \dots, 6\}\}$ de \mathcal{A} , son cardinal est égal à 6, d'où $\mathbb{P}(B) = \frac{6}{36} = \frac{1}{6} > 0$. La probabilité $\mathbb{P}(B)$ étant strictement positive, on peut conditionner par l'événement B .

Soit A l'événement “la somme des dés est strictement supérieure à 6”; A est le sous-ensemble $\{(i, j) \in \Omega : i + j > 6\}$ de \mathcal{A} , et $A \cap B = \{(3, 4), (3, 5), (3, 6)\}$, d'où : $\mathbb{P}(A \cap B) = \frac{3}{36} = \frac{1}{12}$. On conclut que :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{1}{2}.$$

7.1.2 FORMULE DES PROBABILITÉS TOTALES ET FORMULE DE BAYES

Définition. Soit I une partie finie ou dénombrable de \mathbb{N} . Une famille $(B_i)_{i \in I}$ d'événements de Ω forme un *système complet d'événements* de Ω , si

$$\forall i \neq j, B_i \cap B_j = \emptyset, \text{ et } \bigcup_{i \in I} B_i = \Omega.$$

Autrement dit, la famille $(B_i)_{i \in I}$ est une *partition* de Ω .

Théorème 7.1. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

- **Formule des probabilités totales.** Soit $(B_i)_{i \in I}$ un système complet d'événements de Ω , telle que pour tout $i \in I$, $\mathbb{P}(B_i) > 0$, et soit $A \in \mathcal{A}$. Alors :

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i) = \sum_{i \in I} \mathbb{P}_{B_i}(A) \mathbb{P}(B_i).$$

Dans le cas particulier où $I = \{1, \dots, n\}$, on a :

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap B_i) = \sum_{i=1}^n \mathbb{P}_{B_i}(A) \mathbb{P}(B_i).$$

- **Formule de Bayes.** Soient $A \in \mathcal{A}$ et $B \in \mathcal{A}$, tels que $0 < \mathbb{P}(A) < 1$ et $\mathbb{P}(B) > 0$, alors :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}_A(B) \mathbb{P}(A)}{\mathbb{P}_A(B) \mathbb{P}(A) + \mathbb{P}_{A^c}(B) \mathbb{P}(A^c)}.$$

Démonstration. Comme $(B_i)_{i \in I}$ est une partition de Ω , on a :

$$A = A \cap \Omega = A \cap \left(\bigcup_{i \in I} B_i \right) = \bigcup_{i \in I} (A \cap B_i),$$

où les événements $(A \cap B_i)_{i \in I}$ sont deux-à-deux disjoints. Ainsi, en utilisant la σ -additivité de \mathbb{P} :

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \cap B_i) = \sum_{i \in I} \mathbb{P}_{B_i}(A) \mathbb{P}(B_i).$$

Par définition de la probabilité conditionnelle, on a :

$$\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}_A(B) \mathbb{P}(A)}{\mathbb{P}(B)}.$$

La formule de Bayes est obtenue en appliquant la formule des probabilités totales au dénominateur avec la partition $\{A, A^c\}$ puisque $\mathbb{P}(A) \neq 0$ et $\mathbb{P}(A^c) \neq 0$. □

Exemple 7.2. Un sondage est effectué dans une société comprenant 40% de cadres et 60% d'ouvriers. On sait que 20% des cadres et 10% des ouvriers de cette société savent parler anglais. On interroge une personne "au hasard". Quelle est la probabilité que ce soit :

- un cadre sachant parler anglais ?
- un ouvrier sachant parler anglais ?
- une personne sachant parler anglais ?

L'individu interrogé sait parler anglais. Quelle est la probabilité que ce soit un ouvrier ?

On prend pour univers Ω l'ensemble des employés de la société, que l'on munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$. Étant donné que l'on interroge une personne "au hasard", on munit Ω de la probabilité uniforme, notée \mathbb{P} .

Soit C l'événement "l'employé est un cadre", O "l'employé est un ouvrier", A "l'employé parle anglais". Alors, la donnée du problème nous dit que :

$$\mathbb{P}(C) = \frac{4}{10}, \quad \mathbb{P}(O) = \frac{6}{10}.$$

Comme ces deux probabilités sont strictement positives, on peut conditionner par rapport aux événements C et O . D'après la donnée nous savons que :

$$\mathbb{P}_C(A) = \frac{2}{10}, \quad \mathbb{P}_O(A) = \frac{1}{10}.$$

À la première question, on cherche $\mathbb{P}(C \cap A)$. D'après la définition des probabilités conditionnelles, on a :

$$\mathbb{P}(C \cap A) = \mathbb{P}_C(A)\mathbb{P}(C) = \frac{2}{10} \cdot \frac{4}{10} = 0,08.$$

De manière similaire pour la deuxième question :

$$\mathbb{P}(O \cap A) = \mathbb{P}_O(A)\mathbb{P}(O) = \frac{1}{10} \cdot \frac{6}{10} = 0,06.$$

Étant donné que $\{C, O\}$ forme une partition de Ω , d'après la formule des probabilités totales, on a :

$$\mathbb{P}(A) = \mathbb{P}(A \cap C) + \mathbb{P}(A \cap O) = 0,08 + 0,06 = 0,14.$$

Pour répondre à la dernière question, on utilise la formule de Bayes :

$$\mathbb{P}_A(O) = \frac{\mathbb{P}_O(A)\mathbb{P}(O)}{\mathbb{P}(A)} = \frac{1.6}{10.10} \cdot \frac{100}{14} = \frac{3}{7}.$$

Exemple 7.3. (Paradoxe de Simpson). Cet exemple réel¹ montre un paradoxe surprenant qui s'explique grâce aux probabilités conditionnelles et à la formule des probabilités totales. Il vous convaincra de l'importance de bien comprendre ce concept pour interpréter correctement les résultats d'études statistiques. Il provient d'une étude médicale sur le succès de deux traitements contre les calculs rénaux. Le traitement A a été effectué dans les années 1972-1980, et le traitement B dans les années 1980-1985.

La première table montre le succès global et le nombre de traitements pour chaque méthode.

Succès (taux de succès)	
Traitement A	Traitement B
273/350 (78%)	289/350 (83%)

Cela semble révéler que traitement B, qui est nouveau, est plus efficace. Maintenant, en ajoutant des données concernant la taille des calculs, la comparaison prend une autre tournure :

1. Charig CR, ; Webb DR, ; Payne SR, ; Wickham OE . Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy. BMJ 1986 ;292 : 879–82. 3

Résultats en fonction de la taille des calculs

petits calculs		grands calculs	
Traitement A	Traitement B	Traitement A	Traitement B
(81/87) 93%	(234/270) 87%	(192/263) 73%	(55/80) 69%

L'information au sujet de la taille des calculs a inversé les conclusions concernant l'efficacité de chaque traitement. Le traitement A est maintenant considéré comme plus efficace dans les deux cas. Le rebroussement de cette inégalité, qui conduit au paradoxe, se produit à cause de deux effets concurrents :

1. la variable supplémentaire (ici la taille) a un impact significatif sur les rapports ;
2. les tailles des groupes qui sont combinés quand la variable supplémentaire est ignorée sont très différentes. (Les groupes utilisés pour le traitement A et B ont la même taille, mais n'ont pas la même répartition de petits et grands calculs).

Vérifions les calculs pour le traitement A. On choisit comme univers Ω les 350 patients de l'échantillon, que l'on munit de la tribu $\mathcal{P}(\Omega)$. Appelons S l'événement “le traitement est un succès”, P l'événement “les calculs sont petits”. Alors d'après le premier tableau, $\mathbb{P}(S) = \frac{273}{350}$, et d'après le deuxième, $\mathbb{P}(P) = \frac{87}{350}$, $\mathbb{P}(P^c) = \frac{263}{350}$. Ces deux probabilités étant strictement positives, on peut conditionner par les événements correspondants. D'après le deuxième tableau toujours, on a $\mathbb{P}_P(S) = \frac{81}{87}$, $\mathbb{P}_{P^c}(S) = \frac{192}{263}$. Utilisons la formule des probabilités totales pour calculer $\mathbb{P}(S)$.

$$\begin{aligned}\mathbb{P}(S) &= \mathbb{P}_P(S)\mathbb{P}(P) + \mathbb{P}_{P^c}(S)\mathbb{P}(P^c) \\ &= \frac{81}{87} \frac{87}{350} + \frac{192}{263} \frac{263}{350} = \frac{273}{350}.\end{aligned}$$

On retrouve bien le résultat du premier tableau. Des calculs similaires permettent de vérifier les résultats pour le traitement B. Ainsi, ces résultats apparemment contradictoires s'expliquent aisément grâce aux probabilités conditionnelles.

7.2 INDÉPENDANCE DES ÉVÉNEMENTS

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Définition. Deux événements A et B de \mathcal{A} sont dits *indépendants*, si

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

Exemple 7.4.

1. Les événements \emptyset et Ω sont indépendants. En effet,

$$\mathbb{P}(\emptyset \cap \Omega) = \mathbb{P}(\emptyset) = 0, \text{ et } \mathbb{P}(\emptyset)\mathbb{P}(\Omega) = 0 \cdot 1 = 0,$$

d'où, $\mathbb{P}(\emptyset \cap \Omega) = \mathbb{P}(\emptyset)\mathbb{P}(\Omega)$.

2. On jette un dé parfaitement équilibré. Soit A l'événement “obtenir 1,2 ou 3”, et B l'événement “obtenir 1,2,4 ou 5”. On choisit comme espace des états, $\Omega = \{1, \dots, 6\}$, que l'on munit de la probabilité uniforme. On a alors,

$$\mathbb{P}(A) = \frac{1}{2}, \quad \mathbb{P}(B) = \frac{2}{3}, \quad \mathbb{P}(A \cap B) = \mathbb{P}(\{1, 2\}) = \frac{1}{3}.$$

Ainsi, comme $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, on déduit que les événements A et B sont indépendants.

Remarque.

- L'indépendance n'a rien à voir avec le fait que les événements soient disjoints ou non. Dans l'Exemple 2 ci-dessus, les événements sont indépendants, mais non disjoints ($A \cap B \neq \emptyset$).
- Si A et B sont deux événements de probabilité non nulle, alors :

$$A \text{ et } B \text{ sont indépendants} \Leftrightarrow \mathbb{P}(A) = \mathbb{P}_B(A) \Leftrightarrow \mathbb{P}(B) = \mathbb{P}_A(B).$$

Le fait d'avoir une information supplémentaire, à savoir que B est réalisé, ne modifie pas la probabilité de A (de même pour B lorsqu'on sait que A est réalisé) ce qui justifie la terminologie d'*indépendance*. Ces critères ne sont cependant pas utilisés comme définition car ils nécessitent l'hypothèse supplémentaire, $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$.

Proposition 7.1. *Si les événements A et B sont indépendants, il en est de même des événements A^c et B , A et B^c , A^c et B^c .*

Démonstration. Démontrons que A^c et B sont indépendants si A et B le sont.

$$\begin{aligned} \mathbb{P}(A^c \cap B) &= \mathbb{P}(B) - \mathbb{P}(A \cap B), \text{ d'après la formule des probabilités totales} \\ &= \mathbb{P}(B)(1 - \mathbb{P}(A)), \text{ car } A \text{ et } B \text{ sont indépendants} \\ &= \mathbb{P}(B)\mathbb{P}(A^c), \text{ donc } A^c \text{ et } B \text{ sont indépendants.} \end{aligned}$$

En remplaçant A et B par B et A^c , l'implication que l'on vient de prouver entraîne que si A^c et B sont indépendants (et donc B et A^c indépendants), alors B^c et A^c sont indépendants. Et ainsi de suite pour les autres cas. \square

Définition.

- Des événements A_1, \dots, A_n de \mathcal{A} sont dits *mutuellement indépendants*, si pour tout entier $1 \leq k \leq n$ et tout k -uplet d'entiers (i_1, \dots, i_k) , tels que, $1 \leq i_1 < \dots < i_k \leq n$,

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_k}).$$

- Les événements A_1, \dots, A_n sont dits *deux-à-deux indépendants*, si pour tout $i \in \{1, \dots, n\}$, $j \in \{1, \dots, n\}$, tels que $i \neq j$, on a :

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j).$$

CHAPITRE 8

VARIABLES ALÉATOIRES DISCRÈTES

8.1 DÉFINITIONS ET EXEMPLES

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Plutôt que de travailler avec des événements de \mathcal{A} , il est souvent plus commode d'associer une valeur numérique aux résultats d'une expérience aléatoire. Par exemple, lors de n jets de pile ou face, il sera intéressant d'étudier le nombre de piles obtenus. Cela motive l'introduction de la notion de *variable aléatoire*, qui est une application X de Ω dans un ensemble E qui sera typiquement, \mathbb{N} , \mathbb{Z} ou \mathbb{R} . Dans ce cours, nous nous restreignons aux variables aléatoires discrètes.

Définition. Une variable aléatoire X définie sur Ω est dite *discrète* si l'ensemble

$$X(\Omega) = \{x_j : j \in J\} \subset \mathbb{R}$$

des valeurs prises par X est fini ou dénombrable.

Définition (Proposition). Si X est une variable aléatoire discrète définie sur Ω , alors la loi de probabilité de X , notée \mathbb{P}^X , est la probabilité sur $X(\Omega)$ caractérisée par la donnée de :

$$\forall j \in J, \quad \mathbb{P}^X(\{x_j\}) = \mathbb{P}(\{X = x_j\}).$$

Remarque.

- La terminologie de variable aléatoire peut être trompeuse, car il s'agit en fait d'une fonction de Ω dans E .
- Soit $x_j \in X(\Omega)$. Rappelez-vous des notations :

$$X^{-1}(x_j) = \{\omega \in \Omega : X(\omega) = x_j\} = \{X = x_j\}.$$

$\{X = x_j\}$ est l'événement “ X prend la valeur x_j ”. Attention, $X^{-1}(x_j)$ désigne l'image réciproque x_j par l'application X . Cela ne sous-entend nullement que X est bijective !

Exemple 8.1.

1. On considère deux lancers successifs d'un même dé, et on note S la variable aléatoire correspondant à la somme des valeurs obtenues, ainsi S prend ses valeurs dans $\{2, 3, \dots, 12\}$. Calculons la loi de S . Choisissons comme univers $\Omega = \{1, \dots, 6\}^2$, que l'on munit de la tribu $\mathcal{A} = \mathcal{P}(\Omega)$, et de la probabilité uniforme, notée \mathbb{P} . Ainsi pour tout sous-ensemble

A de $\mathcal{P}(\Omega)$, $\mathbb{P}(A) = \frac{\text{card}(A)}{\text{card}(\Omega)}$. Étant donné que S prend ses valeurs dans $\{2, \dots, 12\}$, pour déterminer la loi de probabilité \mathbb{P}^S de S , il suffit de calculer, pour tout $i \in \{2, \dots, 12\}$, $\mathbb{P}^S(\{i\})$:

$$\begin{aligned}\mathbb{P}^S(\{2\}) &= \mathbb{P}(\{S = 2\}) \\ &= \mathbb{P}(\{\omega \in \Omega : S(\omega) = 2\}) = \mathbb{P}(\{(1, 1)\}) = \frac{1}{36}, \\ \mathbb{P}^S(\{3\}) &= \mathbb{P}(\{\omega \in \Omega : S(\omega) = 3\}) = \mathbb{P}(\{(1, 2), (2, 1)\}) = \frac{1}{18}, \\ &\quad \text{etc.}\end{aligned}$$

2. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $A \in \mathcal{A}$ un événement. Alors l'*indicatrice de A* , notée \mathbb{I}_A , est la variable aléatoire définie sur Ω par :

$$\forall \omega \in \Omega, \mathbb{I}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{sinon.} \end{cases}$$

Ainsi l'indicatrice de A prend ses valeurs dans $\{0, 1\}$. En probabilité, il est important de bien savoir manipuler cette notation. En particulier l'indicatrice satisfait aux conditions suivantes. Si A et B sont deux événements de \mathcal{A} , alors :

$$\mathbb{I}_{A^c} = 1 - \mathbb{I}_A, \quad \mathbb{I}_{A \cap B} = \mathbb{I}_A \mathbb{I}_B, \quad \mathbb{I}_{A \cup B} = \mathbb{I}_A + \mathbb{I}_B - \mathbb{I}_{A \cap B}.$$

Calculons la loi de l'indicatrice de A . Étant donné qu'elle prend ses valeurs dans $\{0, 1\}$, il suffit de déterminer, pour $i \in \{0, 1\}$, $\mathbb{P}^{\mathbb{I}_A}(\{i\})$:

$$\begin{aligned}\mathbb{P}^{\mathbb{I}_A}(\{1\}) &= \mathbb{P}(\mathbb{I}_A \in \{1\}) = \mathbb{P}(\{\omega \in \Omega : \mathbb{I}_A(\omega) = 1\}) \\ &= \mathbb{P}(\{\omega \in \Omega : \omega \in A\}) = \mathbb{P}(A), \\ \mathbb{P}^{\mathbb{I}_A}(\{0\}) &= \mathbb{P}(\mathbb{I}_A \in \{0\}) = \mathbb{P}(\{\omega \in \Omega : \mathbb{I}_A(\omega) = 0\}) \\ &= \mathbb{P}(\{\omega \in \Omega : \omega \notin A\}) = \mathbb{P}(A^c) = 1 - \mathbb{P}(A).\end{aligned}$$

Voici une liste de lois classiques à connaître.

Loi uniforme. Soit $n \in \mathbb{N}^*$ et une variable aléatoire, $X : \Omega \rightarrow \{x_1, \dots, x_n\}$, où pour tout $i \neq j$, $x_i \neq x_j$. Supposons que la loi de probabilité de X est donnée par :

$$\forall j \in \{1, \dots, n\}, \mathbb{P}^X(\{x_j\}) = \frac{1}{n}.$$

Alors la loi de X est appelée la *loi uniforme discrète* sur $\{x_1, \dots, x_n\}$.

Exemple 8.2. On lance une pièce équilibrée. Soit $\Omega = \{P, F\}$ que l'on munit de la probabilité uniforme \mathbb{P} . On définit la variable aléatoire $X : \Omega \rightarrow \{-1, 1\}$, par $X(\{P\}) = 1$, $X(\{F\}) = -1$. Ainsi, la loi de probabilité de X est :

$$\mathbb{P}^X(\{1\}) = \mathbb{P}(\{X = 1\}) = \mathbb{P}(\{P\}) = \frac{1}{2}, \quad \mathbb{P}^X(\{-1\}) = \mathbb{P}(\{X = -1\}) = \mathbb{P}(\{F\}) = \frac{1}{2},$$

et la variable aléatoire X suit une loi uniforme sur $\{-1, 1\}$.

Loi de Bernoulli. Soit $0 < p < 1$ et une variable aléatoire $X : \Omega \rightarrow \{0, 1\}$ de loi de probabilité :

$$\mathbb{P}^X(\{1\}) = p, \quad \mathbb{P}^X(\{0\}) = 1 - p.$$

Alors la loi de X est appelée *loi de Bernoulli de paramètre p* .

Exemple 8.3. Une *épreuve de Bernoulli de paramètre p* est une expérience aléatoire admettant deux issues succès/échec, telle que p est la probabilité d'obtenir un succès. Soit Ω représentant les issues de l'expérience, \mathbb{P} une probabilité sur Ω et A l'événement représentant le succès. D'après la description de l'expérience, on a $\mathbb{P}(A) = p$. Dans ce cas, l'indicatrice de A , \mathbb{I}_A , suit une loi de Bernoulli de paramètre p . En effet, \mathbb{I}_A est à valeurs dans $\{0, 1\}$ et nous avons déjà vu que :

$$\mathbb{P}^{\mathbb{I}_A}(\{1\}) = \mathbb{P}(A), \quad \mathbb{P}^{\mathbb{I}_A}(\{0\}) = 1 - \mathbb{P}(A).$$

On conclut en utilisant le fait que $\mathbb{P}(A) = p$.

Par exemple, on jette un dé équilibré une fois. On appelle “succès” l'événement A “obtenir un nombre plus grand ou égal à 2”. On choisit comme univers $\Omega = \{1, \dots, 6\}$, alors $\mathbb{P}(A) = \mathbb{P}(\{2, 3, 4, 5, 6\}) = \frac{5}{6}$ et \mathbb{I}_A suit une loi de Bernoulli de paramètre $5/6$.

Loi binomiale. Soit $n \in \mathbb{N}^*$, $0 < p < 1$ et une variable aléatoire $X : \Omega \rightarrow \{0, \dots, n\}$ de loi de probabilité :

$$\forall k \in \{0, \dots, n\}, \quad \mathbb{P}^X(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Alors la loi de X est appelée *loi binomiale de paramètres n et p* .

Exemple 8.4. On appelle *schéma de Bernoulli de paramètres n et p* l'expérience qui consiste à répéter n fois de manière indépendante une épreuve de Bernoulli de paramètre p . On choisit comme univers Ω^n , que l'on munit d'une probabilité \mathbb{P}_n . Pour $i \in \{1, \dots, n\}$, on note A_i l'événement “obtenir un succès lors de la i -ième expérience”. Soit $k \in \{0, \dots, n\}$, comme les expériences sont indépendantes, la probabilité d'obtenir un succès lors des k premières expériences et un échec lors des $n - k$ dernières est :

$$\begin{aligned} \mathbb{P}_n(A_1 \cap \dots \cap A_k \cap A_{k+1}^c \cap \dots \cap A_n^c) &= \mathbb{P}(A_1) \cdots \mathbb{P}(A_k) \mathbb{P}(A_{k+1}^c) \cdots \mathbb{P}(A_n^c) \\ &= p^k (1-p)^{n-k}. \end{aligned}$$

Soit X la variable aléatoire qui compte le nombre de succès dans un schéma de Bernoulli, alors X suit une loi binomiale de paramètres n et p . En effet,

$$\mathbb{P}^X(\{k\}) = \mathbb{P}_n(X = k) = \mathbb{P}_n(\{\text{obtenir } k \text{ succès sur les } n \text{ expériences}\}).$$

Il y a $\binom{n}{k}$ façons d'obtenir ces succès, et pour chacune des façons la probabilité est $p^k (1-p)^{n-k}$. Ainsi,

$$\mathbb{P}^X(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Par exemple, si on jette n fois une pièce de monnaie équilibrée, et on appelle X la variable aléatoire qui compte le nombre de fois où l'on obtient un nombre plus grand ou égal à 2, alors X suit une loi binomiale de paramètres n et $5/6$.

Loi géométrique. Soit $0 < p < 1$ et une variable aléatoire $X : \Omega \rightarrow \mathbb{N}^*$ de loi de probabilité :

$$\forall k \in \mathbb{N}^*, \mathbb{P}^X(\{k\}) = (1-p)^{k-1}p.$$

Alors la loi de X est appelée *loi géométrique de paramètre p* . En translatant de 1 les valeurs de la variable aléatoire, on obtient la loi géométrique sur \mathbb{N} .

Exemple 8.5. Considérons un schéma de Bernoulli où l'expérience est répétée indéfiniment. Si X est la variable aléatoire égale au temps passé jusqu'au premier succès, alors X suit une loi géométrique de paramètre p . En effet, choisissons comme univers $\Omega^{\mathbb{N}^*}$ et calculons la loi de probabilité de X . Pour tout $k \in \mathbb{N}^*$,

$$\begin{aligned} \mathbb{P}^X(\{k\}) &= \mathbb{P}_k(A_1^c \cap \dots \cap A_{k-1}^c \cap A_k) \\ &= \mathbb{P}(A_1^c) \cdots \mathbb{P}(A_{k-1}^c) \mathbb{P}(A_k), \text{ par indépendance} \\ &= (1-p)^{k-1}p. \end{aligned}$$

Loi de Poisson. Soit $\theta > 0$ et une variable aléatoire $X : \Omega \rightarrow \mathbb{N}$ de loi :

$$\forall k \in \mathbb{N}, \mathbb{P}^X(\{k\}) = e^{-\theta} \frac{\theta^k}{k!}.$$

Alors la loi de X est appelée *loi de Poisson de paramètre θ* .

Exemple 8.6. La loi de Poisson modélise le nombre d'autobus passés à un arrêt avant un instant T donné, et θ représente le nombre moyen d'arrivées dans cet intervalle.

8.2 FONCTION DE RÉPARTITION

Définition. Soit X une variable aléatoire de loi \mathbb{P}^X à valeurs dans \mathbb{R} (ou une partie de \mathbb{R}). On appelle *fonction de répartition de la loi \mathbb{P}^X* ou encore, par abus, *fonction de répartition de X* , l'application F_X définie sur \mathbb{R} par :

$$\forall t \in \mathbb{R}, F_X(t) = \mathbb{P}^X(]-\infty, t]) = \mathbb{P}(X \leq t).$$

Propriété 8.1. La fonction de répartition de la loi \mathbb{P}^X satisfait aux propriétés suivantes :

1. F_X prend ses valeurs dans $[0, 1]$,
2. F_X est une application croissante,
3. F_X est continue à droite et admet une limite à gauche,
4. $\lim_{t \rightarrow -\infty} F_X(t) = 0, \lim_{t \rightarrow +\infty} F_X(t) = 1$.

Les propriétés ci-dessus sont satisfaites en général, que la variable aléatoire soit discrète ou non. Celles ci-dessous ne sont satisfaites que pour les variables aléatoires discrètes.

Propriété 8.2. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et soit X une variable aléatoire discrète réelle définie sur Ω . Alors la fonction de répartition F_X de la loi \mathbb{P}^X , vérifie :

1. $F_X(x) = \sum_{\{y \in X(\Omega) : y \leq x\}} \mathbb{P}^X(\{y\}) ;$
2. si x et y sont deux points consécutifs de $X(\Omega)$, alors F_X est constante sur $[x, y[$;
3. la hauteur du saut en $x \in X(\Omega)$ est donnée par $\mathbb{P}^X(\{x\})$.

La fonction de répartition permet de calculer les probabilités suivantes :

Propriété 8.3. Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et X une variable aléatoire sur Ω de fonction de répartition F_X , alors :

- $\mathbb{P}(X > x) = 1 - F_X(x),$
- $\mathbb{P}(x < X \leq y) = F_X(y) - F_X(x),$
- $\mathbb{P}(X < x) = F_X(x^-).$

8.2.1 ESPÉRANCE

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète définie sur Ω .

L'espérance de la variable aléatoire X représente sa moyenne pondérée par la probabilité de chacune des issues. Lors d'un jeu de hasard par exemple, elle permet de déterminer si le jeu est équitable ou non.

Définition.

- La variable aléatoire discrète X est dite *intégrable*, si la série de terme général $|x_j| p_j$ converge.
- Si X est une variable aléatoire discrète intégrable, on définit son *espérance*, notée $\mathbb{E}(X)$, par :

$$\mathbb{E}(X) = \sum_{j \in J} x_j \mathbb{P}^X(\{x_j\}).$$

Remarque.

- Lorsque l'univers Ω est fini, toute variable aléatoire définie sur Ω est intégrable.
- Si X n'est pas intégrable, elle n'admet pas d'espérance.

L'espérance satisfait aux propriétés suivantes.

Propriété 8.4.

1. Une variable aléatoire discrète X est intégrable si et seulement si $|X|$ est intégrable.
2. Si la variable aléatoire X est bornée, alors elle est intégrable.
3. (Linéarité). Si X et Y sont deux variables aléatoires discrètes intégrables, et si a, b sont deux constantes réelles, alors $aX + bY$ est intégrable et on a :

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

4. (Monotonie). Si X et Y sont deux variables aléatoires discrètes telles que $X \leq Y$, alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$. En particulier si X est intégrable, alors $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$.
5. Si X est une variable aléatoire constante, $X \equiv a$, alors X est intégrable et $\mathbb{E}(X) = a$.

6. (Théorème de transfert). Si X est une variable aléatoire discrète et si f est une application définie sur $X(\Omega)$ telle que la série de terme général $|f(x_j)|\mathbb{P}^X(\{x_j\})$ converge, alors $f(X)$ est une variable aléatoire discrète intégrable et

$$\mathbb{E}(f(X)) = \sum_{j \in J} f(x_j) \mathbb{P}^X(\{x_j\}).$$

Définition. Soit X une variable aléatoire discrète intégrable. Si X est d'espérance nulle, on dit que X est *centrée*.

8.2.2 VARIANCE, MOMENTS D'ORDRES SUPÉRIEURS

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire discrète définie sur Ω .

Définition. La variable aléatoire X est dite de *carré intégrable*, si X^2 admet une espérance. La quantité $\mathbb{E}(X^2)$ est alors bien définie et est appelée *moment d'ordre 2* de X .

Remarque.

1. Grâce au théorème de transfert, la variable aléatoire X est de carré intégrable si et seulement si la série de terme général $x_j^2 \mathbb{P}^X(\{x_j\})$ converge.
2. Si la variable aléatoire X est de carré intégrable, alors elle est intégrable.

En effet, supposons que la variable aléatoire X prenne les valeurs $\{x_i : i \in \mathbb{N}\}$. Fixons $n \in \mathbb{N}^*$, et notons I_n l'ensemble des entiers compris entre 0 et n . Alors,

$$\begin{aligned} \sum_{i \in I_n} |x_i| \mathbb{P}^X(\{x_i\}) &= \sum_{\{i \in I_n : |x_i| \leq 1\}} |x_i| \mathbb{P}^X(\{x_i\}) + \sum_{\{i \in I_n : |x_i| > 1\}} |x_i| \mathbb{P}^X(\{x_i\}) \\ &\leq \sum_{\{i \in I_n : |x_i| \leq 1\}} 1 \cdot \mathbb{P}^X(\{x_i\}) + \sum_{\{i \in I_n : |x_i| > 1\}} x_i^2 \mathbb{P}^X(\{x_i\}) \\ &\leq 1 + \sum_{i \in I_n} x_i^2 \mathbb{P}^X(\{x_i\}). \end{aligned}$$

On conclut en utilisant le critère de comparaison des séries à termes positifs.

3. Soit $\lambda \in \mathbb{R}$. Si la variable aléatoire X est de carré intégrable, il en est de même pour les variables aléatoires $(X + \lambda)$ et $(|X| + \lambda)$.

Propriété 8.5. Si X est une variable aléatoire discrète de carré intégrable, alors :

$$\mathbb{E}(|X|)^2 \leq \mathbb{E}(X^2).$$

Démonstration. Soit $\lambda \in \mathbb{R}$. Si la variable aléatoire X est de carré intégrable, alors il en est de même pour $|X| + \lambda$ et nous posons : $f(\lambda) = \mathbb{E}[(|X| + \lambda)^2]$. En utilisant la linéarité de l'espérance,

$$f(\lambda) = \mathbb{E}(X^2) + 2\lambda \mathbb{E}(|X|) + \lambda^2.$$

Ainsi, f est un polynôme de degré 2, positif ou nul. Son discriminant est donc négatif ou nul, ce qui implique le résultat. \square

Définition. Si X est une variable aléatoire discrète de carré intégrable, on définit sa *variance*, notée $\text{Var}(X)$, par :

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

L'*écart-type*, noté $\sigma(X)$, est défini par $\sigma(X) = \sqrt{\text{Var}(X)}$.

Remarque.

- La variance de X mesure la façon dont X s'écarte de sa moyenne $\mathbb{E}(X)$.
- L'écart-type s'utilise surtout en statistique. Remarquer que si la variable aléatoire X a une unité, alors l'écart-type a la même unité, tandis que la variance a l'unité au carré, d'où l'intérêt dans la pratique de travailler avec l'écart-type.

Propriété 8.6. Soit X une variable aléatoire discrète de carré intégrable, alors :

1. $\text{Var}(X) \geq 0$,
2. $\forall a \in \mathbb{R}, \text{Var}(X + a) = \text{Var}(X)$,
3. $\forall a \in \mathbb{R}, \text{Var}(aX) = a^2 \text{Var}(X)$,
4. $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$,
5. $\text{Var}(X) = 0 \Leftrightarrow \forall j \in J \text{ tel que } \mathbb{P}^X(\{x_j\}) > 0, x_j = \mathbb{E}[X]$.

Démonstration.

1. Découle de la propriété de monotonie de l'espérance et du fait que $(X - \mathbb{E}(X))^2 \geq 0$.
2. 3. et 4. En utilisant la définition de la variance et la linéarité de l'espérance, on a :

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

5. D'après le théorème de transfert,

$$\text{Var}(X) = \sum_{j \in J} (x_j - \mathbb{E}(X))^2 \mathbb{P}^X(\{x_j\}).$$

C'est une somme de termes positifs. Ainsi, cette somme est nulle si et seulement si chacun de ses termes l'est, c'est-à-dire si et seulement si,

$$\forall j \in J, \text{ tel que } \mathbb{P}^X(\{x_j\}) > 0, x_j = \mathbb{E}(X).$$

□

Définition. Soit X une variable aléatoire discrète de carré intégrable. Si X est de variance égale à 1, on dit que X est *réduite*.

Définition. Soit X une variable aléatoire discrète, et soit $n \in \mathbb{N}^*$. Si X^n est intégrable, la quantité $\mathbb{E}(X^n)$ est bien définie et on l'appelle *moment d'ordre n* de la variable aléatoire X .

Remarque. D'après le théorème de transfert, la variable aléatoire X^n est intégrable si et seulement si la série de terme général $|x_j|^n \mathbb{P}^X(\{x_j\})$ converge.

8.2.3 INÉGALITÉ DE MARKOV ET DE BIENAYMÉ TCHEBYCHEV

Voici deux inégalités classiques.

Proposition 8.1 (Inégalité de Markov). *Soit X une variable aléatoire admettant un moment d'ordre $n \geq 1$. Alors,*

$$\forall a > 0, \quad \mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[|X|^n]}{a^n}.$$

Démonstration. Une manière courte d'écrire cette preuve est d'utiliser une variable aléatoire indicatrice. Remarquer que l'on peut écrire la fonction constante égale à 1 sur Ω de la manière suivante : pour tout $\omega \in \Omega$, pour tout réel $a > 0$,

$$1(\omega) = \mathbb{I}_{\{|X| \geq a\}}(\omega) + \mathbb{I}_{\{|X| < a\}}(\omega).$$

Ainsi, pour tout $\omega \in \Omega$, pour tout $a > 0$, on a :

$$\begin{aligned} |X(\omega)|^n &= |X(\omega)|^n \mathbb{I}_{\{|X| \geq a\}}(\omega) + |X(\omega)|^n \mathbb{I}_{\{|X| < a\}}(\omega) \\ &\geq |X(\omega)|^n \mathbb{I}_{\{|X| \geq a\}}(\omega), \quad \text{car } |X(\omega)|^n \mathbb{I}_{\{|X| < a\}}(\omega) \geq 0 \\ &\geq a^n \mathbb{I}_{\{|X| \geq a\}}(\omega), \quad \text{car } a \text{ est positif.} \end{aligned}$$

Autrement dit, on a l'inégalité $|X|^n \geq a^n \mathbb{I}_{\{|X| \geq a\}}$. On conclut en utilisant la monotonie de l'espérance :

$$\mathbb{E}[|X|^n] \geq \mathbb{E}[a^n \mathbb{I}_{\{|X| \geq a\}}] = a^n \mathbb{P}[|X| \geq a].$$

□

Proposition 8.2 (Inégalité de Bienaym -Tchebychev). *Soit X une variable al atoire discr e de carr  int grable. Alors,*

$$\forall a > 0, \quad \mathbb{P}[|X - \mathbb{E}(X)| \geq a] \leq \frac{\text{Var}(X)}{a^2}.$$

Démonstration. L'in galit  de Bienaym -Tch bychev est une cons quence de l'in galit  de Markov. La variable al atoire X  tant de carr  int grable, il en est de m me pour la variable al atoire $X - \mathbb{E}(X)$. On applique l'in galit  de Markov avec la variable al atoire $Y = X - \mathbb{E}(X)$, et $n = 2$. Pour tout $a > 0$,

$$\begin{aligned} \mathbb{P}(|Y| \geq a) &\leq \frac{\mathbb{E}[|Y|^2]}{a^2} \\ \Leftrightarrow \mathbb{P}(|X - \mathbb{E}(X)| \geq a) &\leq \frac{\mathbb{E}[(X - \mathbb{E}(X))^2]}{a^2} = \frac{\text{Var}(X)}{a^2}. \end{aligned}$$

□