

LABORATOIRE



EN PROCESSUS INTELLIGENTS

*Despina BASKIOTIS*

*STATISTIQUE*  
*À UNE ET PLUSIEURS DIMENSIONS*



*EISTI, 2005-06*

Copyright © 2005 Despina BASKIOTIS

Le contenu de ce document peut être redistribué dans son intégralité, sous les conditions énoncées dans la Licence pour Documents Libres (LDL) version 1.1 ou ultérieure.

En particulier, il ne doit, sous aucun prétexte, être modifié.

## AVANT-PROPOS

Par « statistique » on entend trois choses différentes :

- La collecte des données concernant un phénomène.
- Les données elles-mêmes.
- Le traitement de données.

Bien sûr ces trois notions ne sont pas indépendantes car, si nous pouvons faire un traitement de données, ces données doivent exister et, par conséquent, il faut au préalable procéder à une collecte de ces données. On conçoit aussi aisément que les résultats du traitement de données sont tributaires des règles utilisées pour le recueil de données de sorte que le même traitement appliqué sur deux ensembles de données, concernant le même phénomène mais collectés selon des règles différentes, peut aboutir à des résultats non convergents. Pendant ce cours on s'occupera exclusivement du traitement de données. La collecte de données, les sondages, l'échantillonnage, etc. ont leurs propres règles, dont la présentation est en dehors des buts du cours.

Bien que la collecte de données, en tant qu'activité humaine, se perd dans la nuit des temps, le mot statistique est récent. Il fut créé au milieu du 18<sup>e</sup> siècle et le formalisme de la statistique mathématique commença à se développer depuis le début de notre siècle.

La statistique mathématique est la branche de mathématiques qui fournit les méthodes qui permettent de traiter des données relatives à un phénomène et d'obtenir, à l'aide de l'inférence (ou induction) statistique, des modèles du comportement du phénomène étudié.

Essentiellement il s'agit d'étudier un phénomène par rapport à une ou deux de ses caractéristiques, qui sont supposées indépendantes. Le fondement théorique de cette étude est la théorie des probabilités, en particulier les lois de probabilités, les convergences des lois, la loi de grands nombres et le théorème central limite. Cette statistique s'appelle statistique classique. Quant à nous nous préférons l'appeler – pour des raisons évidentes –

statistique à une dimension ou monodimensionnelle.

À partir des années soixante l'utilisation de l'ordinateur a permis de traiter, pour un phénomène donné, plusieurs de ses caractéristiques à la fois, sans même être obligé de faire l'hypothèse de leur indépendance. Cette partie de la statistique s'appelle analyse de données. La cohérence nous oblige à l'appeler statistique à plusieurs dimensions ou multidimensionnelle.

L'objectif du cours de la Statistique Mathématique, donné à l'EISTI pendant la 2e année, est de permettre à l'élève-ingénieur d'acquérir les éléments théoriques de base et les techniques fondamentales. De plus, il doit inciter l'élève à avoir dans sa démarche de résolution des problèmes, une approche statistique afin d'enrichir sa réflexion.

Ce support de cours est destiné à fournir, sous forme écrite, les principaux résultats et à favoriser l'étude, chez soi, des certains points qui ne paraissent pas, ou qui ne sont pas, faciles. De plus pour la présentation des techniques statistiques, nous avons délibérément choisi une forme algorithmique, assez inhabituelle en statistique. Ainsi, pendant les séances des T.D., la consultation de ce photocopié permettra un travail plus confortable et, espérons-le, plus efficace. Afin que le photocopié soit (relativement) autonome, nous reprenons en une annexe les résultats fondamentaux de la Théorie des Probabilités, ce qui règle, aussi, le problème de l'homogénéité des notations.

# 1

## INTRODUCTION

Le but de la *statistique mathématique* est de développer un formalisme qui permet l'étude des *caractéristiques* d'un ensemble d'éléments en se fondant sur l'étude d'une partie (en général, petite) extraite de cet ensemble.

L'ensemble dont nous voulons faire l'étude s'appelle *population* et il est supposé être potentiellement infini. La partie de cet ensemble que nous utilisons pour faire l'étude s'appelle *échantillon*. Les caractéristiques (on dit aussi *caractères*) étudiées de la population peuvent être de deux types :

- Caractéristiques quantitatives.- Ce sont des résultats issus de mesures. À une caractéristique quantitative on associe une variable continue, qui représente toutes les valeurs possibles de la mesure considérée. Selon la structure de cet ensemble de valeurs de la mesure, on distingue :
  - *caractéristique quantitative ordinale*, quand dans l'ensemble de mesures nous pouvons définir seulement un ordre (e.g. mesure de température).
  - *caractéristique quantitative normale*, quand, en plus de l'ordre, nous pouvons définir, dans l'ensemble de mesures, des opérations arithmétiques, comme l'addition ou la multiplication (e.g. teneur en nitrates de l'eau potable).
- Caractéristiques qualitatives.- Ce sont des résultats issus d'observations. Une telle caractéristique peut prendre plusieurs valeurs qui sont mutuellement exclusives et dont le nombre dépend du point de vue adopté. Ces valeurs, dans le cas des caractéristiques qualitatives, s'appellent des *modalités*. À une caractéristique qualitative on associe une variable discrète, qui représente toutes les modalités. Selon la structure de l'ensemble des modalités on distingue :

- *caractéristique qualitative ordinale*, quand dans l'ensemble des modalités nous pouvons définir seulement un ordre (e.g. âge).
- *caractéristique qualitative nominale*, quand dans l'ensemble des modalités nous ne pouvons pas définir une structure (e.g. lieu d'habitation).

Un échantillon peut être représenté à l'aide d'un tableau  $\mathbf{X}$  à deux dimensions. Les lignes de ce tableau représentent les éléments de l'échantillon. Chaque colonne du tableau correspond à une caractéristique. Le tableau  $\mathbf{X}$  est appelé le *tableau de données*. Actuellement il est à la mode de l'appeler *corpus*, comme si l'échantillon d'une population pourrait nous donner l'ensemble des œuvres de la population ! Une forme particulière de ce tableau s'appelle *tableau de contingence*. Si on note par  $\mathcal{E}$  l'échantillon, par  $\mathcal{C}$  l'ensemble de caractéristiques, avec  $\text{card}(\mathcal{E}) = n$  et  $\text{card}(\mathcal{C}) = m$ , alors un élément du tableau  $X$  sera noté  $x_{ij}$ , où  $i = 1, \dots, n \stackrel{\Delta}{=} ]n]$  et  $j = 1, \dots, m \stackrel{\Delta}{=} ]m]$ . La  $i$ -ième ligne du tableau correspond à l' $i$ -ième élément de l'échantillon et sera notée  $\mathbf{x}_i$ . Il s'agit d'un vecteur-ligne à  $m$  dimensions qui contient toutes les valeurs des mesures et observations faites sur cet élément. De la même façon la  $j$ -ième colonne du tableau correspond à la  $j$ -ième caractéristique de l'échantillon et sera notée  $\mathbf{x}_j$ . Il s'agit d'un vecteur-colonne à  $n$  dimensions qui contient tous les résultats de la mesure (ou de l'observation) correspondants à cette caractéristique. Nous pouvons considérer que cette mesure (ou observation) peut être modélisée par l'intermédiaire d'une variable aléatoire  $X_j$ , qui prend les valeurs  $X_j = x_{ij}$ ;  $i \in ]n]$ . C'est la raison pour laquelle dorénavant, à la place des caractéristiques, nous allons parler des *variables*, en gardant bien sûr la distinction entre variables quantitatives et variables qualitatives. Distinction qui, soit dit en passant, ne nous servira point, car dans la suite nous nous occuperons uniquement des variables quantitatives.

Le travail statistique consiste, comme nous l'avons déjà dit au début, à imaginer, en utilisant des résultats expérimentaux obtenus à partir de l'échantillon, ce qui se passerait pour un nombre infini des résultats. Deux approches sont possibles :

- L'approche descriptive.- Elle consiste essentiellement à utiliser l'histogramme des effectifs, que nous pouvons réaliser comme suit :

Considérons une variable (aléatoire)  $X$  (i.e. une colonne du tableau  $\mathbf{X}$ ) et supposons qu'on découpe l'intervalle de variation de cette variable en sous-intervalles de longueur identique  $[x_a, x_b[, [x_b, x_c[, \dots, [x_k, x_l[, \dots$ . Pour chaque sous-intervalle  $[x_k, x_l[$  on trace un rectangle dont la base est égale à  $x_l - x_k$  et la hauteur proportionnelle au nombre  $n_{kl}$  d'éléments  $\mathbf{x}_i$  de l'échantillon dont les valeurs  $X$  sont telles que  $x_k \leq X < x_l$ .

Si on veut faire des comparaisons entre plusieurs histogrammes, il est d'usage de normaliser, c'est-à-dire de construire des rectangles dont la hauteur est proportionnelle à la fréquence  $f_{kl} = \frac{n_{kl}}{n}$ .

Pour le passage des résultats expérimentaux, issus de l'échantillon, aux résultats concernant la population, on admet que la fréquence  $f_{kl}$  a une limite pour chaque sous-intervalle  $[x_k, x_l[$ , i.e. on suppose que nous avons  $\lim_{n \rightarrow \infty} f_{kl} = \lim_{n \rightarrow \infty} \frac{n_{kl}}{n} = \overline{f_{kl}}$  où  $\overline{f_{kl}}$  est la valeur de la fréquence correspondante dans la population infinie.

Remarquons que la fonction qui fournit la surface des rectangles de l'histogramme d'effectifs d'une population infinie est la fonction de distribution de la v.a.  $X$ .

- L'approche analytique.- On utilise ici aussi la variable  $X$  et on cherche à résumer la distribution expérimentale (c'est-à-dire la distribution observée et reportée au tableau de données  $\mathbf{X}$ ) par des paramètres qui caractérisent la tendance centrale et la dispersion des valeurs de  $X$ . Ensuite on extrapole les résultats obtenus de l'échantillon à la population totale. Cette extrapolation se fait en utilisant des méthodes de calcul empruntées à l'Analyse, mais, de plus, en Statistique on fixe au préalable la valeur des risques encourus par cette opération.

La première approche est l'objet de la *Statistique Descriptive* qui ne rentre pas dans le cadre de ce cours. Dans la suite nous travaillerons, donc, avec la seconde approche qui nous obligera de passer continuellement de la population à l'échantillon et vice versa.

En utilisant un échantillon on obtient des mesures expérimentales et tout calcul utilisant ces mesures est un calcul certain (c-à-d. non probabiliste) tandis que toute extrapolation de ce calcul à la population totale est de nature probabiliste. Cette extrapolation porte le nom d'*inférence statistique*.

Bien évidemment cette analyse peut s'étendre à plusieurs caractéristiques et à plusieurs populations. Nous entrons ainsi dans ce qu'il est convenu d'appeler l'*Analyse de Données* et qui est, en réalité, la *Statistique Multidimensionnelle*. L'objectif de cette branche de la Statistique est plus général et plus ambitieux : étudier l'influence isolée et simultanée des caractéristiques sur une ou plusieurs populations. Et, par un effet de symétrie parfois délicate à mettre en œuvre, nous sommes aussi amenés à étudier la structuration des variables telle que les observations nous laissent l'appréhender.

Ce polycopié est divisé en deux parties : statistique monodimensionnelle et statistique multidimensionnelle.

La statistique monodimensionnelle débute avec quelques rappels de la théorie des probabilités concernant les théorèmes limites et les lois des grands nombres. Ensuite nous passons en revue les lois des probabilités, avant de présenter la méthode de l'inférence statistique. La première partie se termine avec la présentation des techniques statistiques, i.e. l'estimation ponctuelle et par intervalle de confiance, les tests d'hypothèses, le contrôle de réception et l'analyse de variance.

Dans la deuxième partie on introduit la statistique multidimensionnelle par l'intermédiaire de l'analyse canonique, ce qui permet par la suite de présenter comme des applications la régression linéaire, l'analyse et la structuration de données.



# 2

## LES THÉORÈMES LIMITES

---

2.1	Convergence et mesure de probabilité . . . . .	8
2.1.1	Convergence presque sûre . . . . .	8
2.1.2	Convergence en probabilité . . . . .	8
2.1.3	Convergence en moyenne quadratique . . . . .	9
2.1.4	Convergence en loi . . . . .	10
2.1.5	Relations entre les différentes convergences . . . . .	10
2.2	Quelques inégalités en probabilités . . . . .	10
2.3	Lois des grands nombres . . . . .	11
2.3.1	Loi forte des grands nombres . . . . .	11
2.3.2	Loi faible des grands nombres . . . . .	12
2.3.3	Conséquences des lois des grands nombres . . . . .	12
2.4	Théorème central limite (TCL) . . . . .	13

---

Considérons plusieurs v.a. définies sur le même espace probabilisé  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $\mathbb{R}$ . En analyse nous étudions le comportement des fonctions en utilisant le concept de la convergence et nous savons que la suite de fonctions  $\{f_n\}_n$  converge vers  $f$  si  $\lim_{n \rightarrow \infty} \|f_n - f\| = 0$  où nous avons noté  $\|f\| = \left| \int_{\mathbb{R}} |f(x)|^2 dx \right|^{1/2}$ .

Nous pouvons utiliser le même concept pour l'examen des v.a. voisines, en tenant compte que nous avons comme élément supplémentaire pour l'étude de la convergence, soit la mesure de probabilité, soit la fonction de répartition.

## 2.1 Convergence et mesure de probabilité

Dans ce paragraphe on considère une suite des v.a.  $X_1, \dots, X_n, \dots$  de  $(\Omega, \mathcal{A}, P)$  dans  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ . On notera cette suite des v.a. par  $(X_n)_n$ .

### 2.1.1 Convergence presque sûre

DÉFINITION 2.1.1 *On dit que la suite des v.a.  $(X_n)_n$  converge vers la v.a.  $X$  presque sûrement ou avec probabilité un et l'on note*

$$\lim_{n \rightarrow \infty} X_n \stackrel{ps}{=} X$$

si

$$P \left[ \lim_{n \rightarrow \infty} X_n = X \right] = 1$$

c'est-à-dire si

$$P \left[ \omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right] = 1 \quad \forall \omega \in \Omega - \Omega_0 \quad \text{avec} \quad \Omega_0 \subset \Omega \quad \text{tel que} \quad P(\Omega_0) = 0$$

Par conséquent la convergence presque sûre est la convergence en moyenne sur tout  $\Omega$ , sauf éventuellement sur un sous-ensemble  $\Omega_0$  de mesure nulle.

Nous avons les propriétés suivantes :

PROPRIÉTÉ 2.1.1 *Si  $P \left[ \lim_{n \rightarrow \infty} X_n = X \right] = 1$  et  $P \left[ \lim_{n \rightarrow \infty} X_n = X' \right] = 1$ , alors*

$$P[X = X'] = 1.$$

PROPRIÉTÉ 2.1.2 *Nous avons  $P \left[ \lim_{n \rightarrow \infty} X_n = X \right] = 1$  si et seulement si*

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P \left( \left\{ \sup_{m > n} |X_m - X| < \varepsilon \right\} \right) = 1$$

PROPRIÉTÉ 2.1.3 *Si  $P \left[ \lim_{n \rightarrow \infty} X_n = X \right] = 1$ , alors nous pouvons extraire une sous-suite*

$(X_{n_k})_{n_k}$  *de la suite  $(X_n)$  telle que  $P \left[ \lim_{n \rightarrow \infty} X_{n_k} = X \right] = 1$ .*

PROPRIÉTÉ 2.1.4 *Si  $E(X_n) = 0 \quad \forall n \in \mathbb{N}$  et  $\sum_n V(X) < +\infty$ , alors  $P \left[ \lim_{n \rightarrow \infty} X_n = X \right]$ .*

### 2.1.2 Convergence en probabilité

DÉFINITION 2.1.2 *On dit que la suite des v.a.  $(X_n)_n$  converge en probabilité vers la v.a.  $X$  et l'on note*

$$\lim_{n \rightarrow \infty} X_n \stackrel{p}{=} X$$

si

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} P(\{|X_n - X| < \varepsilon\}) = 1$$

Afin de se rendre compte de la différence entre la convergence en probabilité et la convergence presque sûre, nous allons utiliser le langage de l'analyse. Pour les deux notions nous avons que  $\forall \varepsilon > 0$  et  $\forall \delta > 0$ , il existe un entier  $N(\varepsilon, \delta) > 0$  tel que  $P(\{|X_n - X| < \varepsilon\}) = 1 \quad \forall n \geq N(\varepsilon, \delta)$ . Mais la convergence presque sûre établit en plus que cette relation est valable pour chaque  $\omega \in \Omega - \Omega_0$ , i.e.  $\lim_{n \rightarrow \infty} X_n = X$  signifie que  $\forall \varepsilon > 0, \forall \delta > 0$  et  $\forall \omega \in \Omega - \Omega_0$  il existe un entier  $N(\varepsilon, \delta; \omega) > 0$  tel que  $P(\{|X_n(\omega) - X(\omega)| < \varepsilon\}) = 1 \quad \forall n \geq N(\varepsilon, \delta; \omega)$ . Nous avons les propriétés suivantes :

PROPRIÉTÉ 2.1.5 Si  $\lim_{n \rightarrow \infty} X_n = X$  et  $\lim_{n \rightarrow \infty} X_n = X'$ , alors  $P[X = X'] = 1$ .

PROPRIÉTÉ 2.1.6 Si  $\lim_{n \rightarrow \infty} X_n = X$  et  $f : \mathbb{R} \rightarrow \mathbb{R}$ , alors  $\lim_{n \rightarrow \infty} f(X_n) = f(X)$ .

PROPRIÉTÉ 2.1.7 Si  $\lim_{n \rightarrow \infty} X_n = X$ ,  $\lim_{n \rightarrow \infty} Y_n = Y$  et  $P[X = Y] = 1$ , alors

$$\forall \varepsilon > 0 : P[|X_n - Y_n| \geq \varepsilon] = 0.$$

PROPRIÉTÉ 2.1.8 Si la suite des v.a.  $(X_n)_n$  converge presque sûrement vers  $X$  dans  $(\Omega, \mathcal{A}, P)$ , alors  $(X_n)_n$  converge en probabilité vers  $X$ .

### 2.1.3 Convergence en moyenne quadratique

On fait l'hypothèse que les moments d'ordre 2 de la suite des v.a.  $(X_n)_n$  et de la v.a.  $X$  existent.

DÉFINITION 2.1.3 On dit que la suite  $(X_n)_n$  converge en moyenne quadratique vers  $X$  et l'on note

$$\lim_{n \rightarrow \infty} X_n = X_{mq}$$

si

$$\lim_{n \rightarrow \infty} E[|X_n - X|^2] = 0$$

c'est-à-dire que  $\forall \varepsilon > 0$  il existe un entier  $N(\varepsilon) > 0$  tel que  $E[|X_n - X|^2] < \varepsilon \quad \forall n \geq N(\varepsilon)$ .

Notons que cette convergence est définie pour tout  $\omega \in \Omega$ .

Nous avons la propriété suivante :

PROPRIÉTÉ 2.1.9 Si  $\lim_{n \rightarrow \infty} X_n = X$  et  $\lim_{n \rightarrow \infty} Y_n = Y$ , alors  $\lim_{n \rightarrow \infty} X_n Y_n = X Y$ .

### 2.1.4 Convergence en loi

La convergence en loi utilise la notion de la fonction de répartition. Au lieu donc de s'intéresser au comportement voisin des v.a. on s'intéressera au voisinage des fonctions de répartition des v.a.

DÉFINITION 2.1.4 On dit que la suite  $(X_n)_n$  converge en loi vers  $X$  et l'on note

$$\lim_{n \rightarrow \infty} X_n \stackrel{l}{=} X$$

si

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad \forall x \in \mathbb{R} \text{ pour lequel la fonction } F_X \text{ est continue}$$

Nous avons les deux propriétés suivantes :

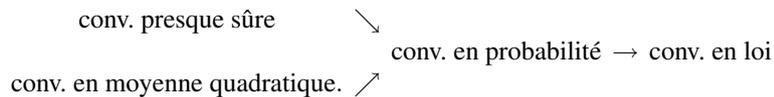
PROPRIÉTÉ 2.1.10 Si  $\lim_{n \rightarrow \infty} X_n \stackrel{l}{=} X$  et  $\lim_{n \rightarrow \infty} Y_n \stackrel{l}{=} Y$ , alors  $X$  et  $Y$  suivent la même loi de probabilité.

Si  $\lim_{n \rightarrow \infty} X_n \stackrel{mq}{=} X$ , alors pour tout intervalle  $[a, b[ \subset \mathbb{R}$  pour lequel la fonction de répartition  $F_X$  est continue nous avons

$$\lim_{n \rightarrow \infty} P[a \leq X_n(\omega) < b] = P[a \leq X(\omega) < b]$$

### 2.1.5 Relations entre les différentes convergences

Nous avons le schéma d'implication suivant :



## 2.2 Quelques inégalités en probabilités

Nous avons le

THÉORÈME 2.2.1 Soit  $X$  une v.a. sur  $(\Omega, \mathcal{A}, P)$ . Considérons une fonction réelle  $g(X)$ . Alors  $g(X)$  est une v.a. telle que

$$P[g(X) \geq c] \leq \frac{E[g(X)]}{c} ; \forall c > 0 \tag{2.2.1}$$

Cas particuliers :

(1) Posons  $g(X) = |X - E(X)|^r$ . Alors

$$P[|X - E(X)|^r \geq c^r] \leq \frac{E[|X - E(X)|^r]}{c^r} ; \forall c > 0 \quad (\text{Inégalité de Markov}) \tag{2.2.2}$$

(2) Si  $g(X) = |X - E(X)|^2$ , alors :

$$P \left[ |X - E(X)|^2 \geq c^2 \right] \leq \frac{E \left[ |X - E(X)|^2 \right]}{c^2} = \frac{V(X)}{c^2} ; \forall c > 0 \quad (2.2.3)$$

(3) Si  $c = \delta \sqrt{V(X)}$ , alors :

$$P \left[ |X - E(X)| \geq \delta \sqrt{V(X)} \right] \leq \frac{1}{\delta^2} ; \forall c > 0 \quad (\text{Inégalité de Chebychev}) \quad (2.2.4)$$

Nous avons aussi la forme équivalentes :

$$P \left[ |X - E(X)| \leq \delta \sqrt{V(X)} \right] \geq 1 - \frac{1}{\delta^2} ; \forall c > 0 \quad (2.2.5)$$

**THÉORÈME 2.2.2 (Inégalité de Kolmogorov).** - Si  $(X_n)_n$  suite de v.a. indépendantes, centrées et de variance finie, alors nous avons :

$$P \left[ \max_{1 \leq k \leq n} |S_k| > c \right] \leq \frac{V(S_n)}{c^2} ; \forall c > 0, n = 1, 2, \dots \quad (2.2.6)$$

où on a posé :  $S_k = \sum_{i=1}^k X_i$ .

## 2.3 Lois des grands nombres

### 2.3.1 Loi forte des grands nombres

Le théorème suivant est le critère de Kolmogorov pour la loi forte des grands nombres :

**THÉORÈME 2.3.1** *Considérons la suite  $(X_n)_n$  de v.a. indépendantes. Si ces v.a. sont centrées et telles que  $\sum_{i=1}^k \frac{V(X_k)}{k^2}$  converge ;  $\forall k = 1, 2, \dots, n$ , alors la suite  $(X_n)_n$  obéit à la loi forte des grands nombres.*

Dans le cas où la suite  $(X_n)_n$  est composée de v.a. indépendantes et identiquement distribuées (v.i.i.d.), alors nous avons le théorème de Khinchine :

**THÉORÈME 2.3.2** *Considérons la suite  $(X_n)_n$  de v.a.i.i.d. . Si ces v.a. sont telles que  $E(X_k) = m < \infty$  ;  $\forall k = 1, 2, \dots, n$ , alors la suite  $(X_n)_n$  obéit à la loi forte des grands nombres.*

Dans le cas où les v.a. de la suite  $(X_n)_n$  sont dépendantes entre elles, nous avons le théorème de Parzen :

**THÉORÈME 2.3.3** *Considérons la suite  $(X_n)_n$  de v.a. dépendantes, telles que*

$V(X_k) \leq M < \infty ; \forall k = 1, 2, \dots, n$ , avec  $M$  une constante. Alors la suite  $(X_n)_n$  obéit à la loi forte des grands nombres s'il existe des constantes positives  $Q$  et  $q$  telles que

$$|E\left(X_k \cdot \frac{1}{k} \sum_{i=1}^k X_i\right)| \leq \frac{Q}{k^q} ; \forall k = 1, 2, \dots, n \quad (2.3.1)$$

### 2.3.2 Loi faible des grands nombres

Nous avons les résultats suivants :

**THÉORÈME 2.3.4** Soit la suite  $(X_n)_n$  de v.a. indépendantes telles que  $|E(X_k)| < \infty ; \forall k = 1, 2, \dots$ . Posons

$$S_n^2 = \sum_{k=1}^n V(X_k)$$

on suppose que

$$\lim_{n \rightarrow \infty} \frac{S_n^2}{n^2} = 0$$

alors la suite  $(X_n)_n$  obéit à la loi faible des grands nombres.

**THÉORÈME 2.3.5** Soit la suite  $(X_n)_n$  de v.a. indépendantes telles que  $|E(X_k)| < \infty ; \forall k = 1, 2, \dots$ . S'il existe une constante  $\delta > 0$  telle que

$$\lim_{n \rightarrow \infty} \frac{1}{n^{1+\delta}} \sum_{k=1}^n E\left(|X_k - E(X_k)|^{1+\delta}\right) = 0$$

alors la suite  $(X_n)_n$  obéit à la loi faible des grands nombres.

### 2.3.3 Conséquences des lois des grands nombres

Nous donnons ci-après les principaux résultats concernant les applications des lois des grands nombres.

**THÉORÈME 2.3.6** Soient  $(X_n)_n$  et  $(Y_n)_n$  deux suites de v.a. indépendantes telles que

$$\lim_{n \rightarrow \infty} X_n = X \quad \text{et} \quad \lim_{n \rightarrow \infty} Y_n = Y$$

Alors

- (i)  $\lim_{n \rightarrow \infty} (aX_n + bY_n) = aX + bY$
- (ii)  $\lim_{n \rightarrow \infty} (X_n \cdot Y_n) = X \cdot Y$
- (iii)  $\lim_{n \rightarrow \infty} (X_n / Y_n) = X / Y$  si  $p(\{Y_n \neq 0\}) = 1$  et  $p(\{Y \neq 0\}) = 1$ .

THÉORÈME 2.3.7 Soient  $(X_n)_n$  et  $(Y_n)_n$  deux suites de v.a. indépendantes telles que

$$\lim_{n \rightarrow \infty} X_n \underset{l}{=} X \quad \text{et} \quad \lim_{n \rightarrow \infty} Y_n \underset{p}{=} c$$

avec  $c$  constante non nulle. Nous avons alors

- (i)  $\lim_{n \rightarrow \infty} (X_n + Y_n) \underset{l}{=} X + c$
- (ii)  $\lim_{n \rightarrow \infty} (X_n \cdot Y_n) \underset{l}{=} c \cdot X$
- (iii)  $\lim_{n \rightarrow \infty} (X_n / Y_n) \underset{l}{=} X / c$  si  $p(\{Y_n \neq 0\}) = 1$ .

## 2.4 Théorème central limite (TCL)

Les résultats du paragraphe précédent assurent que, sous certaines conditions,  $\frac{1}{n} \sum_{k=1}^n X_k$  converge vers  $\frac{1}{n} \sum_{k=1}^n E(X_k)$ . Mais, en règle générale,  $E(X_k)$  est une v.a., de sorte que ces résultats, bien qu'ils valident la démarche statistique, n'apportent pas une aide pour le calcul de  $\frac{1}{n} \sum_{k=1}^n X_k$ .

Le théorème suivant nous fournit plus des précisions dans le cas particulier des v.a.i.i.d.

THÉORÈME 2.4.1 (Th. central limite).- Soit la suite  $(X_n)_n$  de v.a.i.i.d. telles que  $E(X_k) = \mu$  avec  $|\mu| < \infty$  et  $V(X_k) = \sigma^2 < \infty$ ;  $\forall k = 1, 2, \dots$ . Notons par

$$S_n = \frac{1}{n} \sum_{k=1}^n X_k$$

et par

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

la v.a. centrée réduite de  $S_n$ . Alors nous avons

$$\lim_{n \rightarrow \infty} Z_n \underset{l}{=} Z \sim \mathcal{N}(0, 1)$$

De ce théorème nous pouvons conclure que

$$\lim_{n \rightarrow \infty} S_n \underset{l}{=} S \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Dans le cas où les v.a. ne suivent pas la même loi, nous avons le :

THÉORÈME 2.4.2 (Théorème de Lyapounov).- Soit la suite  $(X_n)_n$  de v.a. telles que  $E(X_k) = \mu_k$  avec  $|\mu_k| < \infty$ ;  $\forall k = 1, 2, \dots$ . S'il existe une valeur  $\delta > 0$  telle que

$$E(|X_k - \mu_k|^{2+\delta}) < \infty; \forall k = 1, 2, \dots$$

et

$$\lim_{n \rightarrow \infty} \left[ \frac{1}{\left( \sum_{k=1}^n V(X_k) \right)^{1+\delta/2}} \cdot \sum_{k=1}^n E \left( |X_k - \mu_k|^{2+\delta} \right) \right] = 0$$

alors

$$\lim_{n \rightarrow \infty} Z_n = Z \sim \mathcal{N}(0, 1)$$

avec

$$Z_n = \frac{\sum_{k=1}^n E(X_k - \mu_k)}{\left( \sum_{k=1}^n V(X_k) \right)^{0.5}}$$

et où on a noté par  $\mathcal{N}(0, 1)$  la loi normale centrée, réduite (cf. prochain chapitre).

Pour évaluer l'erreur qu'on commet quand on applique le théorème central limite, nous avons le théorème suivant :

**THÉORÈME 2.4.3** Soit la suite  $(X_n)_n$  de v.a.i.i.d. telles que  $E(X_k) = \mu$  avec  $|\mu| < \infty$  et  $V(X_k) = \sigma^2 < \infty ; \forall k = 1, 2, \dots$  Notons

$$c_n = \frac{\sum_{k=1}^n E(|X_k - \mu|^3)}{\sigma^3}$$

Alors pour tout  $b \in \mathbb{R}$  nous avons

$$\left| P \left( \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} \leq b \right) - \frac{1}{2\pi} \int_{-\infty}^b e^{-\frac{1}{2}x^2} dx \right| \leq \frac{2c_n}{\sqrt{n}}$$

# 3

## LOIS DE PROBABILITÉS

---

3.1	Loi de Bernoulli . . . . .	16
3.2	Loi binomiale . . . . .	17
3.3	Loi multinomiale . . . . .	19
3.4	Loi binomiale négative . . . . .	20
3.5	Loi hypergéométrique . . . . .	22
3.6	Loi de Poisson . . . . .	23
3.7	Loi uniforme discrète . . . . .	25
3.8	Loi uniforme continue . . . . .	26
3.9	Loi normale centrée, réduite (standardisée) . . . . .	27
3.10	Loi normale (loi de Laplace - Gauss) . . . . .	29
3.11	Loi log-normale . . . . .	31
3.12	Loi exponentielle . . . . .	32
3.13	Loi gamma . . . . .	34
3.14	La loi du $\chi^2$ (chi deux) . . . . .	36
3.15	Loi de Student (loi de t) . . . . .	37
3.16	Loi de Fisher (loi f) . . . . .	39
3.17	Loi de Weibull . . . . .	40

---

L'étude d'une expérience aléatoire consiste au calcul des ses caractéristiques, surtout espérance mathématique et variance, à l'examen de la forme de sa fonction de densité, éventuellement à d'autres calculs en fonction de la spécificité du problème posé. Par conséquent chaque fois que nous avons une telle expérience à étudier, nous devons la répéter plusieurs fois et, en utilisant les résultats de ces répétitions, procéder aux calculs et examens. Outre le fait que les répétitions d'une expérience ne sont pas toujours possibles, il faut prendre en considération leur coût. Il est, donc, naturel de chercher à mettre ensemble différentes expériences aléatoires, en formant ainsi des classes ayant des caractéristiques statistiques identiques. Nous pouvons remarquer que cette démarche est possible à partir du moment où nous mettons dans une telle classe des expériences qui sont par leur aspect différentes,

mais pour lesquelles se dégage une identité certaine de leur nature. Ainsi, e.g. jouer à « pile ou face » ou tirer une balle d'une urne contenant des balles de deux couleurs distinctes, sont deux expériences différentes mais dont la nature est la même : chaque fois il y a un résultat parmi deux possibles.

Les lois de probabilité nous fournissent les moyens pour décrire ces classes d'expériences aléatoires car, comme nous avons vu, elles permettent de calculer aussi bien les moments que les fonctions de densité et de répartition. Nous présentons ci-après et de façon relativement détaillée, les principales lois discrètes et continues. Il est à noter qu'à l'aide de cette présentation nous pouvons aussi faire la typologie des expériences aléatoires, dans la mesure où chaque loi correspond à un type d'expérience.

## LOIS DISCRÈTES

### 3.1 Loi de Bernoulli

Une expérience aléatoire qui se produit une seule fois et qu'elle a deux résultats possibles, suit la loi de Bernoulli.

#### CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) *Nombre de résultats de l'expérience* : 2, notés  $A$  et  $B$  avec probabilités  $p$  et  $q = 1 - p$  respectivement.
- (2) *V.a.* :  $X = \begin{cases} 1 & \text{si le résultat est } A \\ 0 & \text{si le résultat est } B \end{cases}$
- (3) *Nombre de répétitions de l'expérience* : fixé d'avance et égal à 1.

#### RÉSULTATS

- (1) Probabilité :

$$P_X(X = 1) = p \quad \text{et} \quad P_X(X = 0) = q$$

- (2) Espérance mathématique

$$E(X) = p$$

- (3) Variance

$$V(X) = pq = p(1 - p)$$

- (4) Fonction caractéristique

$$\phi_X(x) = (1 - p) + pe^{jx}$$

#### NOTATION

$X$  v.a. qui suit la loi de Bernoulli :  $X \sim B(p)$ .

#### EXPÉRIENCES RELATIVES À LA LOI

- Pile ou face.
- Tirage d'une balle dans une urne contenant deux catégories de balles.
- Prélèvement à partir d'un lot, d'un article pouvant être soit bon soit défectueux.

## 3.2 Loi binomiale

La loi binomiale est une extension de la loi de Bernoulli car elle décrit le comportement d'une expérience qui a deux résultats possibles et qui est répétée plusieurs fois. Ce type d'expérience aléatoire se rencontre très souvent dans les applications.

### CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) *Nombre de résultats de l'expérience* : 2, notés  $A$  et  $B$  avec probabilités  $p$  et  $q = 1 - p$
- (2) *V.a. :  $X$  = nombre de réalisations de l'événement  $A$* , ( $X = 0, 1, \dots, n$ ).
- (3) *Nombre de répétitions de l'expérience* : fixé d'avance et égal à  $n$ .
- (4) *Nombre de réalisations de l'événement  $A$*  : variable.
- (5) *Probabilités des résultats* : Pas de modification avec le temps (i.e. tirage avec remise)<sup>1</sup>
- (6) *Répétitions de l'expérience* : Indépendantes entre elles.

### RÉSULTATS

- (1) Probabilité :

$$P_X(X = x) = C_n^x p^x (1 - p)^{n-x}, \quad x = 0, 1, \dots$$

- (2) Espérance mathématique :

$$E(X) = np$$

- (3) Variance :

$$V(X) = np(1 - p)$$

- (4) Fonction caractéristique

$$\Phi_X(x) = [(1 - p) + pe^{ix}]^n$$

- (5) Calcul itératif :

$$P_X(X = x) = \frac{n - x + 1}{x} \cdot \frac{p}{1 - p} \cdot P_X(X = x - 1)$$

### NOTATION

$X$  v.a. qui suit la loi binomiale :  $X \sim B(n, p)$ .

### TABLES

Les tables de loi binomiale sont de deux types :

---

<sup>1</sup>Dans la pratique on considère qu'un tirage est sans remise dès que  $n/N \leq 0.1$

- Tables avec des probabilités simples

$$b(x; n, p) = C_n^x p^x (1-p)^{n-x}, \text{ pour } x = 0, 1, \dots; n \geq x; p \in [0, 1].$$

- Tables avec des probabilités cumulées

$$B(r; n, p) = \sum_{x=0}^r b(x; n, p) = P(X \leq r) \text{ pour } p \in [0, 0.5].$$

Exemple  $B(6; 10, 0.5) = 0.82812$ .

Si  $p > 0.5$ , alors nous pouvons utiliser la même table car :

$$\begin{aligned} B(r; n, p) &= P_X(X \leq r) \\ &= P_X(Y \geq n - r) \\ &= 1 - P_X(Y \leq n - r - 1) \\ &= 1 - \sum_{y=0}^{n-r-1} b(y; n, 1-p) \\ &= 1 - B(n - r - 1; n, 1-p) \end{aligned}$$

où nous avons noté  $Y = n - X$  la v.a. qui suit la loi binomiale avec probabilité  $1 - p$ .

N.B. Parfois dans des livres on trouve des tables des probabilités cumulées  $P(X \geq r)$  au lieu de  $P(X \leq r)$ .

#### EXPÉRIENCES RELATIVES À LA LOI

- Une succession des « pile ou face ».
- Une succession des tirages d'une balle avec remise dans une urne contenant deux catégories de balles.
- Une succession des prélèvements avec remise, à partir d'un lot, d'un article pouvant être bon ou défectueux.

#### EXEMPLES

- (1) *Partage des ressources.* Supposons que par tranche horaire il y a  $n$  personnes qui peuvent demander une ressource quelconque (e.g. une imprimante, de l'électricité, du pain, les services d'une caissière, etc) avec probabilité  $p$  connue. En utilisant les tables de la probabilité cumulée de la distribution binomiale, nous pouvons calculer la probabilité qu'il y ait au plus  $k \leq n$  personnes qui demanderont effectivement cette ressource.
- (2) *Test de qualité.* Supposons que dans un lot contenant plusieurs articles, il y a des articles défectueux avec une probabilité  $p$  connue. En utilisant les tables de la probabilité cumulée de la distribution binomiale, nous pouvons calculer la probabilité qu'il y ait soit au plus  $k$  articles défectueux, soit exactement  $k$  articles défectueux, dans un échantillon de  $n$  articles tirés au hasard d'une population de taille  $N$ .

## PROPRIÉTÉ

- Si  $m$  variables aléatoires indépendantes  $X_1, \dots, X_m$  sont distribuées selon des lois binomiales indépendantes  $B(n_1, p), \dots,$

$B(n_m, p)$ , alors la somme  $X = X_1 + \dots + X_m$  suit la loi binomiale  $B\left(n = \sum_{i=1}^m n_i, p\right)$ .

## REMARQUES

- La fonction de densité est symétrique si  $p = q = 0.5$ .
- La variance de la loi binomiale est plus petite que sa moyenne (loi sous-dispersée).
- La loi de Bernoulli peut être notée  $B(1, p)$ .

### 3.3 Loi multinomiale

Il s'agit d'une généralisation de la loi binomiale dans le cas où le nombre de résultats de l'expérience est supérieur à deux.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) Nombre de résultats de l'expérience :  $q$ , notés  $A_k$  avec probabilités  $p_k$  ;  $k = 1, 2, \dots, q$ , avec  $\sum_{k=1}^q p_k = 1$ .
- (2) V.a. :  $X_k =$  nombre de réalisations de l'événement  $A_k$  ;  $k = 1, 2, \dots, q$ .
- (3) Nombre de répétitions de l'expérience : fixé d'avance et égal à  $n$ .
- (4) Nombre de réalisations de l'événement  $A_k$  : variable.
- (5) Probabilités des résultats : Pas de modification avec le temps (i.e. tirage avec remise)
- (6) Répétitions de l'expérience : Indépendantes entre elles.

## RÉSULTATS

- (1) Probabilité : Si on note par  $\mathbf{X} = [X_1, \dots, X_q]$ , nous avons :

$$P_{\mathbf{X}}(X_1 = x_1, \dots, X_q = x_q) = \frac{n!}{x_1! \dots x_q!} p_1^{x_1} \dots p_q^{x_q}, \quad x_k = 0, 1, \dots, n; \quad k = 1, \dots, q$$

- (2) Espérance mathématique :

$$E(X_k) = np_k; \quad k = 1, \dots, q$$

- (3) Variance :

$$V(X_k) = np_k(1 - p_k); \quad k = 1, \dots, q$$

- (4) Covariance

$$\Gamma(X_i, X_j) = -np_i p_j; \quad i \neq j, \quad i, j = 1, \dots, q$$

(5) Corrélation

$$r(X_i, X_j) = \sqrt{\frac{p_i p_j}{(1-p_i)(1-p_j)}}$$

(6) Fonction caractéristique

$$\text{On a } \phi_{X_k}(x_1, \dots, x_q) = \sum_{l=1}^q p_l e^{jx_l}, \text{ d'où on obtient } \phi_X(x_1, \dots, x_q) = \left( \sum_{l=1}^q p_l e^{jx_l} \right)^n$$

## NOTATION

$X$  v.a. qui suit la loi multinomiale :  $X \sim B(n; p_1, \dots, p_q)$

## EXPÉRIENCES RELATIVES À LA LOI

- Une succession des lancements d'un dé.
- Une succession des tirages avec remise, d'une balle d'une urne contenant plusieurs catégories de balles.
- Une succession des prélèvements avec remise, à partir d'un lot, d'un article pouvant appartenir à plusieurs classes.

## REMARQUE

- Si  $q = 2$ , on retrouve la loi binomiale  $B(n, p)$ .

### 3.4 Loi binomiale négative

Comme dans le cas de la loi binomiale, on répète plusieurs fois une expérience aléatoire qui a deux résultats possibles, mais on voudrait en plus que l'un des résultats arrive un nombre de fois fixe, défini d'avance. La loi suivie dans ce cas est la loi binomiale négative.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) *Nombre de résultats de l'expérience* : 2, notés  $A$  et  $B$  avec probabilités  $p$  et  $q = 1 - p$
- (2) *V.a. :  $X$*  = nombre de répétitions de l'expérience nécessaire pour avoir le résultat souhaité.
- (3) *Nombre de répétitions de l'expérience* : variable.
- (4) *Nombre de réalisations de l'événement  $A$*  : fixé d'avance et égal à  $c$ .
- (5) *Probabilités des résultats* : Pas de modification avec le temps (i.e. tirage avec remise)
- (6) *Répétitions de l'expérience* : Indépendantes entre elles.

## RÉSULTATS

(1) Probabilité :

$$P_X(X = x) = C_{c+x-1}^{c-1} p^c (1-p)^x, \quad x = 0, 1, \dots$$

(2) Espérance mathématique :

$$E(X) = \frac{c \cdot (1-p)}{p}$$

(3) Variance :

$$V(X) = \frac{c \cdot (1-p)}{p^2}$$

(4) Fonction caractéristique

$$\phi_X(x) = \left( \frac{p}{1 - (1-p)e^{ix}} \right)^c$$

(5) Calcul itératif :

$$P_X(X=x) = \frac{x-1}{x-c} \cdot (1-p) \cdot P_X(X=x-1)$$

#### NOTATION

$X$  v.a. qui suit la loi binomiale négative :  $X \sim B^*(c, p)$ .

#### TABLES

On utilise les tables des probabilités cumulées de la loi binomiale. En effet si on note par

$$b^*(x; c, p) = C_{n-1}^{c-1+x} p^c (1-p)^{x-c}, c = 0, 1, \dots; x \geq c; p \in [0, 1]$$

la probabilité  $P_X(X=x)$ , alors nous avons pour la probabilité cumulée

$$P_X(X \leq r) = B^*(r; c, p) = \sum_{x=c}^r b^*(x; c, p) = 1 - \sum_{x=0}^{c-1} b(x; r, p) = 1 - B(c-1, r, p)$$

#### EXPÉRIENCES RELATIVES À LA LOI

- Une succession des « pile ou face », jusqu'à avoir  $c$  fois « pile ».
- Une succession des tirages avec remise, d'une balle d'une urne contenant deux catégories de balles, jusqu'à avoir  $c$  balles d'une même couleur.

#### EXEMPLE

- *Test de qualité.* Supposons que dans un lot de  $N$  articles il y a des articles défectueux avec une probabilité  $p$  connue. En utilisant les tables de la probabilité cumulée de la distribution binomiale, nous pouvons calculer le nombre  $n$  d'articles qu'il faut extraire de ce lot pour avoir, avec une grande probabilité (e.g. 95%),  $c$  articles non défectueux.

## REMARQUES

- Si  $c = 1$ , la loi négative binomiale négative s'appelle loi de Pascal ou loi géométrique de paramètre  $p$ .
- La variance de la loi binomiale négative est plus grande que sa moyenne (loi surdispersée).

**3.5 Loi hypergéométrique**

On considère, comme pour la loi binomiale, le cas d'une expérience aléatoire qui a deux résultats possibles mais on suppose que les probabilités de réalisation de ces deux événements varient avec le temps. Ce type d'expérience est décrit par la loi hypergéométrique.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) Nombre de résultats de l'expérience : 2, notés  $A$  et  $B$  avec probabilités  $p = \frac{k}{N}$  et  $q = \frac{N-k}{N}$ .
- (2) V.a.  $X$  : nombre de réalisations de l'événement  $A$ .
- (3) Nombre de répétitions de l'expérience : fixé d'avance et égal à  $n \leq N$ .
- (4) Nombre de réalisations de l'événement  $A$  : variable.
- (5) Probabilités des résultats : Modifiées avec le temps (i.e. tirage sans remise)
- (6) Répétitions de l'expérience : Dépendantes entre elles.

## RÉSULTATS

- (1) Probabilité :

$$P_X(X = x) = \frac{C_k^x \cdot C_{N-k}^{n-x}}{C_N^n}; x = 0, 1, \dots$$

- (2) Espérance mathématique :

$$E(X) = n \frac{k}{N}$$

- (3) Variance :

$$V(X) = \frac{n k (N-k) (N-n)}{N^2 (N-1)}$$

- (4) Calcul itératif :

$$P_X(X = x) = \frac{(k-x+1)(n-x+1)}{x(N-k-n+x)} \cdot P_X(X = x-1)$$

## NOTATION

$X$  v.a. qui suit la loi hypergéométrique :  $X \sim H(N, n, k)$ .

## TABLES

Les tables de la loi hypergéométrique fournissent les probabilités cumulées : Si on note par  $p(N, n, k, x)$ ,  $x = 0, 1, \dots$  la probabilité  $P_X(X = x)$ , alors la probabilité cumulée est donnée par

$$H(N, n, k, r) = \sum_{x=\max[0, n-(N-k)]}^r p(N, n, k, x) = P_X(X \leq r).$$

Pour la lecture des tables de la loi hypergéométrique, il faut utiliser les relations :

$$p(N, n, k, x) = p(N, k, n, x),$$

$$H(N, n, k, r) = H(N, k, n, r).$$

## EXPÉRIENCES RELATIVES À LA LOI

- Une succession des tirages sans remise, d'une balle dans une urne contenant deux catégories de balles en proportion connue.
- Une succession des prélèvements sans remise, à partir d'un lot, d'un article pouvant être soit bon, soit défectueux.

## EXEMPLE

- Soit un lot contenant  $N$  articles, dont  $k$  défectueux. On prend au hasard  $n$  articles sans remise et on accepte ce lot si on a trouvé au plus  $r$  articles défectueux parmi les  $n$ . Les tables de la loi hypergéométrique nous permettent d'évaluer la probabilité d'accepter le lot en fonction du nombre  $k$  d'articles défectueux contenus dans le lot.

## REMARQUES

- Si  $N \rightarrow \infty$  et  $n$  reste fini ou si  $n$  est très petit par rapport à  $N$ , alors la loi hypergéométrique tend vers la loi binomiale  $B(n, p)$  avec  $p \approx \frac{k}{N}$ , i.e.  $\lim_{N \rightarrow \infty} P_X(X = x) = C_n^x p^x (1-p)^{n-x}$ . Ainsi la v.a.  $X$ , qui suit la loi hypergéométrique, converge en loi vers une v.a. qui suit la loi binomiale.
- Si  $n > 2$ , la variance de la loi hypergéométrique est toujours inférieure à la variance de la loi binomiale, car si on prend  $p \approx \frac{k}{N}$  nous avons :

$$\text{variance de la loi hypergéométrique} \approx npq \frac{N-n}{N-1} < npq = \text{variance de la loi binomiale}$$

sauf si  $n$  est négligeable devant  $N$ .

### 3.6 Loi de Poisson

Lorsque on s'intéresse au nombre de fois qu'un événement se réalise pendant un intervalle de temps ou au nombre d'articles de type spécifique qui se trouvent soit dans une

région d'un espace soit dans une portion d'un volume, nous pouvons utiliser, pour effectuer nos calculs, la loi de Poisson.

#### CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) *Nombre de résultats de l'expérience* : 1, noté  $A$ .
- (2) *V.a. :  $X$*  = nombre de réalisations de l'événement  $A$ .
- (3) *Nombre d'expériences* :  $n \geq 1$ .
- (4) *Unité de temps, de l'espace ou de volume relative à l'expérience* : connu (par exemple 1s ou 1 m<sup>2</sup> ou 10 cl).
- (5) *Nombre moyen de résultats  $A$  pendant une unité relative à l'expérience* : connu, noté  $\lambda$ .
- (6) *Probabilités des résultats* : Pas de modification avec le temps (i.e. tirage avec remise)
- (7) *Répétitions de l'expérience* : Indépendantes entre elles.

#### RÉSULTATS

- (1) Probabilité :

$$P_X(X = x) = \frac{e^{-\theta} \theta^x}{x!} \quad ; x = 0, 1, \dots$$

avec  $\theta = \lambda n$  ( $\theta > 0$  est le nombre moyen de réalisations de l'événement  $A$  pendant les  $n$  unités relatives à l'expérience).

- (2) Espérance mathématique :

$$E(X) = \theta$$

- (3) Variance :

$$V(X) = \theta$$

- (4) Fonction caractéristique

$$\phi_X(x) = \exp[\theta(e^{ix} - 1)]$$

- (5) Calcul itératif :

$$P_X(X = x) = \frac{\theta}{x} \cdot P_X(X = x - 1)$$

#### NOTATION

$X$  v.a. qui suit la loi de Poisson :  $X \sim P(\theta)$  ou  $X \sim P(\lambda, n)$ .

#### TABLES

Les tables de loi de Poisson sont de deux types :

- Tables avec des probabilités élémentaires  $p(x; \lambda) = e^{-\theta} \theta^x / x!$ , pour  $x = 0, 1, \dots$  ;  
 $p \in [0, 1]$  ;  $\theta > 0$ .

- Tables avec des probabilités cumulées  $P(r; \theta) = \sum_{x=0}^r p(x; \theta) = P_X(X \leq r)$ .

#### EXPÉRIENCES RELATIVES À LA LOI

- Nombre de clients arrivant devant un point de vente pendant une durée de temps fixé d'avance.
- Nombre des substances discrètes (e.g. bactéries, poissons, globules rouges, ...) dans un volume de liquide.
- Nombre d'erreurs de frappe dans une ou plusieurs pages dactylographiées.
- Nombre d'appels téléphoniques pendant une durée de temps fixée.

#### EXEMPLE

- Dans un lot contenant un grand nombre d'articles, la probabilité pour qu'un article soit défectueux est faible ( $\leq 0.1$ ) et égale à  $p$ . Supposons qu'on prend  $n$  articles au hasard. Les tables de la loi de Poisson  $P(r; \theta)$ , avec  $\theta = pn$ , nous fournissent la probabilité pour qu'il y ait au plus deux articles défectueux parmi les  $n$ .

#### PROPRIÉTÉ

- Si  $m$  variables aléatoires indépendantes  $X_1, \dots, X_m$  sont distribuées selon des lois de Poisson  $P(\theta_1, n), \dots,$

$$P(\theta_m, n), \text{ alors la somme } X = X_1 + \dots + X_m \text{ suit la loi de Poisson } P\left(\theta = \sum_{i=1}^m \theta_i, n\right).$$

#### REMARQUES

- Si la loi binomiale  $B(n, p)$  a une valeur de  $n$  modérément élevée ( $n > 10$ ) et une valeur de  $p$  petite ( $p \leq 0.1$ ), avec  $np \leq 30$ , alors la loi de Poisson  $P(n, \theta)$  avec  $\theta = np$ , est une approximation de la loi binomiale. (Si  $np > 30$ , on utilise pour l'approximation la loi normale qui sera présentée par la suite.) De même si  $p$  est grande ( $p \geq 0.9$ ), alors la loi de Poisson  $P(n, \theta)$  avec  $\theta = n(1 - p)$ , est une approximation de la loi binomiale.
- Du fait que quand la probabilité est petite, la loi de Poisson est une approximation de la loi binomiale, on l'appelle *loi des événements rares*.
- Pour la loi de Poisson nous avons  $E(X) = V(X)$ . Nous pouvons utiliser cette relation pour tester une suite de valeurs numériques prises par une v.a., afin de reconnaître si cette v.a. suit la loi de Poisson.

### 3.7 Loi uniforme discrète

Une expérience aléatoire qui se produit une seule fois et qu'elle a plusieurs résultats possibles équiprobables, suit la loi uniforme discrète

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

(1) Nombre de résultats de l'expérience :  $q$ , notés  $A_1, A_2, \dots, A_q$  avec probabilités

$$p_1 = p_2 = \dots = p_q = 1/q.$$

(2) V.a.  $\tilde{X} : X = k$  si le résultat est  $A_k$  ;  $k = 1, 2, \dots, q$ .

(3) Nombre de répétitions de l'expérience : fixé d'avance et égal à 1.

## RÉSULTATS

(1) Probabilité :

$$P_X(X = k) = \frac{1}{q} ; k = 1, 2, \dots, q$$

(2) Espérance mathématique

$$E(X) = \frac{q+1}{2}$$

(3) Variance

$$V(X) = \frac{q^2 - 1}{12}$$

## NOTATION

$X$  v.a. qui suit la loi uniforme discrète :  $X \sim U(q)$ .

## TABLES

Pour simuler le comportement d'une loi uniforme discrète on utilise des tables des nombres aléatoires. Ces tables se présentent sous la forme des nombres entiers en cinq chiffres. Chaque chiffre est identifié par les numéros de sa ligne et de sa colonne.

Pour simuler une suite de résultats d'une expérience aléatoire, e.g. 10 résultats du lancement d'un dé, on commence par un chiffre de la table des nombres aléatoires, dont le numéro de la ligne et celui de la colonne sont choisis au hasard. Ensuite on prend les neuf chiffres consécutifs et éventuellement plus, si certains des chiffres considérés sont supérieurs à 6, auquel cas on les élimine.

## EXPÉRIENCES RELATIVES À LA LOI

- Tirage au hasard d'un nombre entre 1 et  $q$ .
- Loterie.
- Création d'une suite de nombres aléatoires entre 0 et  $q - 1$ .

## LOIS CONTINUES

## 3.8 Loi uniforme continue

Une expérience aléatoire qui peut prendre n'importe quelle valeur dans l'intervalle  $[a, b]$  de façon équiprobable, suit la loi uniforme continue.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) *Résultats de l'expérience* : dans l'intervalle  $[a, b]$   
 (2) V.a. :  $X \in [a, b]$

## RÉSULTATS

- (1) Densité de probabilité :

$$f_X(x) = \frac{1}{b-a} ; x \in [a, b]$$

- (2) Espérance mathématique :

$$E(X) = \frac{a+b}{2}$$

- (3) Variance :

$$V(X) = \frac{(b-a)^2}{12}$$

- (4) Fonction caractéristique

$$\phi_X(x) = \frac{1}{(b-a)x} \sin[(b-a)x] \exp\left[j \frac{a+b}{2} x\right]$$

## NOTATION

$X$  v.a. qui suit la loi uniforme continue :  $X \sim U([a, b])$ .

## EXPÉRIENCES RELATIVES À LA LOI

– Création d'un bruit blanc.

## REMARQUE

– Si  $Y$  est une v.a. uniforme, continue sur  $[a, b]$ , alors la transformation  $X = \frac{Y-a}{b-a}$  est une v.a. uniforme, continue sur  $[0, 1]$

**3.9 Loi normale centrée, réduite (standardisée)**

La loi normale centrée, réduite a une fonction de densité avec un maximum pour la valeur  $z = 0$  et symétrique par rapport à un axe qui passe par  $z = 0$ .

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) *Résultats de l'expérience* : dans  $\mathbb{R}$   
 (2) V.a. :  $Z \in \mathbb{R}$

## RÉSULTATS

(1) Densité de probabilité :

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}; z \in \mathbb{R}$$

(2) Fonction de répartition :

$$F_Z(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\zeta^2/2} d\zeta = P_Z(Z < z)$$

(3) Espérance mathématique

$$E(Z) = 0$$

(4) Variance

$$V(Z) = 1$$

(5) Fonction caractéristique

$$\phi_Z(z) = e^{-z^2/2}$$

## NOTATION

$Z$  v.a. qui suit la loi normale réduite :  $Z \sim \mathcal{N}(0, 1)$ .

## TABLES

La table de la loi normale centrée, réduite peut être utilisée de deux façons :

(1) Comme table des valeurs de la fonction de répartition qui, pour une valeur  $z \geq 0$  de la v.a.  $Z$ , fournit la valeur de la fonction de répartition  $F_Z(z) = P_Z(Z < z)$ . Par exemple, si  $z = 1.96$ , alors  $F_Z(z) = F_Z(1.96) = 0.975$ .

En vertu de la relation  $F_Z(-z) = 1 - F_Z(z)$ , nous pouvons utiliser la même table quand  $z < 0$ . Par exemple, si  $z = -1.96$ , alors  $F_Z(z) = F_Z(-1.96) = 1 - F_Z(1.96) = 1 - 0.975 = 0.025$ .

(2) Comme table de pourcentage de la loi normale standardisée qui, pour une valeur de la fonction de répartition  $F_Z(z) = P_Z(Z < z)$  telle que  $F_Z(z) \geq 0.5$ , fournit la valeur de  $z$  de la v.a.  $Z$ . Par exemple, si  $0.975 = F_Z(z)$ , alors  $z = 1.96$ .

Du fait que si  $F_Z(z) = \zeta$ , alors  $F_Z(-z) = 1 - \zeta$ , nous pouvons utiliser la même table quand  $F_Z(z) < 0.5$ . Par exemple si  $F_Z(z) = 0.025$ , alors  $F_Z(-z) = 1 - 0.025 = 0.975$  et d'après la table nous avons :  $-z = 1.96$ , donc  $z = -1.96$ .

Pour calculer la probabilité que  $Z$  se trouve dans l'intervalle  $]z_1, z_2[$  nous utilisons la relation

$P_Z(z_1 < Z < z_2) = F_Z(z_2) - F_Z(z_1)$ . Dans le cas particulier où  $-z_1 = z_2 = z > 0$  nous avons

$$P_Z(-z < Z < z) = 2 F_Z(z) - 1.$$

## EXPÉRIENCES RELATIVES À LA LOI

- Création d'un bruit blanc gaussien.

## PROPRIÉTÉS

- La distribution normale centrée réduite  $\mathcal{N}(0, 1)$  est la limite en loi :
  - de la distribution binomiale  $P(n, p)$  quand la probabilité  $p$  est fixée et  $n \rightarrow \infty$ , i.e.

$$\lim_{n \rightarrow \infty} \frac{X - np}{\sqrt{np(1-p)}} \stackrel{\text{loi}}{=} \mathcal{N}(0, 1).$$

- de la distribution de Poisson  $P(\lambda, n)$  quand  $\lambda n \rightarrow \infty$ , i.e.

$$\lim_{\lambda n \rightarrow \infty} \frac{X - \lambda n}{\sqrt{\lambda n}} \stackrel{\text{loi}}{=} \mathcal{N}(0, 1).$$

**3.10 Loi normale (loi de Laplace - Gauss)**

La loi normale a une fonction de densité avec un maximum pour la valeur  $x = \mu$  et symétrique par rapport à un axe qui passe par  $x = \mu$ .

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) Résultats de l'expérience : dans  $\mathbb{R}$
- (2) V.a.  $\tilde{X} : X \in \mathbb{R}$

## RÉSULTATS

- (1) Densité de probabilité :

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} (x-\mu)^2}; x \in \mathbb{R}$$

- (2) Fonction de répartition :

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2\sigma^2} (\xi-\mu)^2} d\xi = P_X(X < x)$$

- (3) Espérance mathématique

$$E(X) = \mu$$

- (4) Variance

$$V(X) = \sigma^2$$

- (5) Fonction caractéristique

$$\phi_X(x) = e^{jmx} e^{-\sigma^2 x^2/2}$$

## NOTATION

$X$  v.a. qui suit la loi normale :  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

## TABLES

La table de la loi normale réduite peut être utilisée pour le calcul de la loi normale. Pour ce faire on passe de la v.a.  $X$  à la v.a. centrée, réduite

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

qui suit la loi normale réduite, si  $X$  suit la loi normale. On répercute sur  $X$  les résultats obtenus sur  $Z$  en utilisant la relation

$$X = \sigma Z + \mu$$

Ainsi pour calculer la probabilité  $P_X(x_1 < X < x_2)$  nous avons

$$P_X(x_1 < X < x_2) = F_X\left(\frac{x_2 - \mu}{\sigma}\right) - F_X\left(\frac{x_1 - \mu}{\sigma}\right) = F_Z(z_2) - F_Z(z_1)$$

## EXPÉRIENCES RELATIVES À LA LOI

- Variation de la température de l'air pendant un mois.
- Consommation journalière d'électricité.
- Variation par rapport à une longueur nominale, du diamètre des écrous fabriqués par une machine.

## PROPRIÉTÉS

- Si  $m$  variables aléatoires indépendantes  $X_1, \dots, X_m$  sont distribuées selon des lois normales  $\mathcal{N}(\mu_1, \sigma_1^2), \dots,$

$\mathcal{N}(\mu_m, \sigma_m^2)$ , alors la somme  $X = X_1 + \dots + X_m$  suit la loi normale  $\mathcal{N}\left(\mu = \sum_{i=1}^m \mu_i, \sigma^2 = \sum_{i=1}^m \sigma_i^2\right)$ .

- La densité de probabilité est symétrique :  $f_X\left(\frac{X - \mu}{\sigma}\right) = f_X\left(\frac{X + \mu}{\sigma}\right)$ .
- Soit  $m$  variables aléatoires indépendantes  $X_1, \dots, X_m$ , distribuées selon des lois normales  $\mathcal{N}(\mu_1, \sigma_1^2), \dots, \mathcal{N}(\mu_m, \sigma_m^2)$ . Nous pouvons construire le vecteur aléatoire normal  $\mathbf{X} = [X_1, \dots, X_m]^\top \sim \mathcal{N}(\boldsymbol{\mu}, \Gamma)$ , avec  $\boldsymbol{\mu} = [\mu_1, \dots, \mu_m]^\top$  et  $\Gamma$  la matrice de variance-covariance des  $m$  v.a.  $X_1, \dots, X_m$ . La fonction de densité de ce vecteur aléatoire est donnée par :

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{m/2} |\Gamma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Gamma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

où par  $|\Gamma|$  nous avons noté le déterminant de  $\Gamma$ .

## REMARQUES

- Les points d'inflexion de la densité de probabilité sont  $E(X) \pm \sqrt{V(X)}$ . Entre ces deux points se trouve 68.27% de l'aire de la fonction de répartition. Entre les points  $E(X) \pm 3\sqrt{V(X)}$  se trouve 99.73% de l'aire de la même fonction.
- La loi normale est assez proche de la fonction de l'erreur  $\operatorname{erf}x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\xi^2} d\xi$ , car  $\operatorname{erf}x = 2F_X(x\sqrt{2}) - 1$ . Elle peut, elle même, être considérée comme exprimant l'erreur qu'on fait lorsque on mesure une grandeur dont la vraie valeur est  $E(X)$  et l'instrument de la mesure nous induit à une erreur de variance  $V(X)$ .

## 3.11 Loi log-normale

La loi log-normale a une fonction de densité qui n'est pas symétrique par rapport à sa valeur moyenne. On dit qu'une v.a.  $X$  suit une loi log-normale si la v.a.  $Y = \ln(X - \alpha)$ ,  $\alpha \geq 0$ , suit une loi normale, i.e.  $Y \sim \mathcal{N}(\mu, \sigma)$ .

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) Résultats de l'expérience : dans  $\mathbb{R}_+$
- (2) V.a.  $\tilde{X} : X \in \mathbb{R}_+$

## RÉSULTATS

- (1) Densité de probabilité :

$$f_X(x) = \begin{cases} \frac{1}{(x-\alpha)\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} (\ln(x-\alpha)-\mu)^2} & \text{si } x > \alpha \\ 0 & \text{si } x \leq \alpha \end{cases}$$

- (2) Fonction de répartition :

$$\begin{aligned} F_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \frac{1}{(\xi-\alpha)} \exp\left[-\frac{1}{2\sigma^2} (\ln(\xi-\alpha)-\mu)^2\right] d\xi \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\ln(x-\alpha)} \exp\left[-\frac{1}{2\sigma^2} (y-\mu)^2\right] dy \end{aligned}$$

- (3) Espérance mathématique

$$E(X) = \alpha + \exp\left[\mu + \frac{1}{2}\sigma^2\right]$$

- (4) Variance

$$V(X) = \exp[2\mu + \sigma^2] (\exp[\sigma^2] - 1)$$

## NOTATION

$X$  v.a. qui suit la loi normale :  $X \sim \mathcal{L}(\alpha, \mu, \sigma^2)$ .

## TABLES

La table de la loi normale réduite peut être utilisée pour le calcul de la loi log-normale. Pour ce faire on passe de la v.a.  $X$  à la v.a. centrée, réduite

$$Z = \frac{\ln(X - \alpha) - \mu}{\sigma}$$

qui suit la loi normale réduite, si  $X$  suit la loi log-normale. On répercute sur  $X$  les résultats obtenus sur  $Z$  en utilisant la relation

$$X = \alpha + e^{\mu + \sigma Z}$$

Ainsi pour calculer la probabilité  $P_X(x_1 < X < x_2)$  nous avons :

$$P_X(x_1 < X < x_2) = F_X\left(\frac{\ln(x_2 - \alpha) - \mu}{\sigma}\right) - F_X\left(\frac{\ln(x_1 - \alpha) - \mu}{\sigma}\right) = F_Z(z_2) - F_Z(z_1)$$

## EXPÉRIENCES RELATIVES À LA LOI

- Distribution des revenus dans une économie.
- Développement d'un organisme vivant.
- Distribution des pannes par rapport au temps, lors du fonctionnement d'un appareil.

## REMARQUES

- La distribution log-normale dépend des trois paramètres  $\alpha$ ,  $\mu$  et  $\sigma$ . En règle générale le premier de ces paramètres est connu mais pas les deux autres. Par conséquent le calcul de l'espérance mathématique et de la variance n'est pas toujours possible. Nous pouvons, à la place des vraies valeurs, utiliser des estimations données par les formules suivantes :
- Estimation de l'espérance mathématique :

$$m_X = \alpha + \exp\left[m_Y + \frac{1}{2}s_Y^2\right]$$

- Estimation de la variance :

$$s_X^2 = \exp[2m_Y + s_Y^2] (\exp[s_Y^2] - 1)$$

où nous avons noté

$$m_Y = \frac{1}{n} \sum_{i=1}^n \ln x_i ; s_Y^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - m_Y)^2$$

### 3.12 Loi exponentielle

Considérons une expérience aléatoire  $Y$  qui suit la loi de Poisson :  $Y \sim P(\theta)$ . Le temps d'attente pour la première occurrence du résultat de l'expérience aléatoire est une autre v.a.  $X$  qui suit la loi exponentielle.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) Résultats de l'expérience : dans  $\mathbb{R}_+$   
 (2) V.a. :  $X \in \mathbb{R}_+$

## RÉSULTATS

- (1) Densité de probabilité :

$$f_X(x) = \begin{cases} \theta e^{-\theta x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} ; \theta > 0$$

- (2) Fonction de répartition :

$$F_X(x) = \begin{cases} 1 - e^{-\theta x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} ; \theta > 0$$

- (3) Espérance mathématique

$$E(X) = \frac{1}{\theta}$$

- (4) Variance

$$V(X) = \frac{1}{\theta^2}$$

- (5) Fonction caractéristique

$$\phi(x) = \left(1 - j \frac{1}{\theta} x\right)^{-1}$$

## NOTATION

$X$  v.a. qui suit la loi exponentielle :  $X \sim E(\theta)$ .

## TABLES

Nous n'avons pas besoin des tables car  $P_E(X < x) = 1 - e^{-\theta x}$  ;  $\theta \in \mathbb{R}_+$ .

## EXPÉRIENCES RELATIVES À LA LOI

- Temps d'attente avant le premier événement ou entre deux événements consécutifs.
- Temps de fonctionnement d'un appareil avant la première panne. Dans ce cas  $f_X(x)$  est la probabilité d'avoir une panne,  $x$  est le temps pendant lequel l'appareil a fonctionné,  $1/\theta$  est un paramètre caractéristique de la vie de l'appareil.

## EXEMPLE

- *Taux de défaillance.* Considérons une machine en fonctionnement et supposons que le nombre de pannes de cette machine, pour un intervalle de temps  $t$ , suit une loi de Poisson de paramètre  $\theta = \lambda t$ , i.e.

$$P(\text{nombre de pannes} = k) = e^{-\lambda t} \cdot \frac{(\lambda t)^k}{k!}$$

Si on cherche la probabilité de n'avoir pas de pannes, alors on a :  $P(k=0) = e^{-\lambda t}$ . On peut donc dans ce cas considérer la loi exponentielle avec, à la place du paramètre  $\theta = \lambda t$ , le paramètre  $\lambda$ . En fiabilité le paramètre  $\lambda$  est le taux de défaillance, car, si on considère que  $\lambda$  est fonction du temps, alors nous avons :

$$P(T \leq t < T + dT \mid t \geq T) = \frac{\frac{d}{dt}F(t)}{1 - F(t)} = \frac{\lambda(t) e^{-\int_0^t \lambda(t) dt}}{1 - \left(1 - e^{-\int_0^t \lambda(t) dt}\right)} = \lambda(t)$$

où  $P(T \leq t < T + dT \mid t \geq T)$  exprime la probabilité qu'une machine ait une panne entre  $T$  et  $T + dT$ , sous l'hypothèse qu'elle était en fonctionnement jusqu'à l'instant  $T$ . L'espérance mathématique  $E(t)$  est le MTBF (mean time between failures - temps moyen entre deux pannes). De plus la fonction de répartition  $F(t) = 1 - e^{-\lambda t}$  s'appelle fonction cumulée de défaillance et on définit aussi la fonction  $R(t) = 1 - F(t) = e^{-\lambda t}$  qui est la fonction de fiabilité, c'est-à-dire, comme nous venons de le voir, la probabilité de ne pas avoir de pannes.

## REMARQUES

- Bien que la présentation de la loi exponentielle est fondée sur le temps d'attente, on peut imaginer d'appliquer cette loi pour évaluer la probabilité de parcourir une distance avant la première occurrence d'un événement d'une expérience aléatoire qui suit la loi de Poisson. Si  $R$  représente soit la région, soit le volume sous examen, alors la fonction de répartition est

$$F_X(x) = \begin{cases} 1 - e^{-\theta R x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} ; \theta > 0$$

- La loi exponentielle est plus adaptée pour modéliser les pannes des ensembles complets. Par contre pour les pannes des modules la loi de Weibull (cf. infra) est plus appropriée.

## 3.13 Loi gamma

La loi exponentielle est un cas particulier d'un type plus général de loi, la loi gamma.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) *Résultats de l'expérience* : dans  $\mathbb{R}_+$

(2) V.a. :  $X \in \mathbb{R}_+$

#### RÉSULTATS

(1) Densité de probabilité :

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} ; \alpha > 0, \beta > 0$$

avec  $\Gamma(\alpha)$  la fonction gamma  $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$ .

(2) Fonction de répartition :

$$F_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^x y^{\alpha-1} e^{-y/\beta} dy & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases} ; \theta > 0$$

(3) Espérance mathématique

$$E(X) = \alpha\beta$$

(4) Variance

$$V(X) = \alpha\beta^2$$

(5) Fonction caractéristique

$$\phi_X(x) = \frac{1}{(1 - j\beta x)^\alpha}$$

#### NOTATION

$X$  v.a. qui suit la loi gamma :  $X \sim G(\alpha, \beta)$ .

#### PROPRIÉTÉS

– Si  $m$  variables aléatoires indépendantes  $X_1, \dots, X_m$  sont distribuées selon des lois gamma

$G(\alpha_1, \beta), \dots, G(\alpha_m, \beta)$ , alors la somme  $X = X_1 + \dots + X_m$  suit la loi gamma

$$G\left(\alpha = \sum_{i=1}^m \alpha_i, \beta\right).$$

– Si  $X \sim G(\alpha, \beta)$ , alors  $\lim_{\alpha \rightarrow +\infty} \frac{X - \alpha}{\sqrt{\alpha}} = \mathcal{N}(0, 1)$

#### REMARQUES

– Nous avons

$$- \Gamma(1) = 1 ; \Gamma(1/2) = \sqrt{\pi}$$

$$- \Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$$

$$- \Gamma(\alpha + 1) = \alpha !$$

– La loi exponentielle  $E(\theta)$  est une loi gamma  $G(1, 1/\theta)$ .

### 3.14 La loi du $\chi^2$ (chi deux)

Considérons  $m$  v.a. mutuellement indépendantes qui suivent la loi normale  $\mathcal{N}(\mu_i, \sigma_i^2)$ ;  $i = 1, 2, \dots, m$ . La somme des carrés des v.a. centrées, réduites, correspondantes :

$$u = \sum_{i=1}^m \left( \frac{x_i - \mu_i}{\sigma_i} \right)^2$$

est une variable aléatoire  $U$  qui suit la loi du  $\chi^2$  avec  $m$  degrés de liberté (ddl). Le nombre de degrés de liberté sera noté  $v$ .

#### CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) Résultats de l'expérience : dans  $\mathbb{R}$
- (2) V.a.  $\tilde{\cdot} : U \in \mathbb{R}_+$

#### RÉSULTATS

- (1) Densité de probabilité :

$$f_U(u) = \begin{cases} \frac{1}{[(v/2)-1]!} \cdot \frac{1}{2^{v/2}} \cdot u^{(v/2)-1} \cdot e^{-u/2} & \text{si } u \geq 0 \\ 0 & \text{si } u < 0 \end{cases}$$

- (2) Fonction de répartition :

$$F_U(u) = \frac{1}{[(v/2)-1]!} \cdot \frac{1}{2^{v/2}} \int_0^u \xi^{(v/2)-1} \cdot e^{-\xi/2} d\xi = P_X(U < u)$$

- (3) Espérance mathématique

$$E(U) = v$$

- (4) Variance

$$V(U) = 2v$$

- (5) Fonction caractéristique

$$\phi_U(u) = \frac{1}{(1 - 2ju)^{m/2}}$$

#### NOTATION

$U$  v.a. qui suit la loi du  $\chi^2$  à  $v$  ddl :  $U \sim \chi^2(v)$ .

#### TABLES

La table du  $\chi^2$  fournit la valeur  $\chi^2(v; p)$  de la v.a.  $U$  à  $v$  ddl qui a la probabilité  $p$  de ne pas être dépassée, i.e.

$$p = \frac{1}{[(v/2)-1]!} \cdot \frac{1}{2^{v/2}} \int_0^{\chi^2(v;p)} \xi^{(v/2)-1} \cdot e^{-\xi/2} d\xi$$

Par exemple si  $v = 20$ ;  $p = 0.5$ , alors  $\chi^2(v; p) = 19.34$ .

Pour des valeurs de  $v$  supérieures à 30 nous pouvons calculer  $\chi^2(v; p)$  en utilisant l'approximation

$$\chi^2(v; p) = v \cdot \left[ 1 - \frac{2}{9v} + Z_p \sqrt{\frac{2}{9v}} \right]^2$$

où  $Z_p$  la valeur de la v.a. normale réduite qui ne peut pas être dépassée avec probabilité  $p$ , i.e.  $F_Z(Z_p) = p$ .

De même pour des valeurs de  $v$  supérieures à 100 nous pouvons calculer  $\chi^2(v; p)$  en utilisant l'approximation

$$\chi^2(v; p) = \frac{1}{2} \cdot \left( Z_p + \sqrt{2v-1} \right)^2$$

#### EXPÉRIENCES RELATIVES À LA LOI

- Test de conformité d'une suite de valeurs avec une distribution théorique.
- Comparaison des variances.

#### PROPRIÉTÉS

- Si  $Z \sim \mathcal{N}(0, 1)$ , alors  $Z^2 \sim \chi^2(1)$ .
- Si  $U_i \sim \chi^2(v_i)$ ;  $i = 1, \dots, n$  v.a. mutuellement indépendantes, alors
 
$$U = \sum_{i=1}^n U_i \sim \chi^2 \left( \sum_{i=1}^n v_i \right).$$
- Considérons le vecteur aléatoire  $\mathbf{X} = [X_1, \dots, X_m]^\top \sim \mathcal{N}(\mu, \Gamma)$ . La variable aléatoire  $Y = (\mathbf{x} - \mu)^\top \Gamma^{-1} (\mathbf{x} - \mu)$  suit une loi du  $\chi^2(m)$ .

#### REMARQUES

- La loi du  $\chi^2(v)$  est une loi gamma  $G(v/2, 2)$ .
- Si  $X \sim E(\theta)$ , alors  $U = 2X/\theta \sim \chi^2(2)$ .
- Si  $X \sim U(]0, 1[)$ , alors  $U = -2 \log X \sim \chi^2(2)$ .
- La loi du  $\chi^2$  est utilisée lors de l'étude de la variance empirique  $s^2$  d'une population qui suit la loi normale. Elle est aussi utilisée pour des tests d'association sur des tableaux de contingence.

### 3.15 Loi de Student (loi de t)

Si  $Z \sim \mathcal{N}(0, 1)$  et  $U \sim \chi^2(v)$  alors la v.a.

$$T = \frac{Z}{\sqrt{U/v}}$$

suit la loi de Student avec  $v$  ddl.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) Résultats de l'expérience : dans  $\mathbb{R}$   
 (2) V.a.  $\tilde{\tau} : T \in \mathbb{R}$

## RÉSULTATS

- (1) Densité de probabilité :

$$f_T(t) = \frac{1}{\sqrt{\pi v}} \cdot \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)} \cdot \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2} ; t \in \mathbb{R}, v > 0$$

- (2) Fonction de répartition :

$$F_T(t) = \frac{1}{\sqrt{v} B(1/2, v/2)} \int_{-\infty}^t \left(1 + \frac{x^2}{v}\right)^{-(v+1)/2} dx = P_T(T < t)$$

où nous avons noté

$$x = \frac{v}{v+t^2}$$

et  $B(\alpha, \beta)$  est la fonction beta

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)} = B(\beta, \alpha)$$

- (3) Espérance mathématique

$$E(T) = 0$$

- (4) Variance

$$V(T) = \frac{v}{v-2}$$

- (5) Fonction caractéristique

$$\phi(t) = \frac{1}{\pi \Gamma(v/2)} \left(\frac{|t|}{2\sqrt{v}}\right)^{v/2} T_{v/2}\left(\frac{|t|}{\sqrt{v}}\right)$$

## NOTATION

$T$  v.a. qui suit la loi de Student à  $v$  ddl :  $T \sim t(v)$ .

## TABLES

La table de la loi de Student fournit les valeurs de  $t$  telles que  $F_T(t) = P_T(T(v) < t)$  pour  $v$  et  $p$  fixés.

Du fait que  $F_T(-t) = 1 - F_T(t)$  on peut utiliser la même table pour des valeurs de  $t$  négatives.

Pour des valeurs de  $v$  supérieures à 30 on peut utiliser la loi normale qui offre une bonne approximation.

## EXPÉRIENCES RELATIVES À LA LOI

- Comparaison des moyennes.

## REMARQUES

- La fonction de densité  $f_T(t)$  est symétrique par rapport à  $t = 0$ .
- Si  $T \sim t(\nu)$ , alors  $\lim_{\nu \rightarrow \infty} T \stackrel{L}{=} \mathcal{N}(0, 1)$ .
- La loi de Student est utilisée à l'étude des moyennes empiriques d'une population quand sa variance est inconnue.

**3.16 Loi de Fisher (loi f)**

Si  $U_i \sim \chi^2(\nu_i)$  avec  $i = 1, 2$  alors la v.a.

$$F = \frac{U_1/\nu_1}{U_2/\nu_2}$$

suit la loi de Fisher avec  $\nu_1$  et  $\nu_2$  ddl.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

- (1) Résultats de l'expérience : dans  $\mathbb{R}$
- (2) V.a.  $\tilde{\cdot} : F \in \mathbb{R}$

## RÉSULTATS

- (1) Densité de probabilité :

$$f_F(x) = \frac{[(\nu_1 + \nu_2 - 2) / 2]!}{[(\nu_1 - 2) / 2]! \cdot [(\nu_2 - 2) / 2]!} \cdot \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1 / 2} \cdot \frac{x^{(\nu_1 - 2) / 2}}{(1 + \nu_1 x / \nu_2)^{(\nu_1 + \nu_2) / 2}} ; x \in \mathbb{R}_+, \nu_1, \nu_2 > 0$$

- (2) Espérance mathématique

$$E(F) = \frac{\nu_2}{\nu_2 - 2}$$

- (3) Variance

$$V(F) = \frac{2 \nu_2^2 (\nu_1 + \nu_2 - 2)}{\nu_1 (\nu_2 - 2)^2 (\nu_2 - 4)} ; \nu_2 > 4$$

## NOTATION

$F$  v.a. qui suit la loi de Fisher à  $\nu_1$  et  $\nu_2$  ddl :  $F \sim F(\nu_1, \nu_2)$ .

## TABLES

La table de la loi de Fisher fournit les valeurs  $F(p; \nu_1, \nu_2)$  pour  $p \leq 0.5$ . Si  $p > 0.5$  on peut utiliser la même table car

$$F(p; \nu_1, \nu_2) = \frac{1}{F(1 - p; \nu_2, \nu_1)}$$

## EXPÉRIENCES RELATIVES À LA LOI

– Analyse de variance.

## REMARQUES

– Soit la v.a.  $F$  qui suit la loi de Fisher à  $\nu_1$  et  $\nu_2$  ddl :  $F \sim F(\nu_1, \nu_2)$ . Nous avons :

$$\lim_{\nu_1, \nu_2 \rightarrow \infty} \frac{\frac{1}{2} \left( \ln F + \frac{1}{\nu_1} - \frac{1}{\nu_2} \right)}{\sqrt{\frac{1}{2} \left( \frac{1}{\nu_1} + \frac{1}{\nu_2} \right)}} \stackrel{I}{=} \mathcal{N}(0, 1)$$

– La loi de Fisher est utilisée lorsque on désire comparer les variances empiriques de deux populations, pendant l'analyse de variance et aussi quand on procède à une régression multilinéaire.

## 3.17 Loi de Weibull

Soit une v.a.  $Y$  qui suit la loi exponentielle  $E\left(\frac{!}{(\theta - x_0)}\right)$  de paramètre  $(\theta - x_0)^{-1}$ , avec  $\theta - x_0 > 0$ . Alors la v.a.  $X = Y^b, b > 0$  suit la loi de Weibull.

## CADRE GÉNÉRAL DE L'EXPÉRIENCE

(1) Résultats de l'expérience : dans  $\mathbb{R}_+$

(2) V.a.  $\tilde{\cdot} : X \in \mathbb{R}_+$

## RÉSULTATS

(1) Densité de probabilité :

$$f_X(x) = \frac{b}{\theta - x_0} \left( \frac{x - x_0}{\theta - x_0} \right)^{b-1} \cdot \exp \left[ - \left( \frac{x - x_0}{\theta - x_0} \right)^b \right]; x, \theta, x_0, b (\theta - x_0) \in \mathbb{R}_+$$

Les différents paramètres utilisés pour le calcul de la fonction de densité sont :

- $b$  paramètre de forme ou, encore, de pente (sans unités). Il influe sur la forme de la courbe de la distribution.
- $x_0$  paramètre de position ou de localisation. Il exprime la plus petite valeur que peut prendre la v.a.  $X$  et de ce fait il caractérise l'origine de la distribution. Le plus souvent il est pris égal à zéro.
- $\theta$  paramètre d'échelle (exprimé selon les mêmes unités que la v.a.  $X$ ). Dans le cas des études de fiabilité, il est caractéristique de la durée de vie des appareils.

(2) Fonction de répartition :

$$F_X(x) = 1 - \exp \left[ - \left( \frac{x - x_0}{\theta - x_0} \right)^b \right]$$

(3) Espérance mathématique

$$E(X) = x_0 + (\theta - x_0) \Gamma\left(1 + \frac{1}{b}\right)$$

(4) Variance

$$V(X) = (\theta - x_0)^2 \left[ \Gamma\left(1 + \frac{2}{b}\right) - \Gamma^2\left(1 + \frac{1}{b}\right) \right]$$

#### NOTATION

$X$  v.a. qui suit la loi de Weibull :  $X \sim W(b, \theta, x_0)$ .

#### TABLES

Le calcul de la loi de Weibull peut se faire en utilisant la loi exponentielle, car si  $X \sim W(b, \theta, x_0)$ , alors la v.a.  $Y = \sqrt[b]{X}$  suit la loi exponentielle  $E\left(\frac{1}{\theta - x_0}\right)$  et  $P_E(Y < y) = 1 - e^{-y/(\theta - x_0)}$ .

Si on veut calculer l'espérance et la variance de la loi de Weibull pour un triplet  $b, \theta, x_0$  donné, on utilise des tables qui fournissent les valeurs de deux paramètres  $A$  et  $B$  pour chaque valeur de  $b$ . On obtient l'espérance et l'écart-type par les formules :

$$E(X) = Ab + x_0 \quad ; \quad \sigma_X = B(\theta - x_0)$$

#### EXPÉRIENCES RELATIVES À LA LOI

- Problèmes de fiabilité des modules de différents systèmes.
- Problèmes relatifs à l'usure et à la dégradation de systèmes.

#### REMARQUES

- Si  $b = 1$ , alors la distribution de Weibull est identique à la distribution exponentielle  $E\left(\frac{1}{\theta - x_0}\right)$ . Si  $b = 4$ , alors les distributions normale et de Weibull coïncident.
- En fiabilité on modélise les pannes de jeunesse en posant  $b < 1$  et les pannes de vieillesse en posant  $b > 1$ .
- Dans l'étude du taux de pannes d'une machine, il est raisonnable de prendre  $x_0 = 0$ . Nous avons ainsi une loi de Weibull à deux paramètres  $b$  et  $\theta$ . Dans ce cas la fonction de répartition s'écrit

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\theta}\right)^b\right]$$

d'où on obtient

$$\ln \frac{1}{1 - F(x)} = \left(\frac{x}{\theta}\right)^b \Rightarrow \ln \left[ \ln \frac{1}{1 - F(x)} \right] = b(\ln x) - (b \ln \theta)$$

qui est une équation linéaire du type  $Y = bX + c$ , i.e. une droite dont la pente est  $b$ .

Cette approche permet aussi d'avoir une relation entre  $b$  et  $\theta$  et une caractérisation du rôle joué par  $\theta$ .

Supposons que  $q\%$  d'appareils ont une durée de vie inférieure ou égale à  $B_q$ . Donc, si  $B_q$  est une durée de vie, alors  $q$  est la probabilité de défaillance. Si on prend  $q = 50\%$ , alors, du fait que  $F(x) = 0.5 =$  probabilité de défaillance, on a

$$0.5 = 1 - \exp \left[ - \left( \frac{B_{50}}{\theta} \right)^b \right] \Rightarrow \exp \left[ \left( \frac{B_{50}}{\theta} \right)^b \right] = 2 \Rightarrow \theta = \frac{B_{50}}{(\ln 2)^{1/b}}$$

Posons maintenant  $x = \theta$  et on obtient

$$F(\theta) = 1 - \exp \left[ - \left( \frac{\theta}{\theta} \right)^b \right] = 1 - e^{-1} = 1 - \frac{1}{2.718} = 0.632$$

Donc  $\theta$  est la durée de vie pendant laquelle 63.2% d'appareils sont tombés en panne.

# 4

## APPROXIMATIONS DES LOIS

---

4.1	Approximation des lois en utilisant la loi uniforme . . .	43
4.2	Approximation de $P(a < S_n \leq b)$ . . . . .	44
4.3	Approximation d'une loi discrète . . . . .	45
4.3.1	Approximation de la loi binomiale . . . . .	45
4.3.2	Approximation de la loi de Poisson . . . . .	46

---

Bien que nous avons déjà examiner l'approximation de la loi binomiale ou de la loi de Poisson par la loi normale, le cadre théorique de l'approximation d'une loi par une autre loi a comme fondement le T.C.L. Nous présentons ci-après les résultats fondamentaux de ces approximations.

### 4.1 Approximation des lois en utilisant la loi uniforme

Considérons une v.a.  $X$  qui suit la loi de probabilité  $\mathcal{L}$  de fonction de répartition  $F$ . On suppose que  $F$  est continue à droite, i.e. que  $F(x) = P(X \leq x)$ . On cherche à déterminer une suite des valeurs numériques de la v.a.  $X$ , c'est-à-dire une suite de nombres qui sont répartis selon la loi  $\mathcal{L}$ . Le théorème suivant fournit une méthode de construction.

**THÉORÈME 4.1.1** *Soit une v.a.  $X$  qui suit la loi de probabilité  $\mathcal{L}$  de fonction de répartition  $F$ , continue. Considérons la fonction  $G : ]0, 1] \rightarrow \mathbb{R}$ , définie par la relation*

$$G(y) = \inf \{x \in \mathbb{R} \mid F(x) \geq y\} \quad ; \quad y \in ]0, 1] \tag{4.1.1}$$

*Alors, si  $Y$  est une v.a. qui suit la loi uniforme  $U(0, 1)$ , la v.a.  $G(Y)$  suit la loi  $\mathcal{L}$  de fonction de répartition  $F$ , c'est-à-dire elle représente la v.a.  $X$ .*

Dém. Considérons  $y \in ]0, 1]$  et posons  $B(y) = \{x \in \mathbb{R} \mid F(x) \geq y\} \subset \mathbb{R}$ . Du fait que  $F$  est continue et croissante,  $B(y)$  est en réalité un intervalle de  $\mathbb{R}$  de la forme  $[a, +\infty[$  car  $F$  est continue à droite. Donc si  $F(x) \geq y$  alors  $G(y) \leq x$  et vice versa. Par conséquent la

fonction de répartition de la v.a.  $G(Y)$  qui est donnée par la relation  $F_G(x) = P(G(Y) \leq x)$ , est égale à  $F_G(x) = P(Y \leq F(x))$  et, comme  $Y \sim U(0, 1)$ , nous avons  $P(Y \leq F(x)) = F(x)$ . D'où finalement  $P(G(Y) \leq x) = F(x)$ , i.e. la v.a.  $G(Y)$  suit la loi  $\mathcal{L}$ .

D'après ce théorème, nous pouvons utiliser les valeurs d'une v.a. qui suit la loi uniforme  $U(0, 1)$  pour engendrer d'autres v.a. qui suivent une autre loi  $\mathcal{L}$ .

Comme exemple d'application nous allons utiliser la loi uniforme pour engendrer des valeurs qui suivent la loi normale réduite.

Soit  $(X_n)_n$  suite de v.a. qui suivent la loi uniforme  $U(0, 1)$ . Nous savons que l'espérance mathématique de cette loi est  $E(X_k) = \frac{1}{2}$  et la variance  $V(X_k) = \frac{1}{12}$ . Par conséquent la v.a.  $S_n = \sum_{k=1}^n X_k$  a comme espérance mathématique  $E(S_n) = \frac{n}{2}$  et variance  $V(S_n) = \frac{n}{12}$ . La v.a. standardisée

$$Z_n = \frac{S_n - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

suit la loi normale  $\mathcal{N}(0, 1)$ . Donc la v.a.  $S_n$  suit la loi  $\mathcal{N}(0, \frac{n}{12})$ .

Ce résultat nous permet d'élaborer l'algorithme suivant pour créer des éléments de la loi normale  $\mathcal{N}(0, 1)$  :

- (1) On extrait douze fois un élément de la loi uniforme  $U(0, 1)$  (c'est-à-dire un nombre distribué équiprobablement dans l'intervalle  $[0, 1]$ ).
- (2) On fait la somme de ces douze éléments.
- (3) De cette somme on retranche la valeur 6.

Le résultat obtenu est un élément de la loi normale  $\mathcal{N}(0, 1)$ , que l'on appelle souvent bruit blanc gaussien (c'est-à-dire un nombre distribué selon la loi normale dans l'intervalle  $[0, 1]$ ).

## 4.2 Approximation de $P(a < S_n \leq b)$

Soit la suite  $(X_n)_n$  de v.a.i.i.d. telles que  $E(X_k) = \mu$  avec  $|\mu| < \infty$  et  $V(X_k) = \sigma^2 < \infty$ ;  $\forall k = 1, 2, \dots$ . La v.a.  $S_n = \sum_{k=1}^n X_k$  a comme espérance mathématique  $E(S_n) = n\mu$  et variance  $V(S_n) = \sigma^2 n$ . Nous avons par conséquent :

$$\begin{aligned} P(a < S_n \leq b) &= P\left(\frac{a-n\mu}{\sigma\sqrt{n}} < \frac{S_n-n\mu}{\sigma\sqrt{n}} \leq \frac{b-n\mu}{\sigma\sqrt{n}}\right) \\ &= P\left(\frac{S_n-n\mu}{\sigma\sqrt{n}} \leq \frac{b-n\mu}{\sigma\sqrt{n}}\right) - P\left(\frac{S_n-n\mu}{\sigma\sqrt{n}} \leq \frac{a-n\mu}{\sigma\sqrt{n}}\right) \end{aligned} \quad (4.2.1)$$

Comme

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \quad (4.2.2)$$

est une v.a. standardisée, nous avons, d'après le TCL, que  $Z_n \sim \mathcal{N}(0, 1)$ . Donc

$$P(a < S_n \leq b) = \frac{1}{2\pi} \left[ \int_{-\infty}^{\frac{b-n\mu}{\sigma\sqrt{n}}} e^{-\frac{1}{2}x^2} dx - \int_{-\infty}^{\frac{a-n\mu}{\sigma\sqrt{n}}} e^{-\frac{1}{2}x^2} dx \right] \Rightarrow \quad (4.2.3)$$

$$P(a < S_n \leq b) = \frac{1}{2\pi} \int_{\frac{a-n\mu}{\sigma\sqrt{n}}}^{\frac{b-n\mu}{\sigma\sqrt{n}}} e^{-\frac{1}{2}x^2} dx \quad (4.2.4)$$

### 4.3 Approximation d'une loi discrète

Soit la suite  $(X_n)_n$  de v.a. qui suivent une loi discrète, indexée par  $kh$ . Nous avons

$$P(a < S_n \leq b) = P(a+h < S_n \leq b) \quad (4.3.1)$$

Si on approxime la loi discrète par la loi continue  $f$  nous avons :

$$P(a+h < S_n \leq b) = \sum_{k=a+h}^b f(k) \quad (4.3.2)$$

où  $f(k)$  est l'aire du rectangle centré à  $k$  avec base  $h$ . Donc

$$P(S_n \leq b) = \text{aire de l'histogramme jusqu'à } b + \frac{h}{2}$$

Le terme  $\frac{h}{2}$  est la correction de la continuité. De même à gauche il ne faut pas prendre  $a$  mais  $a - \frac{h}{2}$ . (4.2.4) s'écrit :

$$P(a < S_n \leq b) = \frac{1}{2\pi} \int_{\frac{a-\frac{h}{2}-n\mu}{\sigma\sqrt{n}}}^{\frac{b+\frac{h}{2}-n\mu}{\sigma\sqrt{n}}} e^{-\frac{1}{2}x^2} dx \quad (4.3.3)$$

#### 4.3.1 Approximation de la loi binomiale

Soit la suite  $(X_n)_n$  de v.a. qui suivent la loi de Bernoulli  $B(p)$ , dont l'espérance mathématique est  $p$  et la variance  $p(1-p)$ . Alors la v.a.  $S_n = \sum_{k=1}^n X_k$  a comme espérance mathématique  $E(S_n) = np$  et variance  $V(S_n) = np(1-p)$ , i.e.  $S_n$  suit, comme on devait s'attendre, la loi binomiale  $B(n, p)$ . Cette loi est approximée par la loi normale  $\mathcal{N}(np, np(1-p))$ . Par conséquent nous avons

– sans correction

$$P(a < S_n \leq b) = \frac{1}{2\pi} \int_{\frac{a-n\mu}{\sigma\sqrt{n}}}^{\frac{b-n\mu}{\sigma\sqrt{n}}} e^{-\frac{1}{2}x^2} dx \quad (4.3.4)$$

– avec correction

$$P(a < S_n \leq b) = P((a+1) < S_n \leq b) = \frac{1}{2\pi} \int_{\frac{a-0.5-n\mu}{\sigma\sqrt{n}}}^{\frac{b+0.5-n\mu}{\sigma\sqrt{n}}} e^{-\frac{1}{2}x^2} dx \quad (4.3.5)$$

### 4.3.2 Approximation de la loi de Poisson

Soit la suite  $(X_n)_n$  de v.a. qui suivent la loi de Poisson  $P(\lambda)$ . Alors la v.a.  $S_n = \sum_{k=1}^n X_k$  est approximée par la loi normale  $\mathcal{N}(n\lambda, n\lambda)$ . Par conséquent nous avons

– sans correction

$$P(a < S_n \leq b) = \frac{1}{2\pi} \int_{\frac{a-n\lambda}{\sigma\sqrt{n\lambda}}}^{\frac{b-n\lambda}{\sigma\sqrt{n\lambda}}} e^{-\frac{1}{2}x^2} dx \quad (4.3.6)$$

– avec correction

$$P((a+1) < S_n \leq b) = \frac{1}{2\pi} \int_{\frac{a-0.5-n\lambda}{\sigma\sqrt{n\lambda}}}^{\frac{b+0.5-n\lambda}{\sigma\sqrt{n\lambda}}} e^{-\frac{1}{2}x^2} dx \quad (4.3.7)$$

# 5

## ÉCHANTILLONNAGE

---

5.1	Les relations entre population et échantillon . . . . .	47
5.2	Échantillon avec des v.a. normales . . . . .	48

---

Dans la statistique monodimensionnelle l'étude d'un phénomène commence par l'étude de la loi de probabilité qui approche mieux son comportement. Supposons que nous avons établi que les résultats des expériences aléatoires relatives au phénomène sous étude suivent la loi de probabilité  $\mathcal{L}$ . Alors le résultat de la  $k$ -ième expérience sera représenté par une variable aléatoire  $X_k$  qui suit la loi  $\mathcal{L}$ . L'étude statistique ne se fera pas sur la totalité des expériences, c'est-à-dire sur la population, mais sur un sous-ensemble de celle-ci, l'échantillon qui formellement est un ensemble d'expériences indépendantes et dont les résultats - considérés comme des réalisations des variables aléatoires - suivent la même loi  $\mathcal{L}$ .

Ce chapitre présente les principaux résultats concernant l'échantillonnage.

### 5.1 Les relations entre population et échantillon

Considérons une population  $\mathcal{P}$  contenant un nombre  $N$  de résultats d'expériences aléatoires concernant un phénomène particulier. Supposons que nous voulons étudier du point de vue statistique ces résultats. On peut supposer que chacune de ces  $N$  expériences peut être représentée par une variable aléatoire  $X_k, k = 1, \dots, N$ . Le résultat de la  $k$ -ième

expérience sera noté par  $x_k$  et on aura  $X_k = x_k, k = 1, \dots, N$ .

En règle générale nous ne pouvons pas utiliser pour les calculs toute la population, car soit la taille de la population  $N$  est très grande, soit la population totale n'est pas disponible. On utilise donc ce qu'on appelle un *échantillon*, i.e. un sous-ensemble de la population de taille  $n$  (avec  $n \ll N$ ). Cette utilisation repose sur un certain nombre de règles, parce qu'il faut qu'il y ait concordance entre les grandeurs statistiques (e.g. espérance mathématique, variance, ...) de la population et les mêmes grandeurs calculées en utilisant l'échantillon et qui sont appelés, de ce fait, empiriques.

Un échantillon de taille  $n$  est modélisé par la donnée de  $n$  variables aléatoires  $(X_k)_n$  indépendantes et identiquement distribuées (v.a.i.d.).

Le tableau suivant fournit la correspondance des termes et des notations utilisés en probabilités (population) et statistique (échantillon)

Nous donnons en terminant ce paragraphe le comportement asymptotique des grandeurs statistiques empiriques.

Pour la moyenne empirique nous avons les résultats suivants :

$$\lim_{n \rightarrow \infty} \bar{X}_n \underset{ps}{=} \mu_X \quad ; \quad \lim_{n \rightarrow \infty} \bar{X}_n \underset{p}{=} \mu_X \quad ; \quad \lim_{n \rightarrow \infty} \sqrt{n} \frac{\bar{X}_n - \mu_X}{\sigma_X} \underset{l}{=} \mathcal{N}(0, 1)$$

Pour la variance empirique nous avons si  $E(X^2) < +\infty$  :

$$\lim_{n \rightarrow \infty} S_n^2 \underset{ps}{=} \sigma_X^2$$

## 5.2 Échantillon avec des v.a. normales

Soit un échantillon représenté par  $n$  v.a.i.d.  $(X_k)_n$  avec  $X_k \sim \mathcal{N}(\mu, \sigma^2)$ . Dans ce cas il est possible de connaître la loi suivie par  $\bar{X}_n$  et  $S_n^2$ .

En effet nous avons :

– Pour la moyenne empirique :

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

ou, encore

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Par conséquent pour l'étude de la moyenne empirique  $\bar{X}_n$  d'un échantillon qui suit la loi normale dont on connaît sa moyenne  $\mu$  et sa variance  $\sigma^2$ , il faut utiliser la loi normale  $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$

– Pour la variance empirique :

$$(n-1) \frac{S_n^2}{\sigma^2} \sim \chi^2(n-1)$$

où nous avons utilisé la relation  $s_n^2 = \frac{1}{n-1} \sum_{l=1}^n (x_l - \bar{x}_n)^2$

On voit ainsi que pour l'étude de la variance empirique  $s_n^2$  d'un échantillon qui suit la loi normale, il faut utiliser la loi du  $\chi^2$ . Bien sûr, d'après la formule ci-dessus, cette étude ne peut se faire que si on connaît la variance  $\sigma^2$  de la population.

Dans le cas où cette dernière est inconnue, nous procédons comme suit : En utilisant les deux dernières relations nous avons

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} / \frac{s_n}{\sigma} \sim \mathcal{N}(0, 1) / \sqrt{\frac{\chi^2(n-1)}{n-1}}$$

où le numérateur et le dénominateur du second membre sont des v.a. indépendantes. Par conséquent

$$\sqrt{n} \frac{\bar{X}_n - \mu}{s_n} \sim t(n-1)$$

où  $t(n-1)$  loi de Student à  $n-1$  degrés de liberté.

Ainsi dans le cas où la variance  $\sigma^2$  de la population est inconnue, nous pouvons utiliser pour l'étude de la variance empirique  $S_n^2$  la loi de Student.

Un dernier résultat est que si les v.a. de l'échantillon suivent la loi normale, alors  $\bar{X}_n$  et  $S_n^2$  sont des variables aléatoires indépendantes.

L'ensemble des ces résultats sera utilisé lors de l'élaboration des tests statistiques.

Paramètre	Échantillon	Population
<i>Effectifs</i>	$n$	$N$
<i>Probabilité</i>	Fréquence $f_{kl} = \frac{n_{kl}}{n}$	Probabilité $P(x_k \leq X < x_l) = \lim_{n \rightarrow \infty} \frac{n_{kl}}{n}$ avec $n_{kl}$ nombre d'éléments de la population dont les valeurs sont entre $x_k$ et $x_l$
<i>Moyenne</i>	Moyenne arithmétique $m_x(n) = \bar{x}_n = \frac{\sum_{l=1}^n x_l}{n}$	Espérance mathématique $\mu_X = E(X) = \lim_{n \rightarrow \infty} m_x(n)$
<i>Erreur</i>	Écart algébrique $e_l = x_l - \bar{x}_n$ de moyenne $m_{e_l} = 0$	Écart algébrique $\varepsilon_l = X - \mu_X$ de moyenne $E(X - \mu_X) = 0$
<i>Écart-type</i>	Écart quadratique moyen $s_n = \sqrt{\frac{\sum_{l=1}^n (x_l - \bar{x}_n)^2}{n}}$ Variance empirique $s_n^2 = \frac{1}{n-1} \sum_{l=1}^n (x_l - \bar{x}_n)^2$ ou, encore, $s_n^2 = \frac{1}{n} \sum_{l=1}^n (x_l - \bar{x}_n)^2$	Écart-type $\sigma_X = \lim_{n \rightarrow \infty} \sqrt{\frac{\sum_{l=1}^n (x_l - \mu_X)^2}{n}}$ Variance $\sigma_X^2 = E[(X - \mu_X)^2]$ Coefficient de variation $v_X = \frac{\sigma_X}{\mu_X}$

TAB. 5.1 – Correspondance échantillon – population

# 6

## INFÉRENCE

---

6.1	Statistique exhaustive . . . . .	52
6.2	Famille complète de densités . . . . .	53
6.3	Estimateurs . . . . .	53
6.3.1	Biais d'un estimateur . . . . .	53
6.3.2	Convergence d'un estimateur . . . . .	54
6.3.3	Efficacité d'un estimateur . . . . .	54
6.3.4	Risque quadratique . . . . .	54
6.4	Amélioration d'un estimateur . . . . .	54
6.4.1	Conclusions . . . . .	55
6.5	Éléments pour un choix . . . . .	56

---

Supposons que nous voulons étudier un phénomène quelconque (e.g. physique, économique, ...) dont le résultat est caractérisé par une v.a.  $X$ . En général on associe au phénomène un modèle. Si le modèle réussit à spécifier complètement la loi de  $X$ , alors on peut se fonder sur cette loi pour prendre des décisions. Mais le plus souvent le modèle nous permet de déterminer la forme de la fonction de densité de  $X$ , sans que nous puissions préciser la valeurs des ses paramètres. Ainsi on pourra savoir que la forme de la fonction de densité

est e.g.  $f(x, \theta) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x-\theta_1)^2}{2}}$  mais sans connaître les valeurs de  $\theta = [\theta_1, \theta_2]^T \in \mathbb{R}^2$ .

Afin de pouvoir choisir la « meilleure » valeur de  $\theta$  on est amené à faire des observations sur  $X$ , c'est-à-dire à mesurer la valeur de  $X$  sur un échantillon et, à l'aide de ces valeurs mesurées de  $X$ , procéder à une estimation de la valeur de  $\theta$ .

Il y a deux méthodes d'estimation des paramètres  $\theta$  d'une loi de probabilité :

- L'estimation ponctuelle, qui, à partir des  $n$  v.a.  $X_1, \dots, X_n$  fournit une valeur pour le paramètre  $\theta$ . Il s'agit donc d'établir une application de  $\mathbb{R}^n$  dans  $\mathbb{R}^m$  si  $\theta = [\theta_1, \dots, \theta_m]^T$ . Cette application s'appelle statistique ou estimateur (cf. infra).
- L'estimation par intervalle de confiance, qui associe à chaque paramètre un intervalle, dit de confiance, dans lequel se trouve, avec une probabilité fixée, sa vraie valeur. Dans la mesure où le paramètre inconnu n'est pas une quantité aléatoire, c'est l'intervalle de confiance qui est aléatoire

Parmi les problèmes relatifs à l'estimation, notons en particulier les suivants :

- Comment choisir la « meilleure » valeur de  $\theta$  ?
- Comment déterminer la fonction  $L$  qui mesure la perte engendrée par l'estimation de  $\theta$ , i.e. quel est le coût d'une estimation (mauvaise) du paramètre  $\theta$  ?

Le cadre théorique de l'inférence est le suivant :

**DÉFINITION 6.0.1** Soit une v.a.  $X$  définie sur  $(\Omega, \mathcal{A}, P)$  qui suit une loi de probabilité de densité  $f(x, \theta)$  avec  $\theta = (\theta_1, \dots, \theta_r)^T \in \Theta$ . On définit la famille des fonctions de densité quand  $\theta$  varie dans  $\Theta$  :  $\mathcal{F}(\Theta) = \{f(\cdot, \theta) \mid \theta \in \Theta\}$ .

**DÉFINITION 6.0.2** On appelle modèle statistique la donnée d'un espace  $(\Omega, \mathcal{A})$  probabilisable et d'une famille de densités  $(P)_{\theta \in \Theta}$  quand  $\theta \in \Theta$ .

**DÉFINITION 6.0.3** Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon. On appelle statistique toute variable aléatoire  $\mathbf{U}$  de la forme  $U = \phi(X_1, \dots, X_n)$  où  $\phi$  est une fonction mesurable de  $\mathbb{R}^n \rightarrow \mathbb{R}$ .

Dans le cas multidimensionnel on écrira

$$\mathbf{U}(X_1, \dots, X_n) = (U_1(X_1, \dots, X_n), \dots, U_m(X_1, \dots, X_n))^T.$$

## 6.1 Statistique exhaustive

**DÉFINITION 6.1.1** La statistique  $\mathbf{U}(X_1, \dots, X_n)$  est une statistique exhaustive pour le paramètre  $\theta \in \Theta$ , si la probabilité conditionnelle de  $(X_1 = x_1, \dots, X_n = x_n)$  sachant  $\mathbf{U}(X_1, \dots, X_n)$ , est indépendante de  $\theta$ .

$P_\theta(X_1 = x_1, \dots, X_n = x_n \mid \mathbf{U} = \mathbf{u})$  n'est pas une fonction de  $\theta$ .

Le théorème suivant fournit un critère pour la statistique exhaustive.

**THÉORÈME 6.1.1** (Théorème de factorisation de Fisher-Neyman).- Soient  $X_1, \dots, X_n$  une suite de  $n$  v.a.i.i.d. avec fonction de densité  $f(\cdot, \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}^r$ . Une statistique  $m$ -dimensionnelle  $\mathbf{U} = \mathbf{U}(X_1, \dots, X_n)$  est exhaustive pour  $\theta \in \Theta$ , si et seulement si la densité  $f(\cdot, \theta)$  peut se mettre sous la forme

$$f(x_1, \dots, x_n; \theta) = g[\mathbf{U}(x_1, \dots, x_n); \theta] \cdot h(x_1, \dots, x_n) \quad (6.1.1)$$

où  $g$  dépend de  $x_1, \dots, x_n$  par l'intermédiaire de  $\mathbf{U}$  et  $h$  ne dépend pas de  $\theta$ .

Remarquons que la définition de la statistique exhaustive conduit au résultat suivant :

Soit  $\mathbf{U}$  une statistique exhaustive pour  $\theta \in \Theta$  et considérons une autre statistique  $\mathbf{U}' = (U'_1, \dots, U'_k)^T$ . Alors la distribution conditionnelle de  $\mathbf{U}'$  étant donné  $\mathbf{U} = \mathbf{u}$ , est indépendante de  $\theta$  pour toutes les valeurs de  $\mathbf{u}$ . En effet  $\mathbf{U}'$  est définie dans  $\mathcal{B}_{\mathbb{R}}^k$  (tribu des boréliens sur le pavé de dimension  $k$ ). Alors  $\forall B \in \mathcal{B}_{\mathbb{R}}^k$  on pose  $A = (\mathbf{U}')^{-1}(B)$ . Alors

$$P_\theta(\mathbf{u}' \in B \mid \mathbf{U} = \mathbf{u}) = P_\theta\left((x_1, \dots, x_n)^T \in A \mid \mathbf{U} = \mathbf{u}\right)$$

qui est indépendant de  $\theta$  pour toute valeur de  $\mathbf{u}$ .

Nous voyons ainsi que la connaissance d'une statistique exhaustive pour le paramètre  $\theta \in \Theta$  rend superflue la recherche d'une autre statistique, car la première contient toute l'information concernant  $\theta \in \Theta$ .

Pour compléter la notion de la statistique exhaustive nous donnons le :

**THÉORÈME 6.1.2** *Soit une application bijective  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , mesurable et indépendante de  $\theta$ . Considérons une statistique  $\mathbf{U}$ , exhaustive pour  $\theta \in \Theta$ . Alors la statistique  $\mathbf{U}' = \varphi(\mathbf{U})$  est aussi exhaustive pour  $\theta \in \Theta$  et  $\mathbf{U}$  est exhaustive pour  $\theta = \psi(\theta)$ , où  $\psi : \mathbb{R}^r \rightarrow \mathbb{R}^r$ , application bijective et mesurable.*

## 6.2 Famille complète de densités

Considérons une famille de densités dépendant du paramètre  $\theta \in \Theta : \mathcal{F}(\Theta) = \{f(\cdot, \theta) \mid \theta \in \Theta\}$  et soit  $\mathbf{X}$  un vecteur aléatoire à  $k$  dimensions, dont la densité de probabilité est  $f(\cdot, \theta); \theta \in \Theta$ . Considérons aussi l'application mesurable  $g : \mathbb{R}^k \rightarrow \mathbb{R}$ , qui est telle que  $g(\mathbf{X})$  soit une variable aléatoire. On suppose que  $E_\theta[g(\mathbf{X})]$  existe pour tout  $\theta \in \Theta$ .

**DÉFINITION 6.2.1** *On dit que la statistique  $U$  est complète pour la famille  $\mathcal{F}(\Theta)$  si pour toute fonction  $g$  définie comme ci-dessus et telle que si  $E_\theta[g(\mathbf{U})] = 0 \forall \theta \in \Theta$  implique  $g(\mathbf{U}) = 0$ .*

On montre en particulier que la statistique exhaustive des familles exponentielles est complète.

Une conséquence immédiate de cette définition est qu'il y a une seule fonction de  $\mathbf{X}$  qui a une valeur de l'espérance mathématique fixée. Car si  $g, h$  deux fonctions de  $\mathbf{X}$  avec  $E_\theta[g(\mathbf{X})] = E_\theta[h(\mathbf{X})] \forall \theta \in \Theta$ , alors  $E_\theta[g(\mathbf{X}) - h(\mathbf{X})] = 0 \forall \theta \in \Theta$ . Posons  $v(\mathbf{X}) = g(\mathbf{X}) - h(\mathbf{X})$ . Alors  $E_\theta[v(\mathbf{X})] = 0, \forall \theta \in \Theta$ . Mais comme la famille est complète, alors  $v(\mathbf{X}) = 0 \forall \mathbf{x} \in \mathbb{R}^k - N$  c'est-à-dire  $g(\mathbf{X}) - h(\mathbf{X}) = 0 \forall \mathbf{x} \in \mathbb{R}^k - N$ .

## 6.3 Estimateurs

**DÉFINITION 6.3.1** *On appelle estimateur de  $\theta$  ( ou  $\alpha(\theta)$  ), toute statistique  $U$  dont la valeur observée sur l'échantillon  $(X_1, \dots, X_n)$  est utilisée pour estimer  $\theta$  ( ou  $\alpha(\theta)$  ).*

### 6.3.1 Biais d'un estimateur

**DÉFINITION 6.3.2** *L'estimateur  $U(X_1, \dots, X_n)$  est un estimateur non biaisé pour le paramètre  $\theta$  ( ou  $\alpha(\theta)$  ),  $\theta \in \Theta$ , si  $E_\theta[U] = \theta, \forall \theta \in \Theta \subset \mathbb{R}$ .*

On appelle biais de l'estimateur  $U$  de  $\theta$  ( ou  $\alpha(\theta)$  ) le nombre  $b_\theta(U) = (E_\theta[U] - \theta)$  ( ou  $E_\theta[U] - \alpha(\theta)$  ).

DÉFINITION 6.3.3 L'estimateur  $U(X_1, \dots, X_n)$  est un estimateur asymptotiquement sans biais pour le paramètre  $\theta$  (ou  $\alpha(\theta)$ ),  $\theta \in \Theta$ , si  $\lim_{n \rightarrow \infty} b_\theta(U) = 0$ .

Notons que la notion de la statistique non biaisée n'est pas invariante par rapport à une transformation de la statistique. Ainsi si  $U$  est une statistique non biaisée, alors  $h(U)$  n'est pas en général non biaisée, sauf si  $h$  est linéaire.

### 6.3.2 Convergence d'un estimateur

DÉFINITION 6.3.4 On dit que l'estimateur  $\{U_n\}$  est convergent pour  $\theta \in \Theta$  si  $\lim_{n \rightarrow \infty} U_n = \theta$ .

### 6.3.3 Efficacité d'un estimateur

Le problème qui se pose : En quel sens peut-on dire qu'un estimateur est le meilleur possible ?

C'est le problème que tente à résoudre la notion d'efficacité.

DÉFINITION 6.3.5 Soient  $U$  et  $T$  deux estimateurs sans biais de  $\theta$ , (ou de  $\alpha(\theta)$ ), l'estimateur  $U$  est plus efficace que  $T$ , si  $V(U) \leq V(T)$ ,  $\forall \theta \in \Theta$ .

Si nous avons une suite de statistiques efficaces, alors on réduit le risque d'une mauvaise inférence en augmentant le nombre  $n$  d'observations.

Remarquons que si une statistique est efficace, alors toute fonction linéaire de cette statistique est aussi efficace.

### 6.3.4 Risque quadratique

DÉFINITION 6.3.6 On appelle risque quadratique de l'estimateur  $U$  de  $\theta$ , (ou de  $\alpha(\theta)$ ), le nombre :

$$R_\theta(U) = E_\theta(U - \theta)^2 \left( \text{ou } E_\theta(U - \alpha(\theta))^2 \right)$$

Cette quantité représente la moyenne du carré de l'erreur de la statistique  $U$  qui estime  $\theta$ .

On dit que  $U$  est préférable à  $V$  si  $\forall \theta \in \Theta, R_\theta(U) \leq R_\theta(V)$ .

PROPOSITION 6.3.1 Si  $\lim_{n \rightarrow \infty} R_\theta\{U_n\} = 0$  alors  $U_n$  est convergent et asymptotiquement sans biais.

Remarque :

- Un estimateur sans biais n'est pas toujours meilleur au sens du risque qu'un estimateur biaisé.
- Deux estimateurs ne sont pas nécessairement comparables.

## 6.4 Amélioration d'un estimateur

Il est fréquent d'avoir un estimateur  $T$  d'un paramètre  $g(\theta)$  ne possédant aucune des propriétés essentielles. Il y a deux procédures qu'on utilise pour l'amélioration d'un estimateur : la première consiste à rechercher, à partir de  $T$ , un estimateur qui lui est préférable par réduction de la variance ; la seconde a pour objet de déterminer un estimateur dont le biais est inférieur à celui de  $T$ . On va présenter seulement la première méthode.

Nous avons le théorème suivant :

**THÉORÈME 6.4.1 ( Rao-Blakwell ).-** Soient :

- $X_1, \dots, X_n$  v.a.i.i.d. avec fonction de densité  $f(\cdot, \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}$
- $U = U(X_1, \dots, X_n)$  une statistique exhaustive pour  $\theta$ .
- Considérons aussi une statistique  $V = V(X_1, \dots, X_n)$  non biaisée pour  $\theta$  et qui n'est pas une fonction de  $U$ .
- Posons  $\varphi(U) = E_\theta[V \mid U = u]$ .

Alors nous avons :

- (1)  $\varphi(U)$  est une fonction de la statistique exhaustive  $U$ .
- (2)  $\varphi(U)$  est une statistique non biaisée pour  $\theta$ .
- (3)  $\sigma_\theta^2[\varphi(U)] < \sigma_\theta^2[V]$ ,  $\theta \in \Theta$  si  $E_\theta[V^2] < \infty$ .

Ce théorème ne conduit pas à l'estimateur optimal ; il permet seulement d'obtenir un estimateur préférable à l'estimateur initial.

L'association de la statistique non biaisée et de la famille complète aboutit au résultat suivant :

**THÉORÈME 6.4.2 ( Lehmann -Scheffé ).-** Soient  $X_1, \dots, X_n$  v.a.i.i.d. avec fonction de densité  $f(\cdot, \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}$  et  $\mathcal{F}(\Theta) = \{f(\cdot, \theta) \mid \theta \in \Theta\}$  la famille de densités correspondante. Considérons la statistique exhaustive  $U = U(X_1, \dots, X_n)$  pour  $\theta \in \Theta$  avec fonction de densité  $g(\cdot, \theta)$  et soit  $\mathcal{G}(\Theta) = \{g(\cdot, \theta) \mid \theta \in \Theta\}$  la famille correspondante qu'on suppose complète. Soit, enfin,  $V = V(U)$  une statistique non biaisée pour  $\theta$ , telle que  $E_\theta[V^2] < \infty$ ,  $\forall \theta \in \Theta$ . Alors  $V$  est la seule et unique statistique non biaisée pour  $\theta$  qui a la plus petite variance.

Ce théorème prouve que, si la statistique  $U$  est complète,  $V(U)$  est optimal pour  $\theta$ , c'est-à-dire de variance minimale.

### 6.4.1 Conclusions

Le th. 6.4.2 apporte des précisions sur les résultats du th. 6.1.2. En effet on sait, d'après le th. 6.1.2, que pour avoir une statistique non biaisée de variance petite, il faut prendre des statistiques qui sont fonctions des statistiques exhaustives et le th. 6.4.2 nous permet de choisir parmi toutes ces statistiques non biaisées celle qui a une variance minimale. Elle est fonction d'une statistique exhaustive et complète.

Dans la suite nous appellerons une telle statistique non biaisée, de variance minimale, uniforme (car la variance est minimale pour tout  $\theta \in \Theta$ ).

En conclusion si on dispose d'un estimateur sans biais fonction d'une statistique exhaustive complète c'est le meilleur estimateur possible.

## 6.5 Éléments pour un choix

La meilleure statistique est celle qui a toutes les propriétés. Elle est tellement meilleure qu'elle ne peut pas exister. Il faut donc choisir une statistique en fonction des certaines propriétés, en faisant des compromis. La description des stratégies du choix ne fait pas partie de l'inférence statistique mais plutôt de la décision. Il appartient donc, à ce domaine que d'aucuns appellent l'art de l'ingénieur.

# 7

## ESTIMATION

---

7.1	Estimation avec statistique complète et exhaustive . . .	58
7.2	Estimation en absence de statistique complète et exhaustive . . .	58
7.3	Estimateur du maximum de vraisemblance . . . . .	59
7.3.1	Cas où $\theta$ est scalaire . . . . .	60

---

Dans le cas où une v.a.  $X$  a une fonction de densité  $f(\cdot, \theta)$  qui dépend d'un paramètre  $\theta \in \Theta \subset \mathbb{R}^r$ , alors il faut estimer les valeurs de  $\theta$  si on veut connaître les valeurs des probabilités associées à  $X$ . En général on ne procède pas directement à l'estimation de  $\theta$  mais on construit plutôt une fonction de  $\theta$ ,  $g(\theta)$ , réelle et mesurable. Pour estimer  $g(\theta)$  on utilise une statistique  $U = U(X_1, \dots, X_n)$  qui est l'estimateur de  $g(\theta)$ . La valeur  $U(x_1, \dots, x_n)$  de  $U$  pour les valeurs observées de  $X_1, \dots, X_n$  est appelée l'estimée de  $g(\theta)$ .

Un estimateur  $\theta$  doit être :

- non biaisé, i.e.  $E_\theta[U(X_1, \dots, X_n)] = g(\theta); \forall \theta \in \mathbb{R}^r$ ,
- de variance minimale, dans la mesure où, d'après l'inégalité de Chebychev :

$$P_\theta[|U - g(\theta)| < \varepsilon] \geq 1 - \frac{\sigma_\theta^2(U)}{\varepsilon^2}$$

la variance peut être vue comme une mesure de la concentration des valeurs de l'estimateur autour de la moyenne.

La recherche d'un estimateur non biaisé, de variance minimale (e.n-b.v.m.) s'effectue de deux façons différentes, selon que l'on dispose d'une statistique complète et exhaustive ou non. Une autre voie de recherche pour un estimateur est la méthode du maximum de vraisemblance. Notons, enfin, que nous pouvons appliquer pour l'élaboration d'une estimation d'autres techniques fondées e.g. sur la théorie de la décision, le théorème de Bayes ou l'optimisation minimax. Ces méthodes étant en dehors du cadre de ce cours, ne seront pas présentées ici.

### 7.1 Estimation avec statistique complète et exhaustive

Considérons une suite de v.a.  $X_1, \dots, X_n$  i.i.d. avec fonction de densité  $f(\cdot, \theta)$ ,  $\theta \in \Theta \subset \mathbb{R}^r$ . Supposons que  $\mathbf{V} = (V_1(X_1, \dots, X_n), \dots, V_m(X_1, \dots, X_n))^T$  est une statistique exhaustive pour  $\theta \in \Theta$  et que la famille  $\mathcal{F}(\Theta) = \{f(\cdot, \theta) \mid \theta \in \Theta\}$  est complète. Sous ces hypothèses, nous avons le :

**THÉORÈME 7.1.1** Soit  $g(\theta)$  une fonction réelle et mesurable et supposons que  $U = U(X_1, \dots, X_n)$  une statistique non biaisée pour  $g(\theta)$  avec variance bornée. Posons  $\varphi(V) = E_\theta[U \mid \mathbf{V}]$ . Alors  $\varphi(V)$  est un e.n-b.v.m. pour  $g(\theta)$  et il est unique à un ensemble de mesure nulle près.

### 7.2 Estimation en absence de statistique complète et exhaustive

Dans ce cas on fait les hypothèses suivantes :

- (1)  $f(x, \theta) > 0 \forall x \in S$  où  $S$  ensemble indépendant de  $\theta \in \Theta \subset \mathbb{R}$ .
- (2)  $\Theta$  est un intervalle ouvert de  $\mathbb{R}$ .
- (3)  $\frac{\partial}{\partial \theta} f(x, \theta)$  existe  $\forall \theta \in \Theta$  et  $\forall x \in S$ .
- (4)  $\sum_{x_1 \in S} \dots \sum_{x_n \in S} f(x_1, \theta) \dots f(x_n, \theta)$  est différentiable.
- (5) L'information de Fisher pour  $\theta$ , définie par la relation

$$I(\theta) = E_\theta \left[ \frac{\partial}{\partial \theta} \log f(x, \theta) \right]^2$$

est positive pour tout  $\theta \in \Theta$ .

- (6) Si  $U = U(X_1, \dots, X_n)$  une statistique non biaisée pour  $g(\theta)$  alors  $\sum_{x_1 \in S} \dots \sum_{x_n \in S} U(X_1, \dots, X_n) f(x_1, \theta) \dots f(x_n, \theta)$  est différentiable.

Sous ces conditions, nous avons le :

**THÉORÈME 7.2.1** ( Inégalité de Cramér – Rao ).- Pour tout estimateur non biaisé  $U = U(X_1, \dots, X_n)$  nous avons :

$$\sigma_\theta^2(U) \geq \frac{\left[ \frac{d}{d\theta} g(\theta) \right]^2}{n \cdot I(\theta)}$$

Nous avons aussi le :

**THÉORÈME 7.2.2** *Nous avons l'égalité à la relation du th.7.2.1 si et seulement si il existe une fonction réelle  $h(\theta)$  de  $\theta$  telle que  $U = g(\theta) + h(\theta)V_\theta; \forall \theta \in \Theta$  avec  $V_\theta = V_\theta(X_1, \dots, X_n) = \sum_{j=1}^n \frac{\partial}{\partial \theta} \log f(x_j, \theta)$ .*

**DÉFINITION 7.2.1** *On dit qu'un estimateur sans biais  $U$  de  $\theta$  ou ( de  $\alpha(\theta)$ ) est **efficace** si il atteint la borne de Cramer-Rao. C'est à dire si :*

$$\sigma_\theta^2(U) = \frac{\left[ \frac{d}{d\theta} g(\theta) \right]^2}{n \cdot I(\theta)}$$

### 7.3 Estimateur du maximum de vraisemblance

Cet estimateur, dû à R. A. Fisher, n'a pas une justification mathématique rigoureuse. Le principe est le suivant : Lors d'une série de  $n$  expériences qui met en jeu les v.a.  $X_1, \dots, X_n$ , ayant comme fonction de densité  $f(\cdot, \theta)$ ,  $\theta \in \Theta$ , nous avons observé les valeurs  $X_1 = x_1, \dots, X_n = x_n$ , avec probabilité  $P(X_1 = x_1, \dots, X_n = x_n; \theta)$ . Il y a certaines valeurs de  $\theta \in \Theta$  pour lesquelles cette probabilité est élevée et d'autres pour lesquelles elle est faible, c'est-à-dire pour lesquelles l'événement observé est improbable. Il paraît raisonnable, cherchant à estimer  $\theta$ , de privilégier la valeur de  $\theta$  qui maximise la probabilité  $P(X_1 = x_1, \dots, X_n = x_n; \theta)$ .

En général il y a une seule valeur de  $\theta$  qui maximise  $P(X_1 = x_1, \dots, X_n = x_n; \theta)$ . Pour cette valeur on utilise la notation  $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$  et l'on appelle *l'estimée du maximum de vraisemblance*. L'estimateur  $\hat{\theta}(X_1, \dots, X_n)$  est appelé *l'estimateur du maximum de vraisemblance (e.m.v.)* pour  $\theta$ .

Parfois pour établir la valeur maximale de la probabilité on utilise la fonction

$$L(\theta | x_1, \dots, x_n) = h(x_1, \dots, x_n) \cdot P(X_1 = x_1, \dots, X_n = x_n; \theta)$$

où  $h(x_1, \dots, x_n) > 0; \forall (x_1, \dots, x_n) \in \mathbb{R}^n$  est une fonction qui ne dépend pas de  $\theta$ .  $L(\theta | x_1, \dots, x_n)$  est la fonction du maximum de vraisemblance de  $\theta$  et nous allons estimer  $\hat{\theta}$  à l'aide de cette fonction en posant éventuellement  $h(x_1, \dots, x_n) = 1$ , car la valeur de  $\theta$  qui maximise  $P(X_1 = x_1, \dots, X_n = x_n; \theta)$  maximise aussi  $L(\theta | x_1, \dots, x_n)$  et maximise encore la fonction

$$l(\theta | x_1, \dots, x_n) = \ln L(\theta | x_1, \dots, x_n)$$

qui est le logarithme du maximum de vraisemblance. Cette forme de la fonction du maximum de vraisemblance est utile quand la densité de probabilité a une forme exponentielle, comme e.g. la loi normale.

Le théorème suivant relie l'estimateur du maximum de vraisemblance avec un estimateur exhaustif.

**THÉORÈME 7.3.1** *Soit une suite  $X_1, \dots, X_n$  de v.a.i.i.d. avec fonction de densité de probabilité  $f(\cdot, \theta); \theta \in \Theta \subset \mathbb{R}^r$ . Considérons une statistique  $\mathbf{V} = (V_1(x_1, \dots, x_n), \dots, V_r(x_1, \dots, x_n))^\top$*

exhaustive pour  $\theta = (\theta_1, \dots, \theta_r)^\top$ . Si  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_r)^\top$  est l'unique estimateur du maximum de vraisemblance de  $\theta$ , alors  $\hat{\theta}$  est une fonction de  $\mathbf{V}$ .

Le théorème suivant montre que l'estimateur du maximum de vraisemblance est invariant par rapport à une transformation bijective.

**THÉORÈME 7.3.2** Soit une suite  $X_1, \dots, X_n$  de v.a.i.i.d. avec fonction de densité de probabilité  $f(\cdot, \theta)$ ;  $\theta \in \Theta \subset \mathbb{R}^r$ . Considérons la transformation bijective  $\varphi: \Theta \rightarrow \Theta' \subset \mathbb{R}^m$ . Si  $\hat{\theta}$  est l'estimateur du maximum de vraisemblance de  $\theta$ , alors  $\varphi(\hat{\theta})$  est l'estimateur du maximum de vraisemblance de  $\varphi(\theta)$ .

### 7.3.1 Cas où $\theta$ est scalaire

Le calcul de la valeur  $\hat{\theta}$  de  $\theta = (\theta_1, \dots, \theta_r)^\top$  qui maximise la fonction du maximum de vraisemblance est un problème de maximisation d'une fonction, sous la contrainte  $a_i \leq \theta_i \leq b_i$ ;  $\forall i = 1, \dots, r$ , et, de ce fait, complexe. Dans le cas particulier où  $\theta$  est un scalaire, nous pouvons utiliser le logarithme du maximum de vraisemblance. Si les dérivées

$$l'(\theta | x_1, \dots, x_n) = \frac{\partial}{\partial \theta} l(\theta | x_1, \dots, x_n)$$

$$l''(\theta | x_1, \dots, x_n) = \frac{\partial^2}{\partial \theta^2} l(\theta | x_1, \dots, x_n)$$

existent pour tout  $\theta \in \Theta$ , alors l'estimateur du maximum de vraisemblance est une des racines de l'équation

$$l'(\theta | x_1, \dots, x_n) = 0$$

Une racine  $\xi$  de cette équation pour laquelle nous avons

$$l''(\xi | x_1, \dots, x_n) < 0$$

est un maximum local. Par conséquent nous avons :

$$\hat{\theta} = \arg \max \{ l(\xi | x_1, \dots, x_n) / l'(\xi | x_1, \dots, x_n) = 0, l''(\xi | x_1, \dots, x_n) < 0 \}$$

## INTERVALLE DE CONFIANCE

---

8.1	Définition d'un intervalle de confiance . . . . .	62
8.2	Construction des intervalles de confiance . . . . .	62
8.2.1	Intervalles de confiance pour les moyennes . . . . .	63
8.2.1.1	Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec variance connue . . . . .	64
8.2.1.2	Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec variance inconnue . . . . .	64
8.2.2	Intervalle de confiance pour les variances . . . . .	65
8.2.2.1	Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec moyenne connue . . . . .	65
8.2.2.2	Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec moyenne inconnue . . . . .	66

---

L'estimateur du maximum de vraisemblance fournit la valeur la plus probable du paramètre sous test, en se fondant sur un ensemble d'observations  $X_1 = x_1, \dots, X_n = x_n$  effectuées en utilisant un échantillon de taille  $n$ . Néanmoins on ne peut pas affirmer que cette valeur coïncide toujours avec la vraie valeur. Plutôt que de chercher à savoir si la valeur estimée est proche ou éloignée de la vraie valeur (ce qui aurait exigé d'avoir une métrique ad hoc dans l'espace des paramètres), on cherche à déterminer un intervalle dans lequel il est probable de supposer que se trouve la vraie valeur avec une probabilité de se tromper fixée par l'utilisateur. Cet intervalle s'appelle *intervalle de confiance*.

Pour fixer les idées on prend comme exemple un ensemble de  $n$  observations  $X_1 = x_1, \dots, X_n = x_n$  qui suivent la loi normale  $\mathcal{N}(\mu, \sigma^2)$  où  $\mu$  et  $\sigma^2$  sont inconnues. Supposons qu'on cherche à trouver un intervalle qui contient  $\mu$  avec une probabilité de  $1 - \alpha$ , où  $\alpha$  représente la probabilité de se tromper et qui est fixée par l'utilisateur. Soit  $\theta = \mu$  et posons

$$u(x, \mu) = \sqrt{n} \frac{\bar{x} - \mu}{s} \quad \text{avec} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

La distribution de  $u(x, \mu)$  est la distribution de Student à  $n - 1$  degrés de liberté. Ainsi, sans connaître  $\theta$ , on peut trouver une valeur  $u_\alpha$  telle que :

$$P_\theta [-u_\alpha \leq u(x, \mu) \leq u_\alpha] = 1 - \alpha \quad \text{avec} \quad \alpha \in [0, 1]$$

d'où

$$P_{\theta} \left[ \bar{x} - u_{\alpha} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + u_{\alpha} \frac{s}{\sqrt{n}} \right] = 1 - \alpha$$

et, finalement

$$P_{\theta} \left\{ \mu \in \left[ \bar{x} - u_{\alpha} \frac{s}{\sqrt{n}}, \bar{x} + u_{\alpha} \frac{s}{\sqrt{n}} \right] \right\} = 1 - \alpha$$

L'intervalle  $\left[ \bar{x} - u_{\alpha} \frac{s}{\sqrt{n}}, \bar{x} + u_{\alpha} \frac{s}{\sqrt{n}} \right]$  est l'intervalle de confiance pour la valeur de la moyenne  $\mu$  à  $100 \times (1 - \alpha)\%$ .

### 8.1 Définition d'un intervalle de confiance

Soit l'espace probabilisé  $(\Omega, \mathcal{A}, p)$  et soit la variable aléatoire  $X : \Omega \times \Theta \rightarrow \mathcal{B}_{\mathbb{R}}$ , où  $\Theta$  ensemble de variation du paramètre  $\theta$ . Notons par  $F(\Theta) = \{f(\cdot, \theta) \mid \theta \in \Theta\}$  la famille de densités des probabilités définies avec  $\theta \in \Theta$ .

Considérons une famille  $I(\Theta) = \{I_x(\theta) \mid x \in \mathcal{B}_{\mathbb{R}}\}$  d'intervalles de  $\Theta$  qui ont la propriété suivante :

$$\forall \theta \in \Theta : P_{\theta}[x \mid \theta \in I_x(\theta)] = 1 - \alpha ; \alpha \in [0, 1]$$

où  $\alpha$  est la probabilité que l'intervalle ne contient pas la vraie valeur du paramètre  $\theta$ .

Notons que nous pouvons avoir

$$P_{\theta}[x \mid \theta \in I_x(\theta)] \geq 1 - \alpha ; \alpha \in [0, 1]$$

i.e. un intervalle de confiance avec coefficient minimal de confiance  $1 - \alpha$ .

La détermination des bornes de l'intervalle de confiance dépend du partage de  $\alpha$  en  $\alpha_1$  et  $\alpha_2$ . Nous pouvons avoir deux cas de figure :

- Recherche d'un intervalle bilatéral  $[a, b]$ , correspondant à  $\alpha_1 \neq 0, \alpha_2 \neq 0$ . Dans le cas où la loi est symétrique (e.g. loi normale) nous avons  $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$  et  $b = -a$ .
- Recherche d'un intervalle unilatéral. Deux cas :
  - $[a, +\infty[$  associé à  $\alpha_1 = \alpha$  et  $\alpha_2 = 0$ . Cet intervalle est à utiliser dans le cas de recherche d'un intervalle du paramètre  $\theta$  de la forme  $\theta > a$  (durée de vie, résistance à la rupture, etc.).
  - $]-\infty, b]$  associé à  $\alpha_1 = 0$  et  $\alpha_2 = \alpha$ . Cet intervalle est à utiliser dans le cas de recherche d'un intervalle du paramètre  $\theta$  de la forme  $\theta < b$  (nombre de pièces défectueuses, temps d'attente, etc.).

### 8.2 Construction des intervalles de confiance

Afin de construire un intervalle de confiance, on détermine un estimateur  $u(x, \theta)$  qui pour chaque  $x \in \mathcal{B}_{\mathbb{R}}$  est une fonction monotone de  $\theta \in \Theta$ . On détermine ensuite deux valeurs  $u_1, u_2$  telles que :

$$P_{\theta}[u_1 < u(x, \theta) < u_2] = 1 - \alpha ; \alpha \in [0, 1]$$

et, comme  $u$  est monotone, nous avons :

$$P_{\theta} [\theta_1(x) < \theta < \theta_2(x)] = 1 - \alpha ; \alpha \in [0, 1]$$

$[\theta_1(x), \theta_2(x)]$  est l'intervalle de confiance.

Nous présentons dans la suite quelques exemples des fonctions monotones qui peuvent être utilisées comme des estimateurs pour le calcul des intervalles de confiance.

– Supposons que  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ .

(1) La variance est connue. On pose  $\theta = \mu$  et on construit la fonction

$$t(X_1, \dots, X_n; \theta) = \sqrt{n} \frac{\bar{X} - \theta}{\sigma} \sim \mathcal{N}(0, 1)$$

qui est une fonction monotone par rapport à  $\theta$ .

(2) La variance est inconnue. On utilise l'estimation  $s^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$ . On

pose  $\theta = \mu$  et on construit la fonction

$$t(X_1, \dots, X_n; \theta) = \sqrt{n} \frac{\bar{X} - \theta}{s} \sim t(n-1)$$

qui est une fonction monotone par rapport à  $\theta$ .

(3) La moyenne est connue. On pose  $\theta = \sigma^2$  et on construit la fonction

$$t(X_1, \dots, X_n; \theta) = \frac{1}{\theta} \sum_{j=1}^n (X_j - \mu)^2 \sim \chi_n^2$$

qui est une fonction monotone par rapport à  $\theta$ .

(4) La moyenne est inconnue. On pose  $\theta = \sigma^2$  et on construit la fonction

$$t(X_1, \dots, X_n; \theta) = \frac{1}{\theta} (n-1) \cdot s^2 \sim \chi_{n-1}^2$$

qui est une fonction monotone par rapport à  $\theta$ . Nous avons noté  $s^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$ .

– Supposons que  $X \sim \mathcal{B}(n, p)$ . On pose  $\theta = p$ . Dans ce cas on construit une fonction monotone en considérant la limite

$$t(X; \theta) = \lim_{n \rightarrow \infty} \frac{X - n\theta}{\sqrt{n\theta(1-\theta)}} \sim \mathcal{N}(0, 1)$$

– Supposons que  $X \sim \mathcal{P}(\lambda)$ . On pose  $\theta = \lambda$ . Dans ce cas on construit une fonction monotone en considérant la limite

$$\lim_{n \rightarrow \infty} t(X_1, \dots, X_n; \theta) = \lim_{n \rightarrow \infty} \sqrt{n} \frac{\bar{X} - \theta}{\sqrt{\theta}} \sim \mathcal{N}(0, 1)$$

### 8.2.1 Intervalles de confiance pour les moyennes

Nous donnons ci-après quelques exemples de construction d'intervalle de confiance pour la moyenne d'une suite des variables aléatoires qui suivent la loi normale. La démarche reste la même pour les autres lois.

#### 8.2.1.1 Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec variance connue

On construit la fonction  $Z_n(\mu) = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$  où  $\bar{X}$  est une statistique exhaustive pour la moyenne  $\mu$ . On cherche deux nombres  $a$  et  $b$ , avec  $a < b$  et tels que :

$$P \left[ a \leq \mathcal{N}(\mu, \sigma^2) \leq b \right] = 1 - \alpha \quad \Rightarrow$$

$$P_\mu \left[ a \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq b \right] = 1 - \alpha \quad \Rightarrow$$

$$P_\mu \left[ \bar{X} - b \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} - a \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha \quad \Rightarrow$$

$$P_\mu \left\{ \mu \in \left[ \bar{X} - b \frac{\sigma}{\sqrt{n}}, \bar{X} - a \frac{\sigma}{\sqrt{n}} \right] \right\} = 1 - \alpha$$

La longueur de l'intervalle est  $\frac{b-a}{\sqrt{n}}\sigma$ . Pour que cette longueur soit minimale il faut que la différence  $b-a$  soit minimale. Ceci est possible si  $b=c$  et  $a=-c$ , avec  $c$  le  $\alpha/2$  quantile de  $\mathcal{N}(0, 1)$  qui sera par la suite noté  $z_{\alpha/2}$ . Nous avons donc

$$P_\mu \left\{ \mu \in \left[ \bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \right\} = 1 - \alpha$$

La taille minimale de l'échantillon afin d'avoir un intervalle de longueur  $2c$  est donnée par la relation

$$2z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 2c$$

d'où on obtient

$$n = \left[ \frac{z_{\alpha/2} \cdot \sigma}{c} \right]^2$$

#### 8.2.1.2 Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec variance inconnue

On remplace la variance  $\sigma^2$  par l'estimateur  $s^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$ . On construit la fonction  $T_n(\mu) = \sqrt{n} \frac{\bar{X} - \mu}{s}$ . Cette v.a. est distribuée selon la loi de Student à  $n-1$  degrés de liberté. Par conséquent nous avons :

$$P \left[ a \leq t_{n-1} \leq b \right] = 1 - \alpha \quad \Rightarrow$$

$$P_{\mu} \left[ a \leq \sqrt{n} \frac{\bar{X} - \mu}{s} \leq b \right] = 1 - \alpha \quad \Rightarrow$$

$$P_{\mu} \left[ \bar{X} - b \frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} - a \frac{s}{\sqrt{n}} \right] = 1 - \alpha$$

La longueur de l'intervalle est  $\frac{b-a}{\sqrt{n}}s$ . Cette longueur est minimale si  $b = c$  et  $a = -c$ , avec  $c = t_{n-1; \alpha/2}$ . Nous avons donc :

$$P_{\mu} \left\{ \mu \in \left[ \bar{X} - t_{n-1; \alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1; \alpha/2} \frac{s}{\sqrt{n}} \right] \right\} = 1 - \alpha$$

Si on veut que l'intervalle ait une longueur égale à  $2c$ , alors il faut prendre

$$n = \left[ \frac{t_{n-1; \alpha/2} \cdot s}{c} \right]^2$$

### 8.2.2 Intervalle de confiance pour les variances

Nous donnons ci-après quelques exemples de construction d'intervalle de confiance pour la variance d'une suite des variables aléatoires qui suivent la loi normale. La démarche reste la même pour les autres lois.

#### 8.2.2.1 Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec moyenne connue

On pose  $\theta = \sigma^2$  et on construit la fonction

$$U(\theta) = U(\sigma^2) = \frac{ns_{\mu}^2}{\sigma^2}$$

avec  $s_{\mu}^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \mu)^2$  estimateur exhaustif pour  $\theta = \sigma^2$ . Nous avons que  $U(\sigma^2) \sim \chi_n^2$ .

On cherche donc à calculer deux nombres  $a$  et  $b$  tels que :

$$P[a \leq \chi_n^2 \leq b] = 1 - \alpha \quad \Rightarrow$$

$$P_{\sigma^2} \left[ a \leq \frac{ns_{\mu}^2}{\sigma^2} \leq b \right] = 1 - \alpha \quad \Rightarrow$$

$$P_{\sigma^2} \left[ \frac{ns_{\mu}^2}{b} \leq \sigma^2 \leq \frac{ns_{\mu}^2}{a} \right] = 1 - \alpha \quad \Rightarrow$$

$$P_{\sigma^2} \left\{ \sigma^2 \in \left[ \frac{ns_{\mu}^2}{b}, \frac{ns_{\mu}^2}{a} \right] \right\} = 1 - \alpha$$

La longueur de l'intervalle est  $\left(\frac{1}{a} - \frac{1}{b}\right) ns_{\mu}^2$ . Bien qu'il ne s'agit pas d'un choix optimal, on prend

$$a = \chi_{n;1-\alpha/2}^2 \quad ; \quad b = \chi_{n;\alpha/2}^2$$

### 8.2.2.2 Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec moyenne inconnue

La démarche est la même que dans le cas où la moyenne est connue. Pour  $\mu$  on utilise l'estimateur  $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$ . Ainsi la fonction pour  $\theta = \sigma^2$  devient  $S_n^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$ .

Dans ce cas nous avons  $U(\sigma^2) \sim \chi_{n-1}^2$ . La longueur de l'intervalle est maintenant  $\left(\frac{1}{a} - \frac{1}{b}\right) (n-1) s_{\bar{X}}^2$  et on prend pour les bornes de l'intervalle

$$a = \chi_{n-1;1-\alpha/2}^2 \quad ; \quad b = \chi_{n-1;\alpha/2}^2$$

## TESTS D'HYPOTHÈSES

---

9.1	Choix d'un test . . . . .	69
9.2	Test de la moyenne avec une valeur de référence . . . . .	71
9.2.1	Variance $\sigma^2$ de la population connue . . . . .	72
9.2.1.1	Test simple . . . . .	72
9.2.1.2	Test double . . . . .	72
9.2.2	Variance $\sigma^2$ inconnue . . . . .	73
9.2.2.1	Test simple . . . . .	73
9.2.2.2	Test double . . . . .	74
9.3	Test de la variance avec une valeur de référence . . . . .	75
9.3.0.3	Test simple . . . . .	75
9.3.0.4	Test double . . . . .	76
9.4	Comparaison entre deux moyennes . . . . .	76
9.4.1	Variances $\sigma_x^2, \sigma_y^2$ des populations connues . . . . .	77
9.4.1.1	Test simple . . . . .	77
9.4.1.2	Test double . . . . .	77
9.4.2	Variances des populations inconnues mais égales . . . . .	78
9.4.2.1	Test simple . . . . .	78
9.4.2.2	Test double . . . . .	79
9.4.3	Variances des populations inconnues et non égales . . . . .	80
9.4.3.1	Test simple . . . . .	80
9.4.3.2	Test double . . . . .	80
9.5	Comparaison entre deux variances . . . . .	81
9.5.0.3	Test simple . . . . .	81
9.5.0.4	Test double . . . . .	82

---

On considère un échantillon de taille  $n$  issu d'une population d'effectif beaucoup plus grand. On mesure sur cet échantillon une caractéristique et notons par  $x_i, i = 1; \dots, n$  les mesures ainsi obtenues. On considère que ces mesures sont issues d'une distribution particulière de l'ensemble de variables aléatoires  $X_i, i = 1; \dots, n$  qui sont considérées comme indépendantes et identiquement distribuées avec fonction de densité de probabilité  $f(\cdot, \theta); \theta \in \Theta$ . On suppose que  $\Theta = \Theta_0 \cup \Theta_1$  avec  $\Theta_1 = \Theta - \Theta_0$ . Selon les valeurs des mesures, nous pouvons formuler les hypothèses suivantes :

- $H_0 : \theta \in \Theta_0$   
qui est l'hypothèse nulle. L'hypothèse contraire
- $H_1 : \theta \in \Theta_1$   
est appelée hypothèse alternative.

Afin d'examiner la validité de ces hypothèses on élabore des tests d'hypothèses. Un

Réalité	$H_0$ vraie	$H_1$ vraie
Décision		
<b>Accepter <math>H_0</math></b>	Acceptation à raison de l'hypothèse nulle Prob. $1 - \alpha_\phi(\theta)$	Acceptation à tort de l'hypothèse nulle Prob. $\beta_\phi(\theta)$
<b>Réfuser <math>H_0</math></b>	Réfus à tort de l'hypothèse nulle Prob. $\alpha_\phi(\theta)$	Réfus à raison de l'hypothèse nulle Prob. $1 - \beta_\phi(\theta)$

TAB. 9.2 – Décisions d'un test d'hypothèses

tel test est une application mesurable

$$\phi : \mathbb{R}^n \rightarrow \begin{cases} [0, 1] \\ \{0, 1\} \end{cases}$$

Dans la suite on notera par  $\Phi$  l'ensemble de ces applications. Si  $[x_1, \dots, x_n]$  est la valeur observée de  $(X_1, \dots, X_n)$ , alors  $\phi(x_1, \dots, x_n) = \xi$ . Il y a deux possibilités :

- soit  $\xi \in [0, 1]$ . Dans ce cas on procède à un tirage d'une valeur qui suit la loi uniforme sur  $[0, 1]$ . Si la valeur tirée est inférieure ou égale à  $\xi$ , on accepte le test ;
- soit  $\xi \in \{0, 1\}$ . Dans ce cas on accepte le test si  $\xi = 0$ .

Lors de la prise d'une décision, à l'aide du test  $\phi$ , concernant  $H_0$  (resp.  $H_1$ ) il est possible de commettre une erreur parmi les deux qui sont possibles :

- Erreur de 1<sup>e</sup> espèce.- Rejet de l'hypothèse  $H_0$  bien qu'elle soit vraie. (Rejet à tort de l'hypothèse  $H_0$  ou, ce qui revient au même, acceptation à tort de l'hypothèse  $H_1$ ).
- Erreur de 2<sup>e</sup> espèce.- Acceptation de l'hypothèse  $H_0$  bien qu'elle soit fautive. (Rejet à tort de l'hypothèse  $H_1$  ou, ce qui revient au même, acceptation à tort de l'hypothèse  $H_0$ ).

On suppose que l'erreur de 1<sup>e</sup> espèce se réalise avec une probabilité  $\alpha_\phi(\theta)$  que l'on appelle risque de 1<sup>e</sup> espèce. Par conséquent la probabilité d'accepter à raison  $H_0$  est égale à  $1 - \alpha_\phi(\theta)$ . De même on suppose que l'erreur de 2<sup>e</sup> espèce se réalise avec une probabilité  $\beta_\phi(\theta)$  qui est le risque de 2<sup>e</sup> espèce. Par conséquent la probabilité de refuser à raison  $H_0$  est égale à  $1 - \beta_\phi(\theta)$ . Cette situation est repertoriée au tableau suivant 6.1.

En général les deux erreurs ne conduisent pas à des conséquences de même nature. De ce fait nous ne pouvons pas traiter de façon symétrique les deux hypothèses  $H_0$  et  $H_1$ . Lors de l'élaboration d'un test on prendra comme hypothèse nulle celle dont le rejet à tort a les conséquences les plus pénalisantes.

Dans la suite de ce chapitre nous présentons d'une part une méthode qui permet d'effectuer le choix d'un test et, d'autre part, des algorithmes des tests et quelques applications.

## 9.1 Choix d'un test

Nous avons vu que les probabilités  $\alpha_\varphi(\theta)$  et  $\beta_\varphi(\theta)$  dépendent du test  $\varphi$  et de la valeur de  $\theta$ . Pour choisir le test  $\varphi$  nous allons examiner les valeurs de ces probabilités.

La fonction  $\alpha_\varphi(\theta)$  pour  $\theta \in \Theta_0$  est appelée niveau (de signification) du test et la fonction  $1 - \beta_\varphi(\theta)$  pour  $\theta \in \Theta_1$  est appelée puissance du test.

Considérons un test (application mesurable)  $\varphi \in \Phi$ . La fonction de niveau de signification du test est donnée par

$$\alpha_\varphi(\theta) = E_\theta[\varphi(X_1, \dots, X_n)] ; \theta \in \Theta_0 \quad (9.1.1)$$

Les objectifs qui permettent de choisir un test sont les suivants :

- minimiser le risque de 1e espèce (probabilité du rejet à tort de l'hypothèse nulle)  $\alpha_\varphi(\theta)$ ,  $\theta \in \Theta_0$  ;
- minimiser le risque de 2e espèce (probabilité de l'acceptation à tort de l'hypothèse nulle)  $\beta_\varphi(\theta)$ ,  $\theta \in \Theta_1$ .

On doit donc chercher à choisir un test  $\varphi \in \Phi$  qui minimise ces deux risques. Ce problème n'a pas, en général, de solution car il n'est pas obligatoire qu'une fonction  $\varphi$  assure ces deux valeurs minimales en même temps. On traite donc ces deux risques de manière non symétrique. On choisit d'abord un niveau de signification  $\alpha$  constant pour le test et on cherche un test  $\varphi$  qui, pour  $\alpha = \text{constante}$ , maximise la puissance  $1 - \beta_\varphi(\theta)$  ;  $\theta \in \Theta_1$  du test. Un tel test est appelé uniformément le plus puissant (test UPP). Formellement si  $\Phi_\alpha = \left\{ \varphi \in \Phi \mid \sup_{\theta \in \Theta} \alpha_\varphi(\theta) = \alpha \right\}$ , alors un test  $\varphi^* \in \Phi_\alpha$  est un test UPP si  $1 - \beta_{\varphi^*}(\theta) = \max_{\varphi \in \Phi_\alpha} \{1 - \beta_\varphi(\theta)\}$ . Si  $\text{card}(\Theta_1) = 1$ , alors le test UPP s'appelle le test le plus puissant (test PP).

Nous pouvons avoir trois types de tests :

- (i)  $H_0$  et  $H_1$  simples, c'est-à-dire  $\text{card}(\Theta_0) = 1$  et  $\text{card}(\Theta_1) = 1$  ;
- (ii)  $H_0$  composite et  $H_1$  simple, c'est-à-dire  $\text{card}(\Theta_0) > 1$  et  $\text{card}(\Theta_1) = 1$  ;
- (iii)  $H_0$  et  $H_1$  composites, c'est-à-dire  $\text{card}(\Theta_0) > 1$  et  $\text{card}(\Theta_1) > 1$ .

Dans la suite nous présentons le premier type de test, les deux autres étant des simples extensions de celui-ci.

Soient les variables aléatoires  $X_1, \dots, X_n$ , indépendantes et identiquement distribuées avec fonction de densité de probabilité  $f(\cdot, \theta)$  ;  $\theta \in \Theta$ . Supposons que  $\Theta = \{\theta_0, \theta_1\}$ . Les hypothèses que nous voulons tester sont :

$H_0 : \theta = \theta_0$ , hypothèse nulle ;

$H_1 : \theta = \theta_1$ , hypothèse alternative.

Nous voulons donc tester  $H_0$  contre  $H_1$  en construisant un test au niveau  $\alpha$ . Le théorème suivant nous fournit un test UPP.

**THÉORÈME 9.1.1** (Théorème de Neyman-Pearson) *Pour tester  $H_0$  contre  $H_1$  au niveau  $\alpha$ , on construit un test  $\varphi \in \Phi_\alpha$  défini par la relation*

$$\varphi(x_1, \dots, x_n) = \begin{cases} 1, & \text{si } \prod_{k=1}^n f(x_k, \theta_1) > C \cdot \prod_{k=1}^n f(x_k, \theta_0) \\ \gamma, & \text{si } \prod_{k=1}^n f(x_k, \theta_1) = C \cdot \prod_{k=1}^n f(x_k, \theta_0) \\ 0, & \text{si } \prod_{k=1}^n f(x_k, \theta_1) < C \cdot \prod_{k=1}^n f(x_k, \theta_0) \end{cases} \quad (9.1.2)$$

où les constantes  $\gamma \in [0, 1]$  et  $C > 0$  sont telles que

$$E_{\theta_0}[\varphi(X_1, \dots, X_n)] = \alpha \quad (9.1.3)$$

Alors ce test est le test le plus puissant.

**COROLLAIRE 9.1.1** *Soit un test  $\varphi \in \Phi_\alpha$  défini par (9.1.2) et (9.1.3). Alors nous avons*

$$1 - \beta_\varphi(\theta) \geq \alpha \quad (9.1.4)$$

Pour l'application de ces résultats on considère les v.a.  $X_1, \dots, X_n$ , i.i.d. avec  $X_k \sim \mathcal{L}(\theta)$ , où  $\mathcal{L}$  loi à préciser.

L'algorithme pour la création du test  $\varphi \in \Phi_\alpha$  selon le th. de Neyman-Pearson est le suivant :

**1°.-** On fixe les valeurs du niveau de signification du test  $\alpha$  et du nombre de v.a.  $n$ .

**2°.-** Détermination de la région critique.

**1.1.-** Calcul du quotient

$$R(f(X), \theta_0, \theta_1) = \frac{\prod_{k=1}^n f(x_k, \theta_1)}{\prod_{k=1}^n f(x_k, \theta_0)} \quad (9.1.5)$$

Il faut que nous ayons

$$R(f(X), \theta_0, \theta_1) > C \quad (9.1.6)$$

afin que  $\varphi(x_1, \dots, x_n) = 1$ .

**1.2.-** Calcul du logarithme

$$r(f(X), \theta_0, \theta_1) = \ln R(f(X), \theta_0, \theta_1) \quad (9.1.7)$$

Il faut que

$$r(f(X), \theta_0, \theta_1) > \ln C \quad (9.1.8)$$

pour que  $\varphi(x_1, \dots, x_n) = 1$ .

**1.3.- Résolution de l'inégalité (9.1.8)** par rapport à  $f(X)$ . Supposons que la solution s'exprime sous la forme

$$f(X) > C_0(C, \theta_0, \theta_1) \quad (9.1.9)$$

La région critique est donc donnée par la relation

$$W = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) > C_0(C, \theta_0, \theta_1)\}$$

**3°.-** Calcul de la valeur de  $C_0 = C_0(C, \theta_0, \theta_1)$ . Nous avons d'après (9.1.3)

$$E_{\theta_0}[\varphi(X_1, \dots, X_n)] \alpha = P_{\theta_0}(W) = P_{\theta_0}(f(\mathbf{x}) > C_0) \quad (9.1.10)$$

ce qui donne

$$\alpha = P_{\theta_0}(f(\mathbf{x}) > C_0) + \gamma P_{\theta_0}(f(\mathbf{x}) = C_0) \quad (9.1.11)$$

Comme le facteur de  $\gamma$  est petit, cette relation permet, en raisonnant uniquement sur le premier terme, de calculer la valeur de  $C_0$ .

**4°.-** Calcul de la constante  $\gamma \in [0, 1]$ . En utilisant la relation (9.1.11) et la valeur de  $C_0$ , on évalue la valeur de  $\gamma$ .

**5°.-** Calcul de la puissance  $1 - \beta_\varphi(\theta)$  du test à l'aide de la relation

$$1 - \beta_\varphi(\theta) = P_{\theta_1}(f(\mathbf{x}) > C_0) + \gamma P_{\theta_1}(f(\mathbf{x}) = C_0) \quad (9.1.12)$$

N.B. Cet algorithme est particulièrement adapté dans le cas où la fonction de densité de probabilité  $f(x, \theta)$  de la loi  $\mathcal{L}(\theta)$  se présente sous la forme d'un produit des termes en  $\theta$  comme, e.g. la densité de la loi binomiale  $f(x, \theta) = C_n^x \theta^x (1 - \theta)^{n-x}$ .

## 9.2 Test de la moyenne avec une valeur de référence

Supposons que d'une population, de taille potentiellement infinie, on extrait un échantillon de taille  $n$ . Les résultats des mesures effectuées sur cet échantillon sont représentés par les variables aléatoires  $X_1, \dots, X_n$ . Pour ces variables aléatoires nous faisons l'hypothèse qu'elles sont indépendantes et identiquement distribuées avec moyenne arithmétique  $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ . Le paramètre que nous voulons estimer est la moyenne  $\theta = \mu$  de la

population totale et nous utiliserons comme valeur de référence pour le test l'estimation  $\bar{X}$ .

On prend comme hypothèse nulle

$$H_0 : \theta = \theta_0$$

et comme hypothèse alternative

$$H_1 : \theta \neq \theta_0$$

Parfois on cherche à détecter si  $\theta = \theta_1$ , avec  $\theta_1 \neq \theta_0$ , qui est l'hypothèse de détection :

$$H_D : \theta = \theta_1$$

Notons, enfin, que selon les valeurs que  $\theta$  puisse prendre, nous avons un test simple ou un test double.

### 9.2.1 Variance $\sigma^2$ de la population connue

#### 9.2.1.1 Test simple

1°.-  $H_0 : \theta = \theta_0 ; H_1 : \theta < \theta_0$  (ou  $\theta > \theta_0$ ) ;  $H_D : \theta = \theta_1$

2°.- On fixe les probabilités  $\alpha$  et  $\beta$ .

3°.- Calcul de la taille de l'échantillon :

3.1.- Calcul de la valeur  $Z_\alpha$ . Nous avons

$$P\left(\frac{\theta_0 - \bar{X}}{\sigma/\sqrt{n}} > Z_\alpha\right) = \alpha ; Z_\alpha \sim \mathcal{N}(0, 1)$$

Par lecture des tables de la distribution normale on établit la valeur de  $Z_\alpha$ .

3.2.- Calcul de la valeur  $Z_\beta$ . Nous avons

$$P\left(\frac{\bar{X} - \theta_1}{\sigma/\sqrt{n}} > Z_\beta\right) = \beta ; Z_\beta \sim \mathcal{N}(0, 1)$$

Par lecture des tables de la distribution normale on établit la valeur de  $Z_\beta$ .

3.3.- Calcul de la taille minimale  $n_C$  de l'échantillon :

$$n_C = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\theta_0 - \theta_1)_D^2}$$

Si  $n_C > n$ , où  $n$  la taille de l'échantillon, il faut recommencer l'échantillonnage avec une valeur de  $n$  plus grande.

4°.- Calcul de la valeur du critère  $C$  :

$$C = \pm \frac{1}{2} (Z_\beta - Z_\alpha) \sqrt{\frac{\sigma^2}{n}} + \frac{(\theta_0 + \theta_1)_D}{2}$$

On prend le signe "+" si  $H_1 : \theta < \theta_0$ , on prend le signe "-" si  $H_1 : \theta > \theta_0$

5°.- Décision :

Si  $H_1 : \theta > \theta_0$ , alors l'hypothèse  $H_0$  est acceptée si  $\bar{X} \leq C$ .

Si  $H_1 : \theta < \theta_0$ , alors l'hypothèse  $H_0$  est acceptée si  $\bar{X} \geq C$ .

#### 9.2.1.2 Test double

1°.-  $H_0 : \theta = \theta_0 ; H_1 : \theta \neq \theta_0 ; H_D : |\theta_1 - \theta_0| = d$

2°.- On fixe les probabilités  $\alpha$  et  $\beta$ .

3°.- Calcul de la taille de l'échantillon :

3.1.- Calcul de la valeur  $Z_{\alpha/2}$ . Nous avons

$$P\left(\frac{\theta_0 - \bar{X}}{\sigma/\sqrt{n}} > Z_{\alpha/2}\right) = \frac{\alpha}{2} ; Z_{\alpha/2} \sim \mathcal{N}(0, 1)$$

Par lecture des tables de la distribution normale on établit la valeur de  $Z_\alpha$ .

**3.2.-** Calcul de la valeur  $Z_\beta$ . Nous avons

$$P\left(\frac{\bar{X} - \theta_1}{\sigma/\sqrt{n}} > Z_\beta\right) = \beta; Z_\beta \sim \mathcal{N}(0, 1)$$

Par lecture des tables de la distribution normale on établit la valeur de  $Z_\beta$ .

**3.3.-** Calcul de la taille minimale  $n_C$  de l'échantillon :

$$n_C = \frac{(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{(\theta_0 - \theta_1)_D^2}$$

Si  $n_C > n$ , où  $n$  la taille de l'échantillon, il faut recommencer l'échantillonnage avec une valeur de  $n$  plus grande.

**4°.-** Calcul des valeurs des critères  $C_1$  et  $C_2$  :

$$C_1 = \frac{1}{2} (Z_\beta - Z_{\alpha/2}) \sqrt{\frac{\sigma^2}{n} + \frac{(\theta_0 + \theta_1)_D}{2}}; \quad C_2 = 2\theta_0 - C_1$$

**5°.-** Décision :

Si  $\bar{X} \in [C_1, C_2]$ , alors l'hypothèse  $H_0$  est acceptée.

## 9.2.2 Variance $\sigma^2$ inconnue

On remplace la variance  $\sigma^2$  par son estimation non biaisée  $s^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$

calculée à partir de l'échantillon de taille  $n$ .

### 9.2.2.1 Test simple

**1°.-**  $H_0 : \theta = \theta_0$ ;  $H_1 : \theta > \theta_0$  ou  $\theta < \theta_0$ ;  $H_D : \theta = \theta_1$

**2°.-** On fixe les probabilités  $\alpha$  et  $\beta$ .

**3°.-** Calcul de la taille minimale nécessaire  $N$  de l'échantillon :

**3.1.-** Calcul du nombre de degrés de liberté (ddl)

$$v = n - 1$$

**3.2.-** Calcul de la valeur  $T_{\alpha,v}$ . Nous avons

$$P\left(\frac{\theta_0 - \bar{X}}{s/\sqrt{n}} > T_{\alpha,v}\right) = \alpha; T_{\alpha,v} \sim \text{Student à } v \text{ ddl}$$

Par lecture des tables de la distribution de Student on établit la valeur de  $T_{\alpha,v}$ .

**3.3.-** Calcul de la valeur  $T_{\beta,v}$ . Nous avons

$$P\left(\frac{\bar{X} - \theta_1}{s/\sqrt{n}} > T_{\beta,v}\right) = \beta; T_{\beta,v} \sim \text{Student à } v \text{ ddl}$$

Par lecture des tables de la distribution Student on établit la valeur de  $T_{\beta,v}$ .

**3.4.-** Calcul de la taille minimale  $n_C$  de l'échantillon :

$$n_C = \frac{s^2}{(\theta_0 - \theta_1)_D^2} (T_{\alpha,v} + T_{\beta,v})^2$$

Si  $n_C > n$  recommencer en prenant un échantillon de taille plus grande.

**4°.-** Calcul de la valeur du critère  $C$  :

$$C = \pm \frac{1}{2} (T_{\beta,v} - T_{\alpha,v}) \sqrt{\frac{s^2}{n} + \frac{(\theta_0 + \theta_1)_D}{2}}$$

On prend le signe "+" si  $H_1 : \theta < \theta_0$ , on prend le signe "-" si  $H_1 : \theta > \theta_0$

**5°.-** Décision :

Si  $H_1 : \theta > \theta_0$ , alors l'hypothèse  $H_0$  est acceptée si  $\bar{X} \leq C$ .

Si  $H_1 : \theta < \theta_0$ , alors l'hypothèse  $H_0$  est acceptée si  $\bar{X} \geq C$ .

### 9.2.2.2 Test double

**1°.-**  $H_0 : \theta = \theta_0 ; H_1 : \theta \neq \theta_0 ; H_D : |\theta_1 - \theta_0| = d$

**2°.-** On fixe les probabilités  $\alpha$  et  $\beta$ .

**3°.-** Calcul de la taille minimale  $n_C$  de l'échantillon :

**3.1.-** Calcul du nombre de degrés de liberté

$$v = n - 1$$

**3.2.-** Calcul de la valeur  $T_{\alpha,v}$ . Nous avons

$$P\left(\frac{\theta_0 - \bar{X}}{s/\sqrt{n}} > T_{\alpha/2,v}\right) = \frac{\alpha}{2} ; T_{\alpha/2,v} \sim \text{Student à } v \text{ ddl}$$

Par lecture des tables de la distribution de Student on établit la valeur de  $T_{\alpha/2,v}$ .

**3.3.-** Calcul de la valeur  $T_{\beta,v}$ . Nous avons

$$P\left(\frac{\bar{X} - \theta_1}{s/\sqrt{n}} > T_{\beta,v}\right) = \beta ; T_{\beta,v} \sim \text{Student à } v \text{ ddl}$$

Par lecture des tables de la distribution de Student on établit la valeur de  $T_{\beta,v}$ .

**3.4.-** Calcul de la taille minimale  $n_C$  de l'échantillon :

$$n_C = \frac{s^2}{(\theta_0 - \theta_1)_D^2} (T_{\alpha/2,v} + T_{\beta,v})^2$$

Si  $n_C > n$  recommencer en prenant un échantillon de taille plus grande.

**4°.-** Calcul des valeurs des critères  $C_1, C_2$  :

$$C_1 = \frac{1}{2} (T_{\beta,v} - T_{\alpha/2,v}) \sqrt{\frac{s^2}{n} + \frac{(\theta_0 + \theta_1)_D}{2}} ; C_2 = 2\theta_0 - C_1$$

**5°.-** Décision :

Si  $\bar{X} \in [C_1, C_2]$ , alors l'hypothèse  $H_0$  est acceptée.

### 9.3 Test de la variance avec une valeur de référence

On utilise le même cadre théorique que pour les tests des moyennes. On cherche à estimer  $\theta = \sigma^2$  qui est un indicateur de la variabilité de l'échantillon. Pour cette estimation on utilise comme valeur de référence la variance  $s^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$  de l'échantillon.

#### 9.3.0.3 Test simple

1°.-  $H_0 : \theta = \theta_0 ; H_1 : \theta > \theta_0 ; H_D : \theta = \theta_1$

2°.- On fixe les probabilités  $\alpha$  et  $\beta$ .

3°.- Calculer la valeur du rapport des variances :  $\rho^2 = \frac{\theta_1}{\theta_0}$ .

4°.- Calcul de la taille minimale nécessaire  $N$  de l'échantillon :

4.1.- Calcul du nombre de degrés de liberté (ddl)

$$v = n - 1$$

4.2.- Calcul de la valeur  $\chi_{\alpha,v}^2$ . Nous avons

$$P\left(\frac{(n-1)s^2}{\theta_0} > \chi_{\alpha,v}^2\right) = \alpha ; \chi_{\alpha,v}^2 \sim \text{chi deux à } v \text{ ddl}$$

Par lecture des tables de la distribution du chi deux on établit la valeur de  $\chi_{\alpha,v}^2$ .

4.3.- Calcul de la valeur  $\chi_{1-\beta,v}^2$ . Nous avons

$$P\left(\frac{(n-1)s^2}{\theta_1} > \chi_{1-\beta,v}^2\right) = 1 - \beta ; \chi_{1-\beta,v}^2 \sim \text{chi deux à } v \text{ ddl}$$

Par lecture des tables de la distribution du chi deux on établit la valeur de  $\chi_{1-\beta,v}^2$ .

4.4.- Calcul de la taille nécessaire  $n_C$  de l'échantillon par ajustement du rapport :

$$\frac{\chi_{\alpha,v}^2}{\chi_{1-\beta,v}^2} \simeq \rho^2$$

Si  $n_C > n$  recommencer en prenant un échantillon de taille supérieure à  $n_C$ .

4°.- Calcul de la valeur du critère  $C$  :

$$C = \frac{\theta_0}{n-1} \chi_{\alpha,v}^2$$

5°.- Décision :

Si  $s^2 \leq C$ , alors l'hypothèse  $H_0$  est acceptée.

## 9.3.0.4 Test double

- 1°.-  $H_0 : \theta = \theta_0 ; H_1 : \theta \neq \theta_0 ; H_D : \theta = \theta_1 \text{ ou } \theta = \theta_2$
- 2°.- On fixe les probabilités  $\alpha$  et  $\beta$ .
- 3°.- Calculer la valeur du rapport des variances :  $\rho_1^2 = \frac{\theta_1}{\theta_0} ; \rho_2^2 = \frac{\theta_2}{\theta_0}$ .
- 4°.- Calcul de la taille minimale nécessaire  $N$  de l'échantillon :
- 4.1.- Calcul du nombre de degrés de liberté (ddl)

$$v = n - 1$$

- 4.2.- Calcul de la valeur  $\chi_{\alpha/2,v}^2$ . Nous avons

$$P\left(\frac{(n-1)s^2}{\theta_0} > \chi_{\alpha/2,v}^2\right) = \frac{\alpha}{2} ; \chi_{\alpha/2,v}^2 \sim \text{chi deux à } v \text{ ddl}$$

Par lecture des tables de la distribution du chi deux on établit la valeur de  $\chi_{\alpha,v}^2$ .

- 4.3.- Calcul de la valeur  $\chi_{1-\beta,v}^2$ . Nous avons

$$P\left(\frac{(n-1)s^2}{\theta_1} > \chi_{1-\beta,v}^2\right) = 1 - \beta ; \chi_{1-\beta,v}^2 \sim \text{chi deux à } v \text{ ddl}$$

Par lecture des tables de la distribution du chi deux on établit la valeur de  $\chi_{1-\beta,v}^2$ .

- 4.4.- Calcul de la taille  $n_C$  de l'échantillon par ajustement du rapport :

$$\frac{\chi_{\alpha/2,v}^2}{\chi_{1-\beta,v}^2} \simeq \rho_1^2$$

Si  $N > n$  recommencer en prenant un échantillon de taille supérieure à  $N$ .

- 4°.- Calcul des valeurs des critères  $C_1, C_2$  :

$$C_1 = \frac{\theta_0}{n-1} \chi_{\alpha/2,v}^2 ; C_2 = 2\theta_0 - C_1$$

- 5°.- Décision :

Si  $s^2 \in [C_1, C_2]$ , alors l'hypothèse  $H_0$  est acceptée.

## 9.4 Comparaison entre deux moyennes

Nous avons deux séries de mesures représentées par les deux séries des v.a.  $X_1, \dots, X_n$  et  $Y_1, \dots, Y_m$ . Ces deux séries de mesures sont issues

- soit de la même population et, par conséquent, nous voulons tester l'homogénéité de la population ;
- soit de deux populations différentes et, par conséquent, nous voulons tester l'équivalence de deux populations.

### 9.4.1 Variances $\sigma_X^2, \sigma_Y^2$ des populations connues

#### 9.4.1.1 Test simple

1°.-  $H_0 : \theta_X - \theta_Y = 0 ; H_1 : \theta_X - \theta_Y > 0 ; H_D : \theta_X - \theta_Y = d$

2°.- On fixe les probabilités  $\alpha$  et  $\beta$ .

3°.- Calcul de la taille de l'échantillon :

3.1.- Calcul de la valeur  $Z_\alpha$ . Nous avons

$$P\left(\frac{\bar{X} - \bar{Y} - (\theta_X - \theta_Y)}{\sigma_W} > Z_\alpha\right) = \alpha ; Z_\alpha \sim \mathcal{N}(0, 1)$$

avec

$$\sigma_W^2 = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$$

Par lecture des tables de la distribution normale on établit la valeur de  $Z_\alpha$ .

3.2.- Calcul de la valeur  $Z_\beta$ . Nous avons

$$P\left(\frac{(\theta_X - \theta_Y)_D - (\bar{X} - \bar{Y})}{\sigma_W} > Z_\beta\right) = \beta ; Z_\beta \sim \mathcal{N}(0, 1)$$

Par lecture des tables de la distribution normale on établit la valeur de  $Z_\beta$ .

3.3.- Calcul de la taille minimale  $n_C$  de l'échantillon :

$$n_C = \frac{(Z_\alpha + Z_\beta)^2}{(\theta_X - \theta_Y)_D^2} (\sigma_X^2 + \sigma_Y^2)$$

Si  $n_C > n$ , alors il faut recommencer avec un échantillon plus grand.

4°.- Calcul de la valeur du critère  $C$  :

$$C = \frac{1}{2} (Z_\alpha - Z_\beta) \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} + \frac{(\theta_X - \theta_Y)_D}{2}$$

5°.- Décision :

Si  $\bar{X} - \bar{Y} \leq C$ , alors l'hypothèse  $H_0$  est acceptée.

#### 9.4.1.2 Test double

1°.-  $H_0 : \theta_X - \theta_Y = 0 ; H_1 : \theta_X - \theta_Y \neq 0 ; H_D : \theta_X - \theta_Y = d$

2°.- On fixe les probabilités  $\alpha$  et  $\beta$ .

3°.- Calcul de la taille de l'échantillon :

3.1.- Calcul de la valeur  $Z_{\alpha/2}$ . Nous avons

$$P\left(\frac{\bar{X} - \bar{Y} - (\theta_X - \theta_Y)}{\sigma_W} > Z_{\alpha/2}\right) = \frac{\alpha}{2} ; Z_{\alpha/2} \sim \mathcal{N}(0, 1)$$

avec

$$\sigma_W^2 = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$$

Par lecture des tables de la distribution normale on établit la valeur de  $Z_\alpha$ .

**3.2.-** Calcul de la valeur  $Z_\beta$ . Nous avons

$$P\left(\frac{(\theta_X - \theta_Y)_D - (\bar{X} - \bar{Y})}{\sigma_W} > Z_\beta\right) = \beta; Z_\beta \sim \mathcal{N}(0,1)$$

Par lecture des tables de la distribution normale on établit la valeur de  $Z_\beta$ .

**3.3.-** Calcul de la taille minimale  $n_C$  de l'échantillon :

$$n_C = \frac{(Z_{\alpha/2} + Z_\beta)^2}{(\theta_X - \theta_Y)_D^2} (\sigma_X^2 + \sigma_Y^2)$$

Si  $n_C > n$ , alors il faut recommencer avec un échantillon plus grand.

**4°.-** Calcul de la valeur du critère  $C$  :

$$C = \frac{1}{2} (Z_{\alpha/2} - Z_\beta) \sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} + \frac{(\theta_X - \theta_Y)_D}{2}$$

**5°.-** Décision :

Si  $\bar{X} - \bar{Y} \in [-C, C]$ , alors l'hypothèse  $H_0$  est acceptée.

## 9.4.2 Variances des populations inconnues mais égales

### 9.4.2.1 Test simple

**1°.-**  $H_0 : \theta_X - \theta_Y = 0$ ;  $H_1 : \theta_X - \theta_Y > 0$ ;  $H_D : \theta_X - \theta_Y = d$

**2°.-** On fixe les probabilités  $\alpha$  et  $\beta$ .

**3°.-** Calcul de la taille minimale nécessaire  $N$  de l'échantillon :

**3.1.-** Calcul du nombre de degrés de liberté (ddl)

$$v = n + m - 2$$

**3.2.-** Calcul de la valeur  $T_{\alpha,v}$ . Nous avons

$$P\left(\frac{\bar{X} - \bar{Y}}{\sigma_W} > T_{\alpha,v}\right) = \alpha; T_{\alpha,v} \sim \text{Student à } v \text{ ddl}$$

avec

$$\sigma_W^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m}\right)$$

Par lecture des tables de la distribution de Student on établit la valeur de  $T_{\alpha,v}$ .

**3.3.-** Calcul de la valeur  $T_{\beta,v}$ . Nous avons

$$P\left(\frac{(\theta_X - \theta_Y)_D - (\bar{X} - \bar{Y})}{\sigma_W} > T_{\beta,v}\right) = \beta; T_{\beta,v} \sim \text{Student à } v \text{ ddl}$$

Par lecture des tables de la distribution de Student on établit la valeur de  $T_{\beta,v}$ .

**3.4.-** Calcul de la taille  $N$  de l'échantillon :

$$N = \frac{s_X^2 + s_Y^2}{(\theta_X - \theta_Y)_D^2} (T_{\alpha,v} + T_{\beta,v})^2$$

Si  $N > n$  recommencer en prenant un échantillon de taille supérieure à  $N$ .

**4°.-** Calcul de la valeur du critère  $C$  :

$$C = \frac{1}{2} (T_{\alpha,v} - T_{\beta,v}) \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m} + \frac{(\theta_X - \theta_Y)_D}{2}}$$

**5°.-** Décision :

Si  $\bar{X} - \bar{Y} \leq C$ , alors l'hypothèse  $H_0$  est acceptée.

#### 9.4.2.2 Test double

**1°.-**  $H_0 : \theta_X - \theta_Y = 0$ ;  $H_1 : \theta_X - \theta_Y \neq 0$ ;  $H_D : \theta_X - \theta_Y = d$

**2°.-** On fixe les probabilités  $\alpha$  et  $\beta$ .

**3°.-** Calcul de la taille minimale nécessaire  $N$  de l'échantillon :

**3.1.-** Calcul du nombre de degrés de liberté (ddl)

$$v = n + m - 2$$

**3.2.-** Calcul de la valeur  $T_{\alpha/2,v}$ . Nous avons

$$P\left(\frac{\bar{X} - \bar{Y}}{\sigma_W} > T_{\alpha/2,v}\right) = \frac{\alpha}{2}; T_{\alpha/2,v} \sim \text{Student à } v \text{ ddl}$$

avec

$$\sigma_W^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} \left(\frac{1}{n} + \frac{1}{m}\right)$$

Par lecture des tables de la distribution de Student on établit la valeur de  $T_{\alpha/2,v}$ .

**3.3.-** Calcul de la valeur  $T_{\beta,v}$ . Nous avons

$$P\left(\frac{(\theta_X - \theta_Y)_D - (\bar{X} - \bar{Y})}{\sigma_W} > T_{\beta,v}\right) = \beta; T_{\beta,v} \sim \text{Student à } v \text{ ddl}$$

Par lecture des tables de la distribution de Student on établit la valeur de  $T_{\beta,v}$ .

**3.4.-** Calcul de la taille nécessaire  $N$  de l'échantillon :

$$N = \frac{s_X^2 + s_Y^2}{(\theta_X - \theta_Y)_D^2} (T_{\alpha/2,v} + T_{\beta,v})^2$$

Si  $N > n$  recommencer en prenant un échantillon de taille supérieure à  $N$ .

4°.- Calcul de la valeur du critère  $C$  :

$$C = \frac{1}{2} (T_{\alpha/2, v} - T_{\beta, v}) \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}} + \frac{(\theta_X - \theta_Y)_D}{2}$$

5°.- Décision :

Si  $\bar{X} - \bar{Y} \in [-C, C]$ , alors l'hypothèse  $H_0$  est acceptée.

### 9.4.3 Variances des populations inconnues et non égales

#### 9.4.3.1 Test simple

1°.-  $H_0 : \theta_X - \theta_Y = 0$ ;  $H_1 : \theta_X - \theta_Y > 0$ .

2°.- On fixe la probabilité  $\alpha$ .

3°.- Calcul du nombre de degrés de liberté (ddl)

$$v = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{\left(\frac{s_X^2}{n}\right)^2}{n+1} + \frac{\left(\frac{s_Y^2}{m}\right)^2}{m+1}} - 2$$

Si  $v > n + m - 2$ , alors il faut recommencer avec des échantillons plus grands.

4°.- Évaluer, en utilisant les tables de la distribution de Student, la valeur de  $T_{\alpha, v}$  :

$$P\left(\frac{\bar{X} - \bar{Y}}{s_w} > T_{\alpha, v}\right) = \alpha$$

$$\text{avec } s_w^2 = \frac{s_X^2}{n} + \frac{s_Y^2}{m}$$

5°.- Calcul de la valeur du critère  $C$  :

$$C = T_{\alpha, v}$$

Ensuite on calcule

$$t_1 = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

6°.- Décision :

Si  $t_1 \leq T_{\alpha, v}$ , alors l'hypothèse  $H_0$  est acceptée.

#### 9.4.3.2 Test double

1°.-  $H_0 : \theta_X - \theta_Y = 0$ ;  $H_1 : \theta_X - \theta_Y \neq 0$ .

2°.- On fixe la probabilité  $\alpha$ .

3°.- Calcul du nombre de degrés de liberté (ddl)

$$v = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{\left(\frac{s_X^2}{n}\right)^2}{n+1} + \frac{\left(\frac{s_Y^2}{m}\right)^2}{m+1}} - 2$$

Si  $v > n + m - 2$ , alors il faut recommencer avec des échantillons plus grands.

4°.- Évaluer, en utilisant les tables de la distribution de Student, la valeur de  $T_{\alpha/2,v}$  :

$$P\left(\frac{\bar{X} - \bar{Y}}{s_w} > T_{\alpha,v}\right) = \frac{\alpha}{2}$$

$$\text{avec } s_w^2 = \frac{s_X^2}{n} + \frac{s_Y^2}{m}$$

5°.- Calcul de la valeur du critère :

$$C_1 = -T_{\alpha/2,v}; C_2 = T_{\alpha/2,v}$$

Ensuite on calcule  $t_1$  :

$$t_1 = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

6°.- Décision :

Si  $t_1 \in [-T_{\alpha/2,v}, T_{\alpha/2,v}]$ , alors l'hypothèse  $H_0$  est acceptée.

## 9.5 Comparaison entre deux variances

Nous allons utiliser le même cadre théorique que pour la comparaison entre deux moyennes. On cherche à tester si les variances de deux populations sont identiques. Pour ce test les variances sont estimées en utilisant les valeurs des échantillons à l'aide de la

$$\text{formule } s^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2.$$

### 9.5.0.3 Test simple

1°.-  $H_0 : \theta_X = \theta_Y ; H_1 : \theta_X > \theta_Y ; H_D : \frac{\theta_X}{\theta_Y} = \rho^2.$

2°.- On fixe les probabilités  $\alpha, \beta$ .

3°.- Calcul de la taille minimale nécessaire  $n_C$  de l'échantillon par ajustement de la relation

$$\rho^2 = F_{\alpha;n-1,m-1} \cdot F_{\beta;m-1,n-1}$$

où  $F_{\alpha;n-1,m-1}$  la distribution de Fisher.

Si  $n_C > \min\{n, m\}$ , alors il faut recommencer en prenant des échantillons plus grands.

4°.- Décision :

Si  $\frac{s_X^2}{s_Y^2} \leq F_{\alpha;n-1,m-1}$  , alors l'hypothèse  $H_0$  est acceptée.

#### 9.5.0.4 Test double

1°.-  $H_0 : \theta_X = \theta_Y ; H_1 : \theta_X \neq \theta_Y ; H_D : \frac{\theta_X}{\theta_Y} = \rho^2$  ou  $\frac{\theta_X}{\theta_Y} = \frac{1}{\rho^2}$ .

2°.- On fixe les probabilités  $\alpha$  ,  $\beta$ .

3°.- Calcul de la taille minimale nécessaire  $n_C$  de l'échantillon par ajustement de la relation

$$\rho^2 = F_{\alpha/2;n-1,m-1} \cdot F_{\beta;m-1,n-1}$$

où  $F_{\alpha;n-1,m-1}$  la distribution de Fisher.

Si  $n_C > \min \{n, m\}$  , alors il faut recommencer en prenant des échantillons plus grands.

4°.- Décision :

Si  $\frac{s_X^2}{s_Y^2} \in [F_{1-\alpha/2;n-1,m-1}, F_{\alpha/2;n-1,m-1}]$  , alors l'hypothèse  $H_0$  est acceptée.

# 10

## CONTRÔLE DE RÉCEPTION

---

10.1	Comparaison d'une proportion à une valeur de référence	84
10.1.1	Loi binomiale	84
10.1.2	Loi de Poisson	85
10.1.3	Loi normale	85
10.2	Comparaison de deux proportions	86
10.2.1	Loi hypergéométrique	86
10.2.2	Loi normale	87

---

Le contrôle de réception est destiné à vérifier la conformité d'une production, livrée par lots, à un cahier des charges. Ce contrôle conduit, en général, à la réalisation d'un test statistique, ce qui implique :

- La définition de l'hypothèse nulle  $H_0$ .  
L'élaboration de l'hypothèse nulle se fait en utilisant les spécifications du cahier des charges qui peuvent se rapporter
  - à la valeur moyenne d'une caractéristique du lot ;
  - à la proportion d'un type d'articles dans le lot ;
  - à la dispersion des résultats des mesures d'une caractéristique du lot.
- Le calcul des risques admis  $\alpha$  et  $\beta$ .  
Rappelons que  $\alpha$  est le risque de 1e espèce. Dans le cadre du contrôle de réception, il correspond au risque de refuser le produit quand il est bon, c'est-à-dire  $\alpha$  est le risque du vendeur.  
D'autre part  $\beta$  est le risque de 2e espèce. Il représente le risque d'accepter le produit quand il est mauvais, c'est-à-dire  $\beta$  est le risque de l'acheteur.
- La détermination de la taille  $n$  de l'échantillon.  
Sa valeur est fonction des risques admis  $\alpha$  et  $\beta$  et de l'hypothèse alternative  $H_1$ .
- Le choix du test statistique à utiliser.  
En fonction du cahier des charges nous pouvons être amenés à faire comme test :
  - la comparaison d'une moyenne ou d'une variance à une valeur de référence ;
  - la comparaison d'une proportion à une valeur de référence ;
  - la comparaison de deux ou plusieurs proportions.

Nous présentons dans la suite quelques exemples de contrôle de réception qui sont en complément des tests examinés au chapitre précédent..

## 10.1 Comparaison d'une proportion à une valeur de référence

Supposons qu'une population (un lot) de taille  $N$ , comprend uniquement deux types d'articles  $A$  et  $B$  (e.g. pièces mauvaises :  $A$ , pièces bonnes :  $B$ ). On note par  $p$  la proportion d'articles  $A$  dans le lot. Comparer  $p$  à une proportion de référence  $p_0$  revient à tester l'une des hypothèses suivantes :

$$p \leq p_0 \quad ; \quad p = p_0 \quad ; \quad p \geq p_0$$

Nous prélevons du lot un échantillon de taille  $n$ . Pour que le tirage soit considéré comme un tirage avec remise il faut que  $n \ll N$ . Dans la pratique on prend  $n \leq 0.1 \cdot N$ . On représente par  $x$  le nombre d'articles de type  $A$  que contient l'échantillon. La variable aléatoire correspondante sera notée  $X$ .

Pour effectuer l'un de tests ci-après nous pouvons utiliser trois distributions de probabilité : binomiale, de Poisson et normale.

### 10.1.1 Loi binomiale

On utilise cette loi quand  $n$  est faible (en général  $n < 50$ ). Si on suppose que  $p = p_0$ , nous avons

$$P(X = x | p = p_0) = \frac{p_0^x (1 - p_0)^{n-x} n!}{x! (n-x)!}$$

Pour évaluer le test, on partage le risque de 1e espèce  $\alpha$  en deux parties  $\alpha_1$  et  $\alpha_2$ . Les valeurs de  $\alpha_1$  et  $\alpha_2$  sont fonction de l'hypothèse nulle selon le schéma ci-après :

- Si  $H_0 : p = p_0$ , alors  $\alpha_1 \neq 0, \alpha_2 \neq 0$  et  $\alpha_1 + \alpha_2 = \alpha$ .
- Si  $H_0 : p \leq p_0$ , alors  $\alpha_1 = 0, \alpha_2 = \alpha$ .
- Si  $H_0 : p \geq p_0$ , alors  $\alpha_1 = \alpha, \alpha_2 = 0$ .

On cherche à évaluer l'intervalle  $[x_{\alpha_1}, x_{1-\alpha_2}]$  dans lequel se trouve la probabilité  $p$  avec risque  $\alpha$  et, de ce fait, sera le critère pour l'acceptation du test. Nous avons :

- $x_{\alpha_1}$  est le plus grand entier tel que :

$$P(X < x_{\alpha_1}) = \sum_{x=0}^{x_{\alpha_1}-1} P(X = x | p = p_0) \leq \alpha_1$$

- $x_{1-\alpha_2}$  est le plus petit entier tel que :

$$P(X > x_{1-\alpha_2}) = 1 - \sum_{x=0}^{x_{1-\alpha_2}} P(X = x | p = p_0) \leq \alpha_2$$

$$= \sum_{x=0}^{x_{1-\alpha_2}} P(X = x | p = p_0) \geq 1 - \alpha_2$$

La décision est aussi fonction de l'hypothèse nulle et elle prise comme suit :

- $H_0 : p = p_0$  : Si  $x \in [x_{\alpha_1}, x_{1-\alpha_2}]$  alors l'hypothèse  $H_0$  est acceptée ;

- $H_0 : p \geq p_0$  : Si  $x \geq x_{\alpha_1}$  alors l'hypothèse  $H_0$  est acceptée ;
- $H_0 : p \leq p_0$  : Si  $x \leq x_{1-\alpha_2}$  alors l'hypothèse  $H_0$  est acceptée.

Pour calculer le risque  $\beta$  de 2e espèce, c'est-à-dire le risque de conclure  $p = p_0$  lorsqu'en réalité nous avons  $p = p_1 \neq p_0$ , on utilise la valeur de la probabilité calculée pour  $p = p_1$ . Nous avons ainsi selon l'hypothèse nulle considérée :

- $H_0 : p = p_0$  :  $\beta = P(x_{\alpha_1} \leq X \leq x_{1-\alpha_2}) = P(X \leq x_{1-\alpha_2}) - P(X < x_{\alpha_1})$   

$$= \sum_{x=0}^{x_{1-\alpha_2}} P(X=x | p=p_1) - \sum_{x=0}^{x_{\alpha_1}-1} P(X=x | p=p_1)$$
- $H_0 : p \geq p_0$  :  $\beta = 1 - P(X < x_{\alpha_1}) = 1 - \sum_{x=0}^{x_{\alpha_1}-1} P(X=x | p=p_1)$
- $H_0 : p \leq p_0$  :  $\beta = P(X \leq x_{1-\alpha_2}) = \sum_{x=0}^{x_{1-\alpha_2}} P(X=x | p=p_1)$

### 10.1.2 Loi de Poisson

Nous pouvons appliquer cette loi quand  $p \leq 0.1$  et  $n \geq 50$ . Dans ce cas on suppose que la v.a.  $X$ , qui représente le nombre d'articles de type  $A$  dans l'échantillon de taille  $n$ , suit une loi de Poisson de paramètre  $\lambda_0 = np_0$ . Nous avons donc :

$$P(X = x | p = p_0) = \frac{(np_0)^x}{x!} e^{-np_0}$$

Pour les risques  $\alpha$  et  $\beta$  nous faisons les mêmes calculs que pour la loi binomiale. On utilise la table de la loi de Poisson avec paramètre  $\lambda_0 = np_0$  pour calculer les limites  $x_{\alpha_1}$  et  $x_{1-\alpha_2}$ .

### 10.1.3 Loi normale

Nous pouvons appliquer cette loi quand  $np_0 \geq 30$ . On considère la variable aléatoire centrée réduite  $Z$  dont sa valeur est :

$$z = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$$

où  $x$  est le nombre d'articles de type  $A$  que contient l'échantillon, représentée par la v.a.  $X$ . En utilisant les tables de la loi normale on calcule  $z_{\alpha_1} = z_{1-\alpha_1}$ ,  $z_{1-\alpha_2}$ . La décision est prise comme suit :

- $H_0 : p = p_0$  : Si  $z \in [z_{\alpha_1}, z_{1-\alpha_2}]$  alors l'hypothèse  $H_0$  est acceptée ;
- $H_0 : p \geq p_0$  : Si  $z \geq z_{\alpha_1}$  alors l'hypothèse  $H_0$  est acceptée ;
- $H_0 : p \leq p_0$  : Si  $z \leq z_{1-\alpha_2}$  alors l'hypothèse  $H_0$  est acceptée.

Pour évaluer le risque  $\beta$  on suppose que la v.a.  $X$  suit une loi normale de moyenne  $\mu_1 = np_1$  et de variance  $\sigma_1^2 = np_1(1-p_1)$ . Dans ce cas  $Z$  suit la loi normale de moyenne  $\frac{n-(p_1-p_0)}{\sigma_0}$  et d'écart type  $\frac{\sigma_1}{\sigma_0}$ , où  $\sigma_0 = \sqrt{np_0(1-p_0)}$ . Si on note par  $F(z)$  la fonction de répartition de la loi normale standardisée, nous avons pour la décision :

- $H_0 : p = p_0 : \beta = F \left[ \frac{\sigma_0 z_{1-\alpha_2} - n(p_1 - p_0)}{\sigma_1} \right] - F \left[ \frac{\sigma_0 z_{\alpha_1} - n(p_1 - p_0)}{\sigma_1} \right];$   
 Décision  $H_0$  acceptée si  $x \in \left[ np_0 + z_{\alpha_1} \sqrt{np_0(1-p_0)}, np_0 + z_{1-\alpha_2} \sqrt{np_0(1-p_0)} \right]$
- $H_0 : p \geq p_0 : \beta = 1 - F \left[ \frac{\sigma_0 z_{\alpha_1} - n(p_1 - p_0)}{\sigma_1} \right];$   
 Décision  $H_0$  acceptée si  $x \geq np_0 + z_{\alpha_1} \sqrt{np_0(1-p_0)}$ .
- $H_0 : p \leq p_0 : \beta = F \left[ \frac{\sigma_0 z_{1-\alpha_2} - n(p_1 - p_0)}{\sigma_1} \right];$   
 Décision  $H_0$  acceptée si  $x \leq np_0 + z_{1-\alpha_2} \sqrt{np_0(1-p_0)}$ .

## 10.2 Comparaison de deux proportions

On considère que nous avons deux populations d'effectifs  $N_1$  et  $N_2$ . Chaque population contient deux types d'articles : type A et type B. Supposons que les probabilités réelles (donc inconnues) des articles de type A sont  $p_1$  et  $p_2$  dans les deux populations. L'hypothèse  $H_0$  est l'une de trois suivantes :  $p_1 \leq p_2$  ;  $p_1 = p_2$  ;  $p \geq p_2$ . Afin de tester cette hypothèse on prélève  $n_1$  et  $n_2$  articles de deux populations respectivement. Soient  $x_1$  et  $x_2$  le nombre d'articles de type A dans les deux populations. On calcule les fréquences  $f_1 = \frac{x_1}{n_1}$  et  $f_2 = \frac{x_2}{n_2}$  qui sont des estimations des probabilités  $p_1$  et  $p_2$ . Par convention on affecte l'indice 1 à la population pour laquelle la fréquence des individus A observés est la plus faible, i.e.

$$f_1 = \frac{x_1}{n_1} \leq \frac{x_2}{n_2} = f_2$$

### 10.2.1 Loi hypergéométrique

Nous pouvons l'utiliser quand  $\min \{x_1, x_2, n_1 - x_1, n_2 - x_2\} \leq 18$ .

On transforme le problème comme suit : On réunit les deux populations. On a donc une seule population à  $n = n_1 + n_2$  articles. Cette population contient  $x = x_1 + x_2$  articles de type A. On cherche à évaluer la probabilité que la valeur de la v.a.  $x_1$ , nombre d'articles de type A dans la première population, soit égale à  $x_1$ .  $X_1$  suit une loi hypergéométrique, car

- nous avons une population totale de  $n = n_1 + n_2$  articles ;
- on tire aléatoirement et sans remise  $n_1 = n - n_2$  articles ;
- $x = x_1 + x_2$  articles ont une caractéristique particulière (ils sont de type A).

On cherche donc à calculer la probabilité d'avoir  $x_1$ , articles parmi les  $n_1$ , qui ont cette caractéristique. On a

$$P(X_1 = x_1) = \frac{C_x^n \cdot C_{n-x}^{n_1-x_1}}{C_{n_1}^{n_1}} = \frac{x!(n-x)!n_1!n_2!}{x_1!(x-x_1)!(n_1-x_1)!(n_2-x+x_1)!n!}$$

Pour la décision on fixe le risque  $\alpha$ . Alors on accepte l'hypothèse nulle  $H_0 : p_1 < p_2$

si

$$P(X_1 \leq x_1) = \sum_{k=0}^{x_1} P(X_1 = k) < \alpha$$

Sinon on refuse  $H_0$ . Pour calculer  $\sum_{k=0}^{x_1} P(X_1 = k)$  on utilise les formules

$$P(X_1 = 0) = \frac{(n-x)!n_2!}{(n_2-x)!n!}$$

$$P(X_1 = x_1) = \frac{(x-x_1+1)(n-x_1+1)}{x_1(n_2-x+x_1)} \cdot P(X_1 = x_1 - 1)$$

### 10.2.2 Loi normale

Nous pouvons l'utiliser quand  $\min\{x_1, x_2, n_1 - x_1, n_2 - x_2\} > 18$ . La v.a. normale centrée, réduite qui représente le nombre d'articles de type A, est donnée par

$$z = (n_2x_1 - n_1x_2) \sqrt{\frac{n}{n_1n_2x(n-x)}}$$

On fixe le risque  $\alpha$  et en utilisant les tables de la loi normale on calcule  $z_{\alpha_1} = z_{1-\alpha_1}$ ,  $z_{1-\alpha_2}$ . La décision est prise comme suit :

- $H_0 : p_1 = p_2$  : Si  $z \in [z_{\alpha_1}, z_{1-\alpha_2}]$  alors l'hypothèse  $H_0$  est acceptée ;
- $H_0 : p_1 \geq p_2$  : Si  $z \geq z_{\alpha_1}$  alors l'hypothèse  $H_0$  est acceptée ;
- $H_0 : p_1 \leq p_2$  : Si  $z \leq z_{1-\alpha_2}$  alors l'hypothèse  $H_0$  est acceptée.



## ANALYSE DE VARIANCE

---

11.1 Un facteur contrôlé . . . . .	90
11.2 Deux facteurs contrôlés . . . . .	92

---

Les études que nous faisons en Statistique ont comme but de nous fixer une valeur plausible (*valeur estimée*) d'une des caractéristiques de la population sous test, à partir des valeurs mesurées de cette même caractéristique sur un échantillon issu de la population. Mais il est évident que la valeur mesurée et, a fortiori, la valeur estimée d'une caractéristique peut subir les influences des autres caractéristiques de la population. L'objectif de ce chapitre est de présenter une méthode, l'analyse de variance, qui permet d'étudier ces influences.

Avant de présenter le cadre général de cette méthode, nous allons établir le vocabulaire à utiliser. La caractéristique pour laquelle nous voulons obtenir une estimation est la caractéristique sous test. ( Rappelons que caractéristique se dit aussi variable ou, encore, facteur ). La caractéristique (ou les caractéristiques) dont nous voulons mesurer l'influence sur les résultats, s'appelle caractéristique ou facteur contrôlé. Les différentes valeurs que le facteur contrôlé peut prendre s'appellent modalités.

Pour évaluer l'influence des caractéristiques contrôlés, on effectue des expériences pendant lesquelles les modalités sont fixées et définies d'avance. (Si ce n'est pas le cas, on parle de caractéristique contrôlée à modalités aléatoires.) Lors d'une expérience les résultats obtenus sont dûs non seulement aux valeurs des facteurs contrôlés mais aussi à d'autres caractéristiques connues ou non, mais, en tout cas, qui ne sont pas pris en compte. L'objectif de l'analyse de variance est d'isoler l'influence des caractéristiques non contrôlées et de caractériser l'influence des caractéristiques contrôlées en décomposant la variance globale des résultats obtenus en deux parties :

- variance due aux fluctuations du ou des facteurs contrôlés ;
- variance résiduelle due aux fluctuations des autres facteurs.

Les hypothèses pour l'application de la méthode sont les suivantes

- (H1) L'erreur sur les résultats suit la distribution normale centrée de variance  $\sigma_0$ .
- (H2) Les facteurs influent sur la moyenne des résultats et pas sur leur variance.
- (H3) Les facteurs contrôlés ont des effets additifs.
- (H4) Les mesures effectuées suivent, par modalité, la loi normale.

En théorie le nombre de facteurs à contrôler peut être quelconque. En pratique les formules sont relativement compliquées pour trois facteurs à contrôler et inextricables pour quatre facteurs et plus. On préférera dans ce cas l'utilisation des méthodes d'analyse de données, comme e.g. l'analyse en composantes principales ou l'analyse factorielle des correspondances.

### 11.1 Un facteur contrôlé

On note, en général, par  $A$  le facteur à contrôler. Supposons que le facteur à contrôler a  $p$  modalités et pour chaque modalité nous avons effectué  $n_i$  mesures,  $i = 1, \dots, p$  dont les résultats seront notés  $x_{ik}, k = 1, \dots, n_i$ . Le but de l'analyse de variance est d'isoler l'influence de tous les autres facteurs et de caractériser l'influence du facteur contrôlé en décomposant la variance globale des résultats expérimentaux en deux parties :

- (1)  $s_A^2$  variance représentative des fluctuations dues au facteur  $A$ .
- (2)  $s_R^2$  variance résiduelle, qui représente les fluctuations dues aux autres facteurs.

Pour obtenir cette décomposition il faut pouvoir exprimer l'action sur le résultat du facteur contrôlé par l'intermédiaire d'un modèle mathématique. Le plus simple est de considérer un modèle linéaire. Supposons que les  $n_T = \sum_{i=1}^p n_i$  mesures peuvent s'exprimer par la relation linéaire :

$$x_{ik} = \mu_x + a_i + e_i ; i = 1, \dots, p, k = 1, \dots, n_i$$

où  $\mu_x$  la moyenne de la population pour cette mesure.  $a_i$  représente la fluctuation par rapport à la moyenne qui est due à l'influence du facteur contrôlé quand il a la valeur représentée par la  $i$ -ième modalité. Ainsi la moyenne de la mesure pour la  $i$ -ième modalité est  $\mu_x + a_i$ .  $e_i$  est caractéristique de la  $i$ -ième modalité. Il représente les fluctuations autour de la moyenne  $\mu_x + a_i$  dues à l'influence des autres facteurs.

L'hypothèse  $H_0$  à tester est l'absence d'influence du facteur contrôlé sur les mesures. Avec le modèle linéaire ceci revient à faire une comparaison de moyennes.

Pour l'application pratique des calculs nous construisons le tableau des données :

	<b>Facteur A</b>			
	<i>Modalités <math>1 \leq i \leq p</math></i>			
<i>Mesures</i>	$x_{11}$	$x_{21}$	$\dots$	$x_{p1}$
	$x_{12}$	$x_{22}$	$\dots$	$x_{p2}$
$1 \leq k \leq n_i$	$\dots$	$\dots$	$\dots$	$\dots$
	$x_{1n_1}$	$x_{2n_2}$	$\dots$	$x_{pn_p}$

Posons :

$$n_T = \sum_{i=1}^p n_i ; \bar{x}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{ik}, i = 1, \dots, p ; \bar{x}_T = \frac{1}{n_T} \sum_{i=1}^p \sum_{k=1}^{n_i} x_{ik}$$

On construit le tableau suivant de l'analyse de variance :

Tableau de l'analyse de variance		
Variance	Somme des carrés	Degrés de liberté
Facteur A	$SS_A = \sum_{i=1}^p n_i (\bar{x}_i - \bar{x}_T)^2$	$v_A = p - 1$
Résiduelle	$SS_R = \sum_{i=1}^p \sum_{k=1}^{n_i} (x_{ik} - \bar{x}_i)^2$	$v_R = n_T - p$
Totale	$SS_T = \sum_{i=1}^p \sum_{k=1}^{n_i} (x_{ik} - \bar{x}_T)^2$	$v_T = n_T - 1$

Pour le calcul on utilise les formules suivantes :

(1) Variance totale

$$SS_T = \sum_{i=1}^p \sum_{k=1}^{n_i} x_{ik}^2 - \frac{\left( \sum_{i=1}^p \sum_{k=1}^{n_i} x_{ik} \right)^2}{n_T}$$

$$SS_A = \sum_{i=1}^p \frac{\left( \sum_{k=1}^{n_i} x_{ik} \right)^2}{n_i} - \frac{\left( \sum_{i=1}^p \sum_{k=1}^{n_i} x_{ik} \right)^2}{n_T}$$

$$SS_R = SS_T - SS_A$$

La fonction discriminante est

$$F_C = \frac{SS_A / v_A}{SS_R / v_R}$$

Si  $F_C > F_{\alpha; v_A, v_R}$  on peut conclure, avec risque de se tromper de  $\alpha\%$ , que le facteur contrôlé ait une influence sur les mesures effectuées. Dans ce cas le risque de 2e espèce (i.e. de conclure à tort que le facteur contrôlé n'ait pas une influence sur les mesures) est d'autant plus grand que  $SS_R$  est grand devant  $SS_A$ .

### 11.2 Deux facteurs contrôlés

On considère deux facteurs à contrôler : le facteur  $A$  avec  $p$  modalités et le facteur  $B$  avec  $q$  modalités. On suppose que pour chaque couple de modalités  $(i, j)$  avec  $i = 1, \dots, p$  et  $j = 1, \dots, q$  on effectue  $n_{ij}$  mesures qui seront notées  $x_{ijk}, k = 1, \dots, n_{ij}$ . Ainsi nous avons  $n_T = npq$  mesures au total. Comme précédemment nous utilisons pour étudier l'influence de deux facteurs un modèle linéaire des mesures :

$$x_{ijk} = \mu_x + a_i + b_j + (ab)_{ij} + e_{ij} ; i = 1, \dots, p, j = 1, \dots, q, k = 1, \dots, n_{ij}$$

dont la signification est analogue à celle du paragraphe précédent. Bien sûr ce modèle n'est valable que si les effets de deux facteurs contrôlés  $a$  et  $B$  sont additifs.

Afin d'effectuer les calculs nous construisons le tableau des données est le suivant :

		<b>Facteur A</b> Modalités $1 \leq i \leq p$		
		...	$i$	...
<b>Facteur B</b> $1 \leq j \leq q$	$\vdots$	...	...	...
	$\vdots$	...	$x_{ij1}$	...
	$\vdots$	...	...	...
	$k$	...	$x_{ijk}$	...
	$\vdots$	...	...	...
		...	$x_{ijn}$	...

Le tableau de l'analyse de variance est le suivant :

<b>Tableau de l'analyse de variance</b>		
<i>Variance</i>	<i>Somme des carrés</i>	<i>Degrés de liberté</i>
<i>Facteur A</i>	$SS_A = nq \sum_{i=1}^p (\bar{x}_i - \bar{x}_T)^2$	$\nu_A = p - 1$
<i>Facteur B</i>	$SS_B = np \sum_{j=1}^q (\bar{x}_j - \bar{x}_T)^2$	$\nu_B = q - 1$
<i>Interaction AB</i>	$SS_{AB} = n \sum_{i=1}^p \sum_{j=1}^q (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x}_T)^2$	$\nu_{AB} = (p - 1)(q - 1)$
<i>Résiduelle</i>	$SS_R = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_{ij})^2$	$\nu_R = pq(n - 1)$
<i>Totale</i>	$SS_T = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^{n_{ij}} (x_{ijk} - \bar{x}_T)^2$	$\nu_T = npq - 1$

Pour le calcul des différentes quantités de ce tableau on utilise les formules sui-

vantes :

$$SS_T = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^{n_i} x_{ijk}^2 - \frac{1}{npq} \left( \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^{n_i} x_{ijk} \right)^2$$

$$SS_A = \sum_{i=1}^p \frac{\left( \sum_{j=1}^q \sum_{k=1}^{n_i} x_{ijk} \right)^2}{nq} - \frac{1}{npq} \left( \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^{n_i} x_{ijk} \right)^2$$

$$SS_B = \sum_{j=1}^q \frac{\left( \sum_{i=1}^p \sum_{k=1}^{n_i} x_{ijk} \right)^2}{np} - \frac{1}{npq} \left( \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^{n_i} x_{ijk} \right)^2$$

$$SS_R = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^{n_i} x_{ijk}^2 - \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q \left[ \sum_{k=1}^{n_i} x_{ijk} \right]^2$$

$$SS_{AB} = SS_T - SS_A - SS_B - SS_R$$

La première fonction discriminante qu'on examine est

$$F_{AB} = \frac{SS_{AB}/v_{AB}}{SS_R/v_R}$$

Si  $F_{AB} > F_{\alpha; v_{AB}, v_R}$ , alors on considère, avec risque de se tromper de  $\alpha\%$ , que l'interaction entre les facteurs  $A$  et  $B$  ait une influence sur les mesures.

Dans ce cas on pose

– pour le facteur  $A$  :

$$F_A = \frac{SS_A/v_A}{SS_{AB}/v_{AB}}$$

et si  $F_A > F_{\alpha; v_A, v_{AB}}$  alors le facteur  $A$  a une influence sur les mesures.

– pour le facteur  $B$  :

$$F_B = \frac{SS_B/v_B}{SS_{AB}/v_{AB}}$$

et si  $F_B > F_{\alpha; v_B, v_{AB}}$  alors le facteur  $B$  a une influence sur les mesures.

Si  $F_{AB} < F_{\alpha; v_{AB}, v_R}$  ou si les modalités  $B$  sont fixes, on utilise pour le facteur  $A$  la fonction discriminante suivante :

$$F_A = \frac{SS_A/v_A}{SS_R/v_R}$$

et si  $F_A > F_{\alpha; v_A, v_R}$ , alors on considère, avec risque de se tromper de  $\alpha\%$ , que le facteur  $A$  ait une influence sur les mesures.

De même si  $F_{AB} < F_{\alpha; \nu_{AB}, \nu_R}$  ou si les modalités  $A$  sont fixes, on utilise pour le facteur  $B$  la fonction discriminante suivante :

$$F_B = \frac{SS_B / \nu_B}{SS_R / \nu_R}$$

et si  $F_B > F_{\alpha; \nu_B, \nu_R}$ , alors on considère, avec risque de se tromper de  $\alpha\%$ , que le facteur  $B$  ait une influence sur les mesures.

# 12

## MODÈLE LINÉAIRE À $m$ VARIABLES EXPLICATIVES

---

12.1	Notations	95
12.2	Modèle linéaire	96
12.3	Propriétés de l'estimation	97
12.4	Coefficient de corrélation multiple	97
12.5	Tests de signification	98
12.5.1	Tests sur $R^2$	98
12.5.2	Comparaison d'un coefficient à une valeur de référence	98
12.5.3	Test sur un ensemble des variables explicatives	98

---

La méthode de régression a pour objet l'étude d'une liaison linéaire entre, d'une part, une variable  $Y$  et, d'autre part, un ensemble de variables  $X_1, \dots, X_m$ . Le modèle mathématique de la liaison est donc de la forme :

$$Y = a_1X_1 + a_2X_2 + \dots + a_mX_m + b + e$$

où  $e \sim \mathcal{N}(0, \sigma^2)$ . Pour réaliser la régression il nous faut une série de  $n$  mesures ( $n \gg m$ ) :

$$x_{i1}, x_{i2}, \dots, x_{im}, y_i \quad ; \quad i = 1, 2, \dots, n$$

De plus on fait l'hypothèse selon laquelle les valeurs des variables  $X_i$  sont mesurées sans erreur.

### 12.1 Notations

Dans la suite nous noterons :

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} & 1 \\ x_{21} & x_{22} & \dots & x_{2m} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} & 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ b \end{bmatrix}$$

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}, \dots, \mathbf{x}_m = \begin{bmatrix} x_{1m} \\ x_{2m} \\ \vdots \\ x_{nm} \end{bmatrix}$$

$$m_y = \frac{1}{n} \sum_{i=1}^n y_i, m_{x_1} = \frac{1}{n} \sum_{i=1}^n x_{i1}, \dots, m_{x_m} = \frac{1}{n} \sum_{i=1}^n x_{im}$$

$$\tilde{\mathbf{X}} = \begin{bmatrix} x_{11} - m_{x_1} & x_{12} - m_{x_2} & \cdots & x_{1m} - m_{x_m} \\ x_{21} - m_{x_1} & x_{22} - m_{x_2} & \cdots & x_{2m} - m_{x_m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} - m_{x_1} & x_{n2} - m_{x_2} & \cdots & x_{nm} - m_{x_m} \end{bmatrix}$$

$$\tilde{\mathbf{y}} = \begin{bmatrix} y_1 - m_y \\ y_2 - m_y \\ \vdots \\ y_n - m_y \end{bmatrix}, \tilde{\mathbf{a}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}, \mathbf{m}_x = \begin{bmatrix} m_{x_1} \\ m_{x_2} \\ \vdots \\ m_{x_m} \end{bmatrix}$$

## 12.2 Modèle linéaire

Le modèle linéaire sur l'ensemble des observations  $[\mathbf{y} \ \mathbf{X}]$  s'exprime par la relation :

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{a} + \mathbf{e}$$

$$(resp. \tilde{\mathbf{y}} = \tilde{\mathbf{X}} \cdot \tilde{\mathbf{a}} + \mathbf{e})$$

Le problème de la regression est l'estimation des valeurs des composantes du vecteur  $\mathbf{a}$

(resp.  $\tilde{\mathbf{a}}$ ) de sorte que  $\mathbf{e}^T \mathbf{e} = \sum_{i=1}^n e_i^2$  soit minimal.

Solution :

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

$$(resp. \hat{\tilde{\mathbf{a}}} = (\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}; \hat{\mathbf{b}} = m_y - \hat{\tilde{\mathbf{a}}}^T \mathbf{m}_x)$$

Organisation des calculs :

On construit la matrice des variances-covariances :

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{YY} & \mathbf{V}_{XY}^T \\ \mathbf{V}_{XY} & \mathbf{V}_{XX} \end{bmatrix}$$

Rappelons que pour deux variables aléatoires  $x_1$  et  $x_2$  mises dans le vecteur  $\mathbf{X} = [x_1 \ x_2]$ , la matrice des variances-covariances est donnée par :

$$\mathbf{V} = E\{[\mathbf{X} - E(\mathbf{X})][\mathbf{X} - E(\mathbf{X})]^T\} = \begin{bmatrix} \sigma_{x_1}^2 & E\{[x_1 - E(x_1)][x_2 - E(x_2)]^T\} \\ E\{[x_1 - E(x_1)][x_2 - E(x_2)]^T\} & \sigma_{x_2}^2 \end{bmatrix}$$

et, par conséquent :

$$\mathbf{V}_{XX} = \frac{1}{n} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}; \quad \mathbf{V}_{XY} = \frac{1}{n} \tilde{\mathbf{X}}^T \tilde{\mathbf{Y}}$$

Nous pouvons aussi calculer l'estimation à l'aide de la formule suivante :

$$\hat{\mathbf{a}} = \mathbf{V}_{XX}^{-1} \mathbf{V}_{XY}$$

### 12.3 Propriétés de l'estimation

On peut démontrer que  $\hat{\mathbf{a}}$  est un estimateur sans biais, i.e.  $E(\hat{\mathbf{a}}) = \mathbf{a}$ .

La variance de  $\hat{\mathbf{a}}$  est :

$$V(\hat{\mathbf{a}}) = s^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

avec

$$s^2 = \frac{1}{n-m-1} \sum_{i=1}^n e_i^2 = \frac{n}{n-m-1} V(\mathbf{e})$$

où  $V(\mathbf{e})$  est la variance de  $\mathbf{e}$ .

Pour le terme constant nous avons :

$$V(b) = \frac{s^2}{n} [1 + \mathbf{m}_X^T V_{XX}^{-1} \mathbf{m}_X]$$

Nous avons aussi  $\hat{\mathbf{a}} - \mathbf{a} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{e}$ .

### 12.4 Coefficient de corrélation multiple

C'est le rapport

$$R^2 = \begin{cases} \frac{\text{variance expliquée par la regression}}{\text{variance totale}} \\ \frac{\text{variance totale} - \text{variance résiduelle}}{\text{variance totale}} \end{cases}$$

Nous avons :

$$R^2 = \frac{\mathbf{V}_{XY}^T \mathbf{V}_{XX}^{-1} \mathbf{V}_{XY}}{\mathbf{V}_{YY}}$$

D'après la définition de  $R^2$ , nous avons que

$$1 - R^2 = \frac{V(\mathbf{e})}{V(\mathbf{y})}$$

et, par conséquent :

$$s^2 = \frac{n}{n-m-1} V(\mathbf{y}) (1 - R^2)$$

## 12.5 Tests de signification

### 12.5.1 Tests sur $R^2$

Si

$$\frac{n-m-1}{m} \cdot \frac{R^2}{1-R^2} \geq F_{\alpha; m, n-m-1}$$

alors il existe au moins un coefficient de regression qui est significatif.

### 12.5.2 Comparaison d'un coefficient à une valeur de référence

Si

$$\frac{\hat{a}_i - \alpha_i}{s \sqrt{V_{XX}(i, i)}} = \frac{\hat{a}_i - \alpha_i}{\sqrt{V(\hat{a}_i)}} \leq t_{\alpha; n-m-1}$$

alors  $\hat{a}_i$  est proche de la valeur de référence  $\alpha_i$ .

### 12.5.3 Test sur un ensemble des variables explicatives

Considérons les  $q$  premières variables  $x_1, \dots, x_q$  et soit  $\mathbf{X}_q$  la matrice correspondante des données. Si

$$\frac{n-m-1}{m-q+1} \cdot \frac{\mathbf{y}^T [\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T - \mathbf{X}_q(\mathbf{X}_q^T \mathbf{X}_q)^{-1} \mathbf{X}_q^T] \mathbf{y}}{\mathbf{y}^T [\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] \mathbf{y}} \geq F_{\alpha; m-q+1, n-m-1}$$

alors toutes les variables (et pas seulement les  $q$  premières) sont significatives.

# 13

## ANALYSE DE DONNÉES

---

13.1 Calcul des facteurs . . . . .	99
13.2 Visualisation du nuage des observations . . . . .	102
13.3 Relations entre $\mathbb{R}^n$ et $\mathbb{R}^m$ . . . . .	103

---

Le matériel de base pour l'analyse de données (AdD) est constitué par les mesures effectuées sur des systèmes physiques, donc par les valeurs prises par un certain nombre de variables et/ou par les valeurs calculées d'un certain nombre de paramètres (e.g. à l'aide d'une identification). Si nous avons un ensemble de  $n$  systèmes physiques (resp. un ensemble de  $n$  mesures du même système physique à des dates différentes) qui ont des états structurels différents, nous pouvons avoir un ensemble de  $n$  mesures des variables et/ou des valeurs des paramètres que nous mettrons sous la forme d'une matrice de  $n$  lignes et de  $m$  colonnes, où  $m$  est le nombre des variables et/ou paramètres. (resp. si nous avons un seul système, le formalisme reste le même, car chaque fois que nous procédons à un relevé des mesures nous pouvons considérer que nous avons un nouveau système dont son état structurel peut être différent de l'état du système pendant le relevé précédent).

Nous obtenons, donc, un tableau  $\mathbf{X} = (x_{ij})$  de nombres réels, indicé par les ensembles finis  $I$  (observations) et  $J$  (variables et/ou paramètres), avec  $\text{card } I = n$  et  $\text{card } J = m$ . Une ligne  $\mathbf{x}_i = [x_{i1}, \dots, x_{im}]$  peut être considérée comme un point dans l'espace  $\mathbb{R}^m$ . Par conséquent les  $n$  lignes du tableau  $X$  (en réalité les  $n$  observations) sont des points nous pouvons envisager deux démarches : le calcul des facteurs et l'approximation du nuage des observations dans un sous-espace de  $\mathbb{R}^m$  de dimension réduite. Nous allons présenter ces deux démarches et nous montrerons qu'elles aboutissent aux mêmes résultats.

### 13.1 Calcul des facteurs

Disposant du tableau de données  $X$ , il rest possible de vouloir examiner les corrélations entre les différentes colonnes, afin de dégager les variables qui sont susceptibles de

rendre mieux compte des modifications de l'état structurel des systèmes. Dans ce cas nous avons à examiner  $(m^2 + 3m)/2$  paramètres ( $m$  moyennes,  $m$  variances et  $(m^2 - m)/2$  covariances). Si nous faisons l'hypothèse que les corrélations entre les variables sont nulles, alors nous n'avons que  $2m$  paramètres à examiner (car les  $(m^2 - m)/2$  covariances sont nulles. Bien sûr il n'y a aucune raison de supposer que les données du tableau  $\mathbf{X}$  sont non corrélées. Il s'agit, donc, de faire subir à la matrice  $\mathbf{X}$  une transformation de sorte que la matrice résultante soit composée d'éléments non corrélés. Formellement nous cherchons une matrice  $\mathbf{U}$  de format  $(m \times n)$  telle que, si :

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{U} \quad (13.1.1)$$

alors

$$\mu_{\mathbf{Y}} = 0 \quad (13.1.2)$$

et

$$\Gamma_{\mathbf{Y}} = \text{diagonale} \quad (13.1.3)$$

où  $\Gamma_{\mathbf{Y}}$  la matrice de variance-covariance de  $Y$ .

De cette façon nous remplaçons l'ensemble de variables initiales  $\{x_{1,1}, \dots, x_{1,m}\}$  par un ensemble de variables  $\{y_{1,1}, \dots, y_{1,m}\}$  non corrélées entre elles et qui sont calculées comme suit :

$$\begin{bmatrix} y_{1j} \\ \vdots \\ y_{nj} \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \begin{bmatrix} u_{1j} \\ \vdots \\ u_{mj} \end{bmatrix} \quad (13.1.4)$$

ou encore

$$\mathbf{y}_{.j} = \mathbf{X} \mathbf{u}_j \quad (13.1.5)$$

Nous savons que si  $\mathbf{X}$  est une matrice des données et que  $\mathbf{X}_C$  est la matrice centrée correspondante ( $\mathbf{X}_C = \mathbf{X} - \mu_{\mathbf{X}}$ ), alors la matrice de variance-covariance de  $\mathbf{X}$  est donnée par :

$$\Gamma_{\mathbf{X}_C} = \frac{1}{n} \mathbf{X}_C^T \mathbf{X}_C \quad (13.1.6)$$

Dans la suite on fera l'hypothèse que  $\mathbf{X}$  est centrée. Dans ce cas  $\Gamma_{\mathbf{Y}}$ , à cause de (13.1.1) et (13.1.2) et au coefficient  $\frac{1}{n}$  près, s'écrit :

$$\Gamma_{\mathbf{Y}} = \mathbf{U}^T \mathbf{X}^T \mathbf{X} \mathbf{U} = \mathbf{U}^T \Gamma_{\mathbf{X}} \mathbf{U} \quad (13.1.7)$$

On cherche donc à trouver une matrice  $\mathbf{U}$  telle que la matrice  $\Gamma_{\mathbf{Y}}$ , issue de (13.1.7), soit diagonale. Nous voulons donc, effectuer une orthogonalisation de la matrice  $\Gamma_{\mathbf{X}}$  de variance-covariance de  $\mathbf{X}$ , procédure très connue en algèbre linéaire et pour laquelle il y a une infinité de solutions. Afin d'aboutir à une solution unique on impose deux contraintes, en rapport avec le problème considéré :

- Selon la première contrainte, la variance de chaque facteur doit être maximale. La raison de cette contrainte est la suivante : Considérons le premier facteur  $\mathbf{y}_{.1}$ . Si on se limite pour la description des  $n$  observations aux composantes de ce premier facteur, il est évident que nous représenterons mieux ces observations quand la variance de ce facteur par rapport aux observations est maximale (étant, bien entendu, donné que cette variance ne peut pas être supérieure à la variance des observations calculée dans l'espace initial). Car la variance du facteur est égale à la somme des longueurs des projections des vecteurs initiaux sur le sous-espace engendré par le premier facteur. Et nous savons, d'après la méthode des moindres carrés, que l'approximation des points de l'espace initial dans le sous-espace des facteurs, est optimale quand la somme des longueurs des projections est maximale.

La variance du  $j$ -ième facteur s'écrit :

$$\text{var}(\mathbf{y}_{.j}) = \mathbf{y}_{.j}^\top \mathbf{y}_{.j} = (\mathbf{X} \mathbf{u}_j)^\top (\mathbf{X} \mathbf{u}_j) \quad (13.1.8)$$

d'où

$$\text{var}(\mathbf{y}_{.j}) = \mathbf{u}_j^\top (\mathbf{X}^\top \mathbf{X}) \mathbf{u}_j \quad (13.1.9)$$

- La deuxième contrainte découle de la relation (13.1.9). Comme on cherche à maximiser la quantité  $\text{var}(\mathbf{y}_{.j})$ , il convient d'imposer une valeur à la norme du vecteur  $\mathbf{u}_j$ . Sans perte de généralité, nous pouvons exiger que  $\mathbf{u}_j$  soit unitaire, i.e. que

$$\mathbf{u}_j^\top \mathbf{u}_j = 1 ; j = 1, \dots, m \quad (13.1.10)$$

Ainsi, comme nous voulons que les facteurs soient orthogonaux entre eux, i.e. que :

$$\mathbf{u}_j^\top \mathbf{u}_k = 0 \text{ si } j \neq k \quad (13.1.11)$$

nous imposons la contrainte que la matrice  $\mathbf{U}$  soit orthogonale :

$$\mathbf{U}^\top \mathbf{U} = \mathbf{I} \quad (13.1.12)$$

Par conséquent le problème que nous avons à résoudre est le suivant :

Étant donnée la matrice  $\mathbf{X}$  de format  $(n \times m)$ , trouver une matrice  $\mathbf{U}$  de format  $(n \times m)$ , telle que

$$\Gamma_{\mathbf{Y}} = \mathbf{U}^\top \Gamma_{\mathbf{X}} \mathbf{U} \quad (13.1.13)$$

soit diagonale et satisfasse aux contraintes suivantes :

$$\mathbf{U}^\top \Gamma_{\mathbf{X}} \mathbf{U} = \max \quad (13.1.14)$$

$$\mathbf{U}^\top \mathbf{U} = \mathbf{I} \quad (13.1.15)$$

D'après le théorème 8.2 de l'annexe 1, la solution à ce problème est donnée par la matrice  $\mathbf{U}$  dont les colonnes sont les vecteurs propres orthonormés de la matrice symétrique  $\Gamma_{\mathbf{X}} = \mathbf{C}^\top \mathbf{X}$ .

### 13.2 Visualisation du nuage des observations

Il est possible de vouloir effectuer un travail de synthèse sur les observations afin de dégager des regroupements entre systèmes ayant des vecteurs d'observation voisins. Ce qui donne la possibilité de juger sur la capacité des variables mesurées et des paramètres calculés à répondre aux problèmes posés. Ce travail de synthèse nécessite la visualisation des points correspondants aux observations. Mais comme les observations sont des points de  $\mathbb{R}^m$ , il est évident que tout travail de synthèse dévient difficile, voire impossible, dès que  $m$  dépasse 3. D'où la nécessité d'effectuer une approximation du nuage des observations dans un sous-espace de  $\mathbb{R}^m$ , de dimension aussi faible que possible. De façon formelle, nous supposons que nous voulons approximer le nuage des points de  $\mathbb{R}^m$  par un sous-espace à  $p$  dimensions ( $p < m$ ). Il faut, alors, calculer  $p$  vecteurs  $\mathbb{R}^m$ ,  $\mathbf{u}_1, \dots, \mathbf{u}_p$ , tels que :

$$\widehat{\mathbf{X}} = \mathbf{X} \mathbf{U} \quad (13.2.1)$$

ou encore

$$\begin{bmatrix} \widehat{\mathbf{x}}_1 \\ \vdots \\ \widehat{\mathbf{x}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{bmatrix} [\mathbf{u}_1 \cdots \mathbf{u}_p] \quad (13.2.2)$$

où  $\widehat{\mathbf{x}}_i$  est la projection de  $\mathbf{x}_i \in \mathbb{R}^m$  dans  $\mathbb{R}^p$ , à l'aide de la transformation  $U$ .

Pour calculer les vecteurs  $\mathbf{u}_1, \dots, \mathbf{u}_p$ , nous appliquons la méthode des moindres carrés. Nous essayons donc, de rendre minimale la somme des écarts entre les valeurs exactes et les valeurs approximées, i.e.

$$\min \varepsilon^2 = \sum_{i=1}^m \|\mathbf{x}_i\|^2 - \sum_{i=1}^m \|\widehat{\mathbf{x}}_i\|^2 \quad (13.2.3)$$

Comme  $\sum_{i=1}^m \|\mathbf{x}_i\|^2$  est une quantité qui dépend du nuage des observations et elle est fixe pour un nuage donné de points, il suffit de maximiser la quantité

$$\alpha(\widehat{\mathbf{x}}) = \sum_{i=1}^m \|\widehat{\mathbf{x}}_i\|^2 \quad (13.2.4)$$

Résolvons le problème d'approximation pas à pas. Soit  $p = 1$ . Alors nous avons :

$$\widehat{\mathbf{x}}_i = \mathbf{x}_i \mathbf{u}_1; i = 1, \dots, m \quad (13.2.5)$$

Nous avons donc à maximiser la quantité :

$$\alpha(\widehat{\mathbf{x}}) = \sum_{i=1}^m \|\mathbf{x}_i \mathbf{u}_1\|^2 = (\mathbf{X} \mathbf{u}_1)^\top (\mathbf{X} \mathbf{u}_1) \quad (13.2.6)$$

d'où

$$\alpha(\widehat{\mathbf{x}}) = \mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_1 \quad (13.2.7)$$

Nous imposons la condition supplémentaire que le vecteur  $\mathbf{u}_1$  soit unitaire, i.e :

$$\mathbf{u}_1^\top \mathbf{u}_1 = 1 \quad (13.2.8)$$

Nous avons ainsi à résoudre le problème suivant :

Calculer le vecteur  $\mathbf{u}_1 \in \mathbb{R}^m$  tel que

$$\mathbf{u}_1^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_1 = \max \quad (13.2.9)$$

sous la contrainte

$$\mathbf{u}_1^\top \mathbf{u}_1 = 1 \quad (13.2.10)$$

D'après des résultats de l'algèbre linéaire,  $\mathbf{u}_1$  est le vecteur propre orthonormé de la matrice symétrique  $\mathbf{X}^\top \mathbf{X}$ , correspondant à la plus grande valeur propre.

Supposons maintenant que  $p = 2$ . Alors les deux vecteurs  $\mathbf{u}_1$  et  $\mathbf{u}_2$  que nous cherchons à calculer, sont les vecteurs propres orthonormaux correspondants aux deux plus grandes valeurs propres.

Nous pouvons ainsi obtenir les résultats pour  $p = 1, 2, \dots, m$ . Cette solution est identique à la solution obtenue au paragraphe précédent.

### 13.3 Relations entre $\mathbb{R}^n$ et $\mathbb{R}^m$

Pour calculer les coordonnées des observations  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  dans l'espace factoriel, il suffit d'effectuer la multiplication matricielle :

$$\mathbf{Y} = \mathbf{X} \mathbf{U} \quad (13.3.1)$$

De même il est possible de calculer les facteurs correspondant aux variables  $\{x_1, \dots, x_m\}$ . En effectuant la même démarche qu'auparavant pour les observations, nous trouvons que la matrice à diagonaliser est maintenant  $\mathbf{X} \mathbf{X}^\top$ , qui a les mêmes valeurs propres que  $\mathbf{X}^\top \mathbf{X}$  et dont les vecteurs propres orthonormés forment les colonnes de la matrice  $\mathbf{V}$ . Ainsi les coordonnées des variables dans l'espace factoriel sont données par la relation :

$$\mathbf{Z} = \mathbf{X}^\top \mathbf{V} \quad (13.3.2)$$

Remarquons que si nous avons calculé  $\mathbf{U}$ , il n'est pas nécessaire de calculer  $\mathbf{V}$ . En effet nous avons :

$$\text{dans } \mathbb{R}^m : \quad \mathbf{X}^\top \mathbf{X} \mathbf{u}_j = \lambda_j \mathbf{u}_j \quad (13.3.3)$$

$$\text{dans } \mathbb{R}^n : \quad \mathbf{X} \mathbf{X}^\top \mathbf{v}_j = \lambda_j \mathbf{v}_j \quad (13.3.4)$$

avec  $\lambda_1, \dots, \lambda_m$  valeurs propres de  $\mathbf{X}^\top \mathbf{X}$ .

De (13.3.3) nous obtenons :

$$\mathbf{X} \mathbf{X}^\top (\mathbf{X} \mathbf{u}_j) = \lambda_j (\mathbf{X} \mathbf{u}_j) \quad (13.3.5)$$

(13.3.4) et (13.3.5) permettent d'écrire :

$$\mathbf{v}_j = \mathbf{X} \mathbf{u}_j \quad (13.3.6)$$

Néanmoins le vecteur  $\mathbf{v}_j$  donné par (13.3.6) n'est pas unitaire. En effet nous avons :

$$\mathbf{v}_j^\top \mathbf{v}_j = \mathbf{u}_j^\top \mathbf{X}^\top \mathbf{X} \mathbf{u}_j = \lambda_j \quad (13.3.7)$$

Il faut donc normaliser (13.3.6) en divisant par  $\sqrt{\lambda_j}$ . Nous obtenons :

$$\mathbf{v}_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{X} \mathbf{u}_j; \quad \mathbf{v}_j \quad j\text{-ième facteur dans } \mathbb{R}^n \quad (13.3.8)$$

De façon symétrique nous avons aussi :

$$\mathbf{u}_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{X}^\top \mathbf{v}_j; \quad \mathbf{u}_j \quad j\text{-ième facteur dans } \mathbb{R}^m \quad (13.3.9)$$

Considérons maintenant la matrice

$$\Lambda^{1/2} = \text{diag}(\lambda_1^{1/2}, \dots, \lambda_m^{1/2}) \quad (13.3.10)$$

Nous obtenons ainsi la relation :

$$\mathbf{X} = \mathbf{V} \Lambda^{1/2} \mathbf{U}^\top \quad (13.3.11)$$

car (13.3.8) donne :

$$\mathbf{X} \mathbf{u}_j = \lambda_j^{1/2} \mathbf{v}_j \implies \mathbf{X} \mathbf{u}_j \mathbf{u}_j^\top = \lambda_j^{1/2} \mathbf{v}_j \mathbf{u}_j^\top \implies \mathbf{X} \sum_{j=1}^m \mathbf{u}_j \mathbf{u}_j^\top = \sum_{j=1}^m \lambda_j^{1/2} \mathbf{v}_j \mathbf{u}_j^\top \implies \mathbf{X} \mathbf{U} \mathbf{U}^\top =$$

$$\mathbf{V} \Lambda^{1/2} \mathbf{U}^\top$$

Cette relation fournit la une formule de reconstitution du tableau  $\mathbf{X}$ . Si nous utilisons les  $m$  valeurs et vecteurs propres, la reconstitution est exacte. Dans le cas de visualisation des nuages des observations et des variables, nous devons nous contenter d'un nombre  $p$  de valeurs et vecteurs propres, avec  $p < m$ . Nous avons ainsi :

$$\mathbf{X} \mathbf{X} \mathbf{U} = \mathbf{U} \Lambda$$

$$\widehat{\mathbf{X}}_p = \sum_{j=1}^p \lambda_j^{1/2} \mathbf{v}_j \mathbf{u}_j^\top \quad (13.3.12)$$

Un critère pour la perte de l'information est la norme de la différence de deux matrices :

$$\| \mathbf{X} - \widehat{\mathbf{X}}_p \|^2 = \sum_{j=p+1}^m \lambda_j \quad (13.3.13)$$

où nous avons pris comme norme d'une matrice :

$$\| \mathbf{X} \|^2 = \text{tr}(\mathbf{X}^\top \mathbf{X}) \quad (13.3.14)$$

Pour aboutir à (13.3.13), remarquons que (13.3.3) donne :

$$\mathbf{X}^\top \mathbf{X} \mathbf{U} = \mathbf{U} \Lambda$$

d'où

$$\mathbf{X}^\top \mathbf{X} \mathbf{U} \mathbf{U}^\top = \mathbf{U} \Lambda \mathbf{U}^\top$$

et, à cause de (13.1.15), nous obtenons :

$$\mathbf{X}^\top \mathbf{X} = \mathbf{U} \Lambda \mathbf{U}^\top$$

Par conséquent (13.3.14) s'écrit :

$$\| \mathbf{X} \|^2 = \text{Tr}(\mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top) = \text{Tr}(\mathbf{\Lambda} \mathbf{U} \mathbf{U}^\top)$$

d'où finalement

$$\| \mathbf{X} \|^2 = \text{Tr}(\mathbf{\Lambda}) = \sum_{j=1}^m \lambda_j$$

Un autre critère pour la qualité de la reconstitution est le rapport :

$$\tau_p = \frac{\| \widehat{\mathbf{X}}_p \|^2}{\| \mathbf{X} \|^2} = \frac{\sum_{j=1}^p \lambda_j}{\sum_{j=1}^m \lambda_j} \quad (13.3.15)$$

qui est le taux d'inertie relatif aux  $p$  premiers facteurs.

Le choix du nombre  $p$  de facteurs à retenir est une opération délicate et il n'existe pas une méthode générale permettant d'effectuer ce choix.



# 14

## LES PRINCIPALES MÉTHODES DE L'ANALYSE DE DONNÉES

---

14.1 Analyse en composantes principales . . . . .	107
14.2 Analyse factorielle des correspondances . . . . .	108

---

Devant un problème concret l'utilisateur est amené à effectuer une série de transformations préliminaires sur ses données. Selon les transformations adoptées, nous aboutissons à une méthode particulière de l'analyse de données. Dans la suite nous présentons deux méthodes parmi les plus utilisées. L'analyse en composantes principales et l'analyse factorielle des correspondances. Une autre méthode, de même inspiration, l'analyse factorielle discriminante, sera présentée au chapitre suivant.

### 14.1 Analyse en composantes principales

Du tableau ( $n \times m$ ) des observations initiales  $\mathbf{X} = (x_{ij})$ , nous passons au tableau des observations centrées  $W = (w_{ij})$ , à l'aide de la transformation :

$$w_{ij} = x_{ij} - \frac{1}{n} \sum_{k=1}^n x_{kj} \quad (14.1.1)$$

Notons que cette transformation est dissymétrique par rapport aux observations et aux variables.

- Si nous travaillons avec le tableau  $W$ , nous sommes dans le cadre de l'analyse générale, exposée précédemment. Dans ce cas la matrice de variance-covariance dont on cherchera les valeurs et les vecteurs propres est :

$$\Gamma_{\mathbf{X}} = \frac{1}{n} \mathbf{W}^{\top} \mathbf{W} \quad (14.1.2)$$

Par conséquent, d'après (13.3.8) et (13.3.9), les facteurs sont :

– dans  $\mathbb{R}^m$

$$\mathbf{u}_j = \frac{1}{(n\lambda_j)^{1/2}} \mathbf{W}^\top \mathbf{v}_j = \frac{1}{(n\lambda_j)^{1/2}} \sum_{i=1}^n w_{ij} v_{ij} \quad (14.1.3)$$

d'où

$$\mathbf{u}_j = \frac{1}{(n\lambda_j)^{1/2}} \sum_{i=1}^n (x_{ij} - \hat{\mu}_j) v_{ij} \quad (14.1.4)$$

– dans  $\mathbb{R}^n$

$$\mathbf{v}_j = \frac{1}{(n\lambda_j)^{1/2}} \mathbf{W}^\top \mathbf{u}_j = \frac{1}{(n\lambda_j)^{1/2}} \sum_{k=1}^m w_{jk} u_{jk} \quad (14.1.5)$$

d'où

$$\mathbf{v}_j = \frac{1}{(n\lambda_j)^{1/2}} \sum_{k=1}^m (x_{jk} - \hat{\mu}_k) v_{jk} \quad (14.1.6)$$

$\mathbf{u}_j$  (resp.  $\mathbf{v}_j$ ) constituent les composantes principales du nuage des observations (resp. des variables).

L'inconvenient majeur de l'analyse en composantes principales provient du qu'il s'agit d'une analyse qui n'est pas hors métrique. En effet, si une variable a des valeurs numériques de beaucoup différentes par rapport aux autres variables, alors sa part dans la détermination des composantes principales est particulière : soit prépondérante si ses valeurs sont très grandes par rapport aux valeurs des autres variables, soit négligeable si ses valeurs sont petites. Pour éviter ces inconvénients, plusieurs auteurs (cf. e.g. Lebart, 1975) proposent la méthode en composantes principales normées. Elle consiste à effectuer la transformation :

$$w_{ij} = \frac{x_{ij} - \hat{\mu}_j}{\hat{\sigma}_j} \quad (14.1.7)$$

où  $\hat{\sigma}_j^2$  est la variance estimée de la variable  $j = 1, \dots, m$ .

Dans ce cas la matrice à diagonaliser est la matrice des corrélations de  $m$  variables et la visualisation du nuage des points se fait à l'intérieur d'une sphère centrée à l'origine, la distance entre deux points dans la sphère représentant leur corrélation. Néanmoins en procédant à cette transformation nous avons attribué la même variance à toutes les variables, ce qui a comme résultat de faire disparaître les particularités de la variabilité des variables.

## 14.2 Analyse factorielle des correspondances

L'analyse factorielle des correspondances (AFC) s'applique à des tableaux de fréquence, i.e. des tableaux dont chaque élément  $x$  est issu d'une loi de probabilités sur l'ensemble  $I \times J$ . Toutefois, moyennant un certain nombre de précautions (cf. Benzécri, 1973), nous pouvons envisager l'application de cette méthode à des tableaux des mesures. Essentiellement il s'agit de l'appliquer à des tableaux dont toutes les variables sont exprimées selon la même unité de mesure. Si ce n'est pas le cas, nous devons vérifier a posteriori que

les influences des variables sur les facteurs sont du même ordre. Nous pouvons aussi effectuer un codage logique, c'est-à-dire établir, pour chaque mesure, un intervalle de variation maximale qu'on partage en classes d'effectifs équivalents. Ensuite on remplace la mesure de la variable par l'appartenance à une classe, en mettant e.g. 1 à la classe à laquelle appartient la mesure de la variable et 0 aux autres classe. Il faut aussi pour pouvoir appliquer l'AFC que toutes les mesures soient positives.

Étant donné le tableau  $\mathbf{X} = (x_{ij})$ , nous calculons les quantités :

$$x = \sum_{i=1}^n \sum_{j=1}^m x_{ij} \quad (14.2.1)$$

$$p_{ij} = \frac{x_{ij}}{x} \quad (14.2.2)$$

$$p_{i.} = \frac{1}{x} \sum_{j=1}^m x_{ij} \quad (14.2.3)$$

$$p_{.j} = \frac{1}{x} \sum_{i=1}^n x_{ij} \quad (14.2.4)$$

qui nous permettent de passer au tableau des fréquences, à l'aide des transformations :

– dans  $\mathbb{R}^m$

$$r_{ij} = \frac{p_{ij}}{p_{i.}} ; j = 1, \dots, m \quad (14.2.5)$$

Les  $n$  points  $\mathbf{r}_i$  appartiennent à un espace à  $m - 1$  dimensions, car nous avons la relation :

$$\sum_{j=1}^m r_{ij} = \sum_{j=1}^m \frac{p_{ij}}{p_{i.}} = 1 ; i = 1, \dots, n \quad (14.2.6)$$

– dans  $\mathbb{R}^n$

$$s_{ij} = \frac{p_{ij}}{p_{.j}} ; i = 1, \dots, n \quad (14.2.7)$$

De même ces points appartiennent à un espace de dimension  $n - 1$ , car nous avons :

$$\sum_{i=1}^n s_{ij} = \sum_{i=1}^n \frac{p_{ij}}{p_{.j}} = 1 ; j = 1, \dots, m \quad (14.2.8)$$

Nous remarquons que ces deux transformations conduisent à deux tableaux différents.

Dans  $\mathbb{R}^m$  la distance entre deux observations est donnée par la formule euclidienne :

$$d(\mathbf{r}_i, \mathbf{r}_j) = \sum_{j=1}^m \frac{1}{p_{.j}} (r_{ij} - r_{kj})^2 = \sum_{j=1}^m \frac{1}{p_{.j}} \left( \frac{p_{ij}}{p_{i.}} - \frac{p_{kj}}{p_{k.}} \right)^2 \quad (14.2.9)$$

normalisée par  $p_{.j}$  afin d'éviter les variations d'influence des effectifs inégaux.

Cette distance n'étant pas une somme des carrés, l'analyse générale ne peut pas s'appliquer. Cependant si nous faisons la transformation :

$$w_{ij} = \frac{p_{ij}}{\sqrt{p_i \cdot p_j}} \quad (14.2.10)$$

alors nous pouvons utiliser la distance euclidienne entre deux observations :

$$d(\mathbf{w}_i, \mathbf{w}_j) = \sum_{j=1}^m (w_{ij} - w_{kj})^2 = \sum_{j=1}^m \frac{1}{p_j} \left( \frac{p_{ij}}{p_i} - \frac{p_{kj}}{p_k} \right)^2 \quad (14.2.11)$$

Nous appliquons donc, l'analyse générale en utilisant la matrice symétrique :

$$\mathbf{G} = \mathbf{W}^\top \mathbf{W} \quad (14.2.12)$$

dont le terme général est donné par :

$$g_{jl} = \sum_{i=1}^n \frac{p_{ij} p_{il}}{p_i \sqrt{p_j} p_l} \quad (14.2.13)$$

La matrice des vecteurs propres orthonormés de  $\mathbf{G}$  sera de dimension  $(m \times (m - 1))$  car nous ne prendrons pas en compte le premier vecteur propre correspondant à la valeur propre qui est égale à 1, étant donné que la dispersion des projections des observations sur ce vecteur propre est nulle. En effet si  $\lambda$  est une valeur propre de  $\mathbf{G}$  correspondant au vecteur propre  $\mathbf{u}$ , nous avons :

$$\mathbf{G} \mathbf{u} = \lambda \mathbf{u} \implies \quad (14.2.14)$$

$$\sum_{j=1}^m g_{kj} u_j = \lambda u_k ; \quad k = 1, \dots, m \quad (14.2.15)$$

Si nous prenons comme vecteur propre

$$\mathbf{u} = [\sqrt{p_{.1}}, \sqrt{p_{.2}}, \dots, \sqrt{p_{.m}}]^\top \quad (14.2.16)$$

(14.2.15) donne

$$\sum_{j=1}^m g_{kj} \sqrt{p_{.j}} = \lambda \sqrt{p_{.k}} ; \quad k = 1, \dots, m \implies \quad (14.2.17)$$

$$\sum_{j=1}^m \left( \sum_{i=1}^n \frac{p_{ik} p_{ij}}{p_i \sqrt{p_j} p_k} \right) \sqrt{p_{.j}} = \lambda \sqrt{p_{.k}} ; \quad k = 1, \dots, m \implies \quad (14.2.18)$$

$$\sum_{i=1}^n \frac{p_{ik}}{p_i \sqrt{p_k}} \sum_{j=1}^m p_{ij} = \lambda \sqrt{p_{.k}} ; \quad k = 1, \dots, m \implies \quad (14.2.19)$$

$$\sum_{i=1}^n \frac{p_{ik}}{p_i \sqrt{p_k}} p_i = \lambda \sqrt{p_{.k}} ; \quad k = 1, \dots, m \implies \quad (14.2.20)$$

$$\frac{1}{\sqrt{p_{.k}}} \sum_{i=1}^n p_{ik} = \lambda \sqrt{p_{.k}} ; \quad k = 1, \dots, m \implies \quad (14.2.21)$$

$$\lambda = 1 \quad (14.2.22)$$

Les coordonnées de la  $i$ -ième observation dans l'espace factoriel sont données à l'aide de la formule :

$$y_{ij} = \sum_{k=1}^m \frac{p_{ik}}{p_{i.} \sqrt{p_{.k}}} \quad (14.2.23)$$

La projection, donc, de cette observation sur le  $q$ -ième axe factoriel est égale à

$$f_q(i) = \sum_{k=1}^m u_{kq} \frac{p_{ik}}{p_{i.} \sqrt{p_{.k}}}$$

avec  $u_{kq}$  la  $k$ -ième composante du  $q$ -ième vecteur propre. Si nous appliquons cette formule au vecteur propre donné par (14.2.16), qui correspond à la valeur propre  $\lambda = 1$ , nous avons :

$$f_q(i) = \sum_{k=1}^m \sqrt{p_{.k}} \frac{p_{ik}}{p_{i.} \sqrt{p_{.k}}} = \sum_{k=1}^m \frac{p_{ik}}{p_{i.}} = 1 \quad (14.2.24)$$

i.e. la dispersion des projections est nulle.

Comme cette analyse s'applique aux données initiales  $(r_{ij})$  fournies par la relation (14.2.5), les facteurs ne sont pas les vecteurs propres  $\mathbf{u}_1, \dots, \mathbf{u}_{m-1}$ , mais une transformation des ceux-ci donnée par la formule :

$$\phi_{kj} = \frac{u_{kj}}{\sqrt{p_{.k}}}; \quad j = 1, \dots, m-1 \quad (14.2.25)$$

Le calcul des facteurs  $\mathbf{v}_j$  dans  $\mathbb{R}^n$  peut se faire en utilisant la relation (13.3.8) et en remplaçant le tableau  $\mathbf{X}$  par le tableau  $\mathbf{W}$ . Nous avons donc :

$$\mathbf{v}_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{W} \mathbf{u}_j; \quad j = 1, \dots, m-1 \quad (14.2.26)$$

et comme dans  $\mathbb{R}^n$  les facteurs appliqués aux données initiales  $(s_{ij})$  fournies par la relation (14.2.7) sont calculés à l'aide de la relation :

$$\psi_{ij} = \frac{v_{ij}}{\sqrt{p_{i.}}}; \quad j = 1, \dots, m-1 \quad (14.2.27)$$

Par conséquent les relations entre  $\phi_{kj}$  et  $\psi_{ij}$  sont :

$$\psi_{ij} = \frac{1}{\sqrt{\lambda_j}} \sum_{k=1}^m \frac{p_{ik}}{p_{i.}} \phi_{kj}; \quad i = 1, \dots, n; \quad j = 1, \dots, m-1 \quad (14.2.28)$$

$$\phi_{kj} = \frac{1}{\sqrt{\lambda_j}} \sum_{i=1}^n \frac{p_{ik}}{p_{.k}}; \quad k = 1, \dots, n-1; \quad j = 1, \dots, m-1 \quad (14.2.29)$$

Les contributions absolues, qui nous renseignent sur l'influence des variables, sont les carrés des composants des vecteurs propres  $\mathbf{u}_j$  et  $\mathbf{v}_j$ .



## STRUCTURATION DES DONNÉES

---

15.1	Analyse factorielle discriminante . . . . .	114
15.2	Classification . . . . .	115
15.2.1	Algorithme de K-moyennes (K-Means) . . . . .	116
15.2.2	Nuées dynamiques . . . . .	116
15.3	Classement . . . . .	118
15.3.1	N plus proches voisins . . . . .	118
15.3.2	Classement flou . . . . .	119

---

Une question que nous pouvons nous poser à propos d'un ensemble d'observations concerne la présence (ou l'absence) d'une structure dans cet ensemble. Il y a présence d'une structure chaque fois qu'un ensemble d'observations se décompose en parties plus ou moins homogènes. Cette décomposition est caractéristique de la structuration des données et elle doit, en principe, se retrouver pour chaque nouvelle observation. Trois approches de la structuration d'un ensemble de données sont possibles :

- *Discrimination.*- On voudrait mettre en évidence le pouvoir discriminant des variables mesurées, i.e. la capacité des variables de fournir pour les observations une bonne séparation en classes
- *Classification.*- On voudrait pouvoir se rendre compte de la possibilité de séparation en classes distinctes des observations et, dans ce cas, faire apparaître le mieux possible ces classes.
- *Classement.*- On souhaite prendre une décision quant à l'appartenance d'une observation à une classe, en utilisant les mesures des variables pour cette observation.

Pour chacun de ces problèmes il y a plusieurs méthodes dont quelques-unes sont présentées ci-après.

### 15.1 Analyse factorielle discriminante

Nous faisons l'hypothèse que les  $n$  observations de la population d'apprentissage appartiennent à  $q$  classes différentes :  $C_1, \dots, C_q$  suivant leur état structurel et que cette appartenance nous est connue. Nous voulons visualiser le nuage des observations, et ceci en tenant compte de l'appartenance aux classes, de sorte que nous puissions émettre un jugement concernant la capacité des variables mesurées à reproduire la classification établie. Nous serons, par exemple, d'autant plus satisfaits des nos données, que la visualisation fera apparaître une bonne séparation des classes. Si ce n'est pas le cas, nous chercherons à trouver les variables dont l'élimination rendrait meilleure la séparation des classes.

Nous pouvons aussi, en utilisant la même méthode, de faire l'estimation de l'état structurel d'un système. En effet ayant procédé à une première analyse discriminante, nous avons calculé les facteurs correspondants. Pour un nouveau système, dont nous voulons faire le diagnostic, il suffit de le projeter dans l'espace factoriel en utilisant les facteurs. L'estimation de son état structurel se fera à l'aide de la classe dans laquelle le système se trouvera projeté.

Soit donc le tableau ( $n \times m$ ) de données  $\mathbf{X} = (x_{ij})$ . Nous supposons que la classe  $C_k$  a comme effectifs  $n_k$  observations, avec bien entendu :

$$\sum_{k=1}^q n_k = n \quad (15.1.1)$$

Nous notons :

$$\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (15.1.2)$$

la moyenne totale de la variable  $j$  ( $j = 1, \dots, m$ );

$$\hat{\mu}_{kj} = \frac{1}{n} \sum_{x_i \in C_k} x_{ij} \quad (15.1.3)$$

la moyenne de la variable  $j$  ( $j = 1, \dots, m$ ) pour la classe  $C_k$  ( $k = 1, \dots, q$ ).

La matrice de variance-covariance a comme terme général :

$$\gamma_{jl} = \frac{1}{n} \sum_{k=1}^q \sum_{x_i \in C_k} (x_{ij} - \hat{\mu}_{kj}) (x_{il} - \hat{\mu}_{kl}) + \frac{1}{n} \sum_{k=1}^q n_k (\hat{\mu}_{kj} - \hat{\mu}_j) (\hat{\mu}_{kl} - \hat{\mu}_l) \quad (15.1.4)$$

Cette somme peut se mettre sous forme matricielle :

$$\Gamma_{\mathbf{X}} = \mathbf{W} + \mathbf{B} \quad (15.1.5)$$

avec  $\mathbf{W}$  matrice de covariance intra-classe de terme général :

$$w_{jl} = \frac{1}{n} \sum_{k=1}^q \sum_{x_i \in C_k} (x_{ij} - \hat{\mu}_{kj}) (x_{il} - \hat{\mu}_{kl}) \quad (15.1.6)$$

et  $\mathbf{B}$  matrice de covariance inter-classe, de terme général :

$$b_{jl} = \frac{1}{n} \sum_{k=1}^q n_k (\hat{\mu}_{kj} - \hat{\mu}_j) (\hat{\mu}_{kl} - \hat{\mu}_l) \quad (15.1.7)$$

D'après l'analyse générale un facteur  $\mathbf{u}$  a une variance qui se décompose comme suit :

$$\mathbf{u}^\top \Gamma_X \mathbf{u} = \mathbf{u}^\top \mathbf{W} \mathbf{u} + \mathbf{u}' op \mathbf{B} \mathbf{u} \quad (15.1.8)$$

Au vu de cette relation il est naturel de chercher les facteurs qui maximisent la variance inter-classe  $\mathbf{u}^\top \mathbf{B} \mathbf{u}$  et qui minimisent la variance intra-classe  $\mathbf{u}^\top \mathbf{W} \mathbf{u}$ , car de cette façon les classes se séparent au mieux et, en même temps, chaque classe prend la forme la plus compacte possible. Le problème posé est donc le suivant :

Étant données les matrices de variance-covariance  $\mathbf{W}$  et  $\mathbf{B}$  on cherche à trouver un vecteur  $\mathbf{u}$  tel que

$$\frac{\mathbf{u}^\top \mathbf{B} \mathbf{u}}{\mathbf{u}^\top \Gamma_X \mathbf{u}} = \max \quad (15.1.9)$$

ce qui revient à chercher un vecteur  $\mathbf{u}$  tel que

$$\mathbf{u}^\top \mathbf{B} \mathbf{u} = \max \quad (15.1.10)$$

avec la contrainte

$$\mathbf{u}^\top \Gamma_X \mathbf{u} = \mathbf{1} \quad (15.1.11)$$

Nous pouvons facilement voir que le vecteur cherché est le vecteur propre de la matrice  $\Gamma_X^{-1} \mathbf{B}$  correspondant à la plus grande valeur propre.

Remarquons que la matrice  $\Gamma_X^{-1} \mathbf{B}$  n'est pas symétrique, ce qui peut poser des problèmes lors de son diagonalisation. Dans ce cas nous pouvons diagonaliser la matrice :

$$\mathbf{W} = \mathbf{Y} \Gamma_X^{-1} \mathbf{Y} \quad (15.1.12)$$

avec  $\mathbf{Y}$  matrice  $(m \times q)$  de terme général

$$: b_{jk} = \sqrt{\frac{n_k}{n}} (\hat{\mu}_{kj} - \hat{\mu}_j) \quad (15.1.13)$$

et ensuite nous calculons le facteur  $\mathbf{u}$  à l'aide de la transformation :

$$\mathbf{u} = \mathbf{V} \mathbf{v} \quad (15.1.14)$$

où  $\mathbf{v}$  le vecteur propre de la matrice symétrique  $\mathbf{W}$ .

## 15.2 Classification

Le problème de classification peut se définir de la manière suivante : Soit un ensemble fini d'observations, caractérisé par un nombre fini de variables. Il s'agit de rechercher la partition la plus "typée" relativement à la structure de l'ensemble des observations. Par partition la plus typée on entend la partition pour laquelle chaque classe pourrait être représentée, sans que la qualité de cette représentation soit affectée, par un "type" d'observation. Il existe une multitude de méthodes de classification. Nous présentons dans la suite deux de ces méthodes.

### 15.2.1 Algorithme de K-moyennes (K-Means)

Il s'agit d'une vieille méthode toujours en vogue outre-Atlantique. Nous considérons que nous avons une population d'apprentissage formée de  $n$  observations  $\mathbf{x}_i$  réparties en  $K$  classes. Cette répartition nous est inconnue et la tâche de la méthode est de nous en fournir une qui sera optimale par rapport à un critère qui reste à préciser. Par contre le nombre  $K$  de classes est connu. Pour démarrer l'algorithme on forme  $K$  classes en utilisant les  $n$  observations, notées  $C_1(0), \dots, C_K(0)$ , où 0 indique le numéro d'itération. Les étapes de l'algorithme sont les suivants :

1.- On calcule les moyennes  $\mathbf{m}_1(0), \dots, \mathbf{m}_K(0)$  des classes  $C_1(0), \dots, C_K(0)$ .

On pose  $l = 0$ .

2.-  $l \leftarrow l + 1$ .

3.- Affectation des observations aux classes :

L'observation  $\mathbf{x}_i, i = 1, \dots, n$  est affectée à la classe  $C_j(l)$  si  $\|\mathbf{x}_i - \mathbf{m}_j(l-1)\| < \|\mathbf{x}_i - \mathbf{m}_{j'}(l-1)\| \forall j' = 1, \dots, n, j' \neq j$ .

4.- On calcule les nouveaux centres pour chaque classe :

$$\mathbf{m}_j(l) = \frac{1}{n_j} \sum_{\mathbf{x} \in C_j(l)} \mathbf{x}; j = 1, \dots, K$$

avec  $n_j = \text{card}(C_j(l))$ .

5.- Si  $\mathbf{m}_j(l) = \mathbf{m}_j(l-1) \forall j = 1, \dots, K$ , alors l'algorithme a convergé. Sinon on retourne en 2.

Nous voyons ainsi que le critère pour le calcul de la partition optimale est le minimum de la distance entre les observations et leurs classes.

Le comportement de cet algorithme dépend :

- du nombre de classes spécifié par l'utilisateur ;
- du choix des centres initiaux ;
- de l'ordre selon lequel se présentent les observations ;

### 15.2.2 Nuées dynamiques

Afin d'atténuer les inconvénients de l'utilisation d'une métrique euclidienne pour la classification, Diday a proposé l'utilisation de trois métriques :

- Une métrique entre deux observations  $\mathbf{x}$  et  $\mathbf{y}$  qui peut être soit euclidienne, soit euclidienne réduite, soit encore la métrique du  $\chi^2$  ..
- une métrique entre une observation  $\mathbf{x}$  et une classe  $C$ . Nous avons le choix entre trois métriques :
  - métrique du  $\chi^2$  :

$$D(\mathbf{x}_i, C) = \sum_{j=1}^m \frac{1}{x_{.j}} \left[ \frac{x_{ij}}{x_i} - \frac{1}{\sum_{\mathbf{x}_k \in C} x_k} \sum_{\mathbf{x}_k \in C} x_{kj} \right]^2$$

– métrique euclidienne :

$$D(\mathbf{x}_i, C) = \sum_{j=1}^m [x_{ij} - m_{kj}]^2$$

– métrique de Mahalanobis :

$$D(\mathbf{x}_i, C) = (\mathbf{x}_i - \mathbf{m}_C) \Gamma^{-1} (\mathbf{x}_i - \mathbf{m}_C)^\top$$

où  $\Gamma$  matrice de variance-covariance des classes  $C_1, \dots, C_K$  et  $\mathbf{m}_C$  le centre de gravité de la classe  $C$ .

– une métrique de l'agrégation-écartement des classes :

$$R(\mathbf{x}, C) = \frac{D(\mathbf{x}, C)}{\sum_{j=1}^K D(\mathbf{x}, C_j)} \quad \forall \mathbf{x} \in C$$

Dans la suite on suppose que les observations  $\mathbf{x}$  appartiennent à l'ensemble  $\mathcal{X}$ . Les étapes de l'algorithme sont les suivantes :

1.- On tire au hasard  $\sum_{k=1}^K n_k$  élément et on forme les  $K$  noyaux  $E_1, \dots, E_K$  des classes. On initialise  $S_1$  à une très grande valeur.

2.- calcul du centre de gravité  $\mathbf{m}_l = [m_{l1}, \dots, m_{lm}]$  pour chaque noyau  $E_l : m_{lj} = \frac{1}{\text{card}(E_l)} \sum_{\mathbf{x}_i \in E_l} x_{ij}$ .

3.- Calcul des distances  $D(\mathbf{x}, C_l) \quad \forall \mathbf{x} \in \mathcal{X}; \forall l = 1, \dots, K$ .

4.- Formation des classes

$$C_k = \{\mathbf{x} \in \mathcal{X} / D(\mathbf{x}, C_k) \} = \min_{l=1, \dots, K} D(\mathbf{x}, C_l)$$

5.- Calcul de la mesure d'agrégation-écartement  $R(\mathbf{x}, C_l)$  pour chaque classe  $C_l; l = 1, \dots, K \quad \forall \mathbf{x} \in \mathcal{X}$ .

6.- Calcul de

$$S_2 = \sum_{l=1}^K \sum_{\mathbf{x} \in C_l} R(\mathbf{x}, C_l)$$

7.- Si  $\frac{|S_1 - S_2|}{S_1} < \text{seuil}$  alors convergence de l'algorithme.

Sinon  $S_1 \leftarrow S_2$ .

8.- Calcul des nouveaux noyaux :  $E_l$  ensemble ayant  $n_l$  éléments qui réalisent les plus petites valeurs  $R(\mathbf{x}, C_l); l = 1, \dots, K$ .

Retour à l'étape 2.

Il est à remarquer que le choix des noyaux initiaux (amorçage de l'algorithme) n'est pas toujours évident.

### 15.3 Classement

L'objectif du classement est d'affecter une observation à une classe. Il faut donc avoir, au préalable, effectuer une classification, afin de séparer l'ensemble des observations d'apprentissage en classes. Le critère généralement utilisé pour le classement est une mesure de similitude entre l'observation et les classes.

Comme pour la classification, il y a pléthore des méthodes de classement. Nous avons une présentation détaillée dans [6]. Ces méthodes affectent une observation à une et une seule classe. Cette démarche n'est pas toujours très appropriée et ceci pour deux raisons :

- Les classes ne sont pas nécessairement mutuellement exclusives. Une observation sera affectée à une classe, bien qu'elle puisse faire partie des plusieurs classes.
- Toutes les classes ne sont pas bien représentées (ou même pas du tout) dans la population d'apprentissage. Une observation qui appartient à une de ces classes qui sont insuffisamment représentées, risque d'être affectée à une autre classe, plus ou moins similaire.

Il est, donc, préférable dans pareilles situations d'avoir une méthode qui, étant donnée une observation à classer, fournit pour chaque classe de la population d'apprentissage, une "note" de l'appartenance de cette observation à la classe. La valeur de cette note mesure, en quelque sorte, le degré de vraisemblance de l'appartenance de l'observation à la classe. Pour une telle méthode de classement le cadre théorique peut être constitué par les ensembles flous. Parmi plusieurs méthodes de classement qui utilisent les ensembles flous, nous présentons dans la suite deux méthodes.

#### 15.3.1 N plus proches voisins

C'est une méthode de classement qui évalue le degré d'appartenance d'une observation à une classe en utilisant les N plus proches éléments de cette classe avec l'élément à classer (cf. [54]).

Soient q classes  $C_1, \dots, C_q$ . La classe  $C_k$  a  $n_k$  éléments notés  $\mathbf{x}^i \in \mathbb{R}^n, i = 1, \dots, n_k$ . Soit une nouvelle observation  $\mathbf{y} \in \mathbb{R}^n$ . Chaque classe est partagée en deux sous-classes :

- $C_k^S(\mathbf{y})$  de cardinal N ;
- $C_k^D(\mathbf{y})$  de cardinal  $n_k - N$  ;

et telles que :

$$\forall \mathbf{x}^k \in C_k \text{ si } \mathbf{x}^k \in C_k^S(\mathbf{y}) \text{ alors } d(\mathbf{y}, \mathbf{x}^k) \leq \min_{\mathbf{z}^k \in C_k^D(\mathbf{y})} \{d(\mathbf{y}, \mathbf{z}^k)\}$$

En utilisant la sous-classe  $C_k^S(\mathbf{y})$  nous pouvons calculer un indice de distance entre  $\mathbf{y}$  et la classe  $C_k$  :

$$D(\mathbf{x}^k, C_k) = \frac{1}{N} \sum_{\mathbf{x}^k \in C_k^S(\mathbf{y})} d(\mathbf{y}, \mathbf{x}^k) \quad (15.3.1)$$

Afin de calculer le degré d'appartenance  $\mu(\mathbf{y}, C_k)$  nous imposons les contraintes suivantes :

–  $\mu(\mathbf{y}, C_k)$  est inversement proportionnel à l'indice de distance  $D(\mathbf{y}, C_k)$  :

$$\frac{\mu(\mathbf{y}, C_k)}{D(\mathbf{y}, C_k)} = \frac{\mu(\mathbf{y}, C_l)}{D(\mathbf{y}, C_l)} ; \quad k, l = 1, \dots, q ; \quad k \neq l \quad (15.3.2)$$

– Les degrés d'appartenance sont normalisés :

$$\sum_{k=1}^q \mu(\mathbf{y}, C_k) = 1 \quad (15.3.3)$$

(15.3.2) s'écrit :

$$\mu(\mathbf{y}, C_l) = \frac{D(\mathbf{y}, C_k)}{D(\mathbf{y}, C_l)} \mu(\mathbf{y}, C_k) \quad (15.3.4)$$

et (15.3.3) devient

$$\sum_{l=1}^q \frac{D(\mathbf{y}, C_k)}{D(\mathbf{y}, C_l)} \mu(\mathbf{y}, C_k) = 1 \quad (15.3.5)$$

d'où nous obtenons pour le degré d'appartenance de  $\mathbf{y}$  à la classe  $C_k$ , la formule suivante :

$$\mu(\mathbf{y}, C_k) = \frac{\prod_{l=1, l \neq k}^q D(\mathbf{y}, C_l)}{\sum_{l=1}^q \prod_{m=1, m \neq k}^q D(\mathbf{y}, C_m)} ; \quad k = 1, \dots, q \quad (15.3.6)$$

Les paramètres de réglage sont au nombre de deux : le nombre de classes à retenir pour le calcul du degré d'appartenance et le nombre  $N$  de plus proches éléments à considérer dans chaque classe.

Notons que cette méthode peut s'appliquer avec succès quand les classes ont, à peu près, la même dispersion.

### 15.3.2 Classement flou

Si dans la population d'apprentissage les classes sont fixées a priori selon l'état structurel des systèmes observés, il n'est pas rare qu'elles aient des formes et des dispersions qui diffèrent. (Pour une discussion de ces concepts cf. [7]). Dans ce cas le classement doit tenir compte de ces différences. Nous allons présenter une méthode qui calcule le degré d'appartenance d'un élément à une classe en se fondant sur la disposition spatiale de la classe et dont les détails se trouvent dans [4].

Soit une population d'apprentissage composée de  $q$  classes  $C_1, \dots, C_q$  et dont les éléments appartiennent à l'espace  $\mathbb{R}^n$ . Pour chaque élément  $\mathbf{z} \in \mathbb{R}^n$  on définit un indice de distance avec la classe  $C_k$  comme suit :

$$d(\mathbf{z}, C_k) = \frac{1}{\text{card}(C_k) - 1} \sum_{\mathbf{x} \in C_k} d(\mathbf{z}, \mathbf{x}) ; \quad k = 1, \dots, q \quad (15.3.7)$$

Soit maintenant  $\mathbf{y} \in \mathbb{R}^n$  un élément à classer. Chaque classe est partagée, par rapport à  $\mathbf{y}$ , en deux sous-classes, à savoir :

$$- C_k^S(\mathbf{y}) = \{\mathbf{z} \in C_k / d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{z}, C_k)\}$$

$$- C_k^D(\mathbf{y}) = \{\mathbf{z} \in C_k / d(\mathbf{x}, \mathbf{y}) > d(\mathbf{z}, C_k)\}$$

En utilisant ces deux sous-classes, nous pouvons calculer :

– Un indice de distance de similarité entre  $\mathbf{y}$  et  $C_k$  ;  $k = 1, \dots, q$  :

$$S(\mathbf{y}, C_k) = \begin{cases} \frac{1}{\text{card}(C_k^S(\mathbf{y}))} \sum_{\mathbf{z} \in C_k^S(\mathbf{y})} \frac{d(\mathbf{y}, \mathbf{x})}{d(\mathbf{y}, C_k)} & \text{si } C_k^S(\mathbf{y}) \neq \emptyset \\ 1 & \text{si } C_k^S(\mathbf{y}) = \emptyset \end{cases} \quad (15.3.8)$$

– Un indice de distance de dissimilarité entre  $\mathbf{y}$  et  $C_k$  ;  $k = 1, \dots, q$  :

$$D(\mathbf{y}, C_k) = \begin{cases} \frac{1}{\text{card}(C_k^D(\mathbf{y}))} \sum_{\mathbf{z} \in C_k^D(\mathbf{y})} \frac{d(\mathbf{x}, C_k)}{d(\mathbf{y}, \mathbf{x})} & \text{si } C_k^D(\mathbf{y}) \neq \emptyset \\ 1 & \text{si } C_k^D(\mathbf{y}) = \emptyset \end{cases} \quad (15.3.9)$$

Nous remarquons que  $S(\mathbf{y}, C_k) \in [0, 1]$  et  $D(\mathbf{y}, C_k) \in [0, 1]$ .

Le degré d'appartenance  $\mu(\mathbf{y}, C_k)$  de  $\mathbf{y}$  à une classe  $C_k$  doit être fonction de ses relations de proximité avec tous les éléments de la classe. Il convient donc de considérer le degré d'appartenance comme une somme pondérée de deux quantités suivantes :

– Degré d'appartenance  $\mu(\mathbf{y}, C_k^S(\mathbf{y}))$  qui sera fonction de la similarité entre  $\mathbf{y}$  et  $C_k^S(\mathbf{y})$  et de ce fait nous devons avoir :

$$\mu(\mathbf{y}, C_k^S(\mathbf{y})) \geq 1 - S(\mathbf{y}, C_k) \quad (15.3.10)$$

– Degré d'appartenance  $\mu(\mathbf{y}, C_k^D(\mathbf{y}))$  qui sera fonction de la dissimilarité entre  $\mathbf{y}$  et  $C_k^D(\mathbf{y})$  et de ce fait nous devons avoir :

$$\mu(\mathbf{y}, C_k^D(\mathbf{y})) \leq D(\mathbf{y}, C_k) \quad (15.3.11)$$

Nous avons donc que le degré d'appartenance de  $\mathbf{y}$  à  $C_k$  s'exprime comme suit :

$$\mu(\mathbf{y}, C_k) = \lambda [1 - \alpha S(\mathbf{y}, C_k)] + (1 - \lambda) \beta D(\mathbf{y}, C_k) \quad (15.3.12)$$

avec  $\alpha, \beta, \lambda \in [0, 1]$ .

$\alpha$  et  $\beta$  sont des facteurs de pondération afin que les relations (15.3.10) et (15.3.11) soient vérifiées. Leur valeur détermine la forme du degré d'appartenance. Deux valeurs sont particulièrement utiles :

$$\alpha = \frac{\text{card}(C_k^D(\mathbf{y}))}{\text{card}(C_k)} ; \quad \beta = \frac{\text{card}(C_k^S(\mathbf{y}))}{\text{card}(C_k)} \quad (15.3.13)$$

D'autre part dans la majorité de cas nous pouvons prendre  $\lambda = 0.5$  ce qui revient à considérer le degré d'appartenance comme une moyenne entre  $\mu(\mathbf{y}, C_k^S(\mathbf{y}))$  et  $\mu(\mathbf{y}, C_k^D(\mathbf{y}))$ .

# Annexe A

## RAPPELS DES PROBABILITÉS

---

A.1	Définition de la probabilité . . . . .	122
A.1.1	Définition selon le modèle uniforme . . . . .	122
A.1.2	Définition fréquentielle . . . . .	123
A.1.3	Définition axiomatique . . . . .	124
A.2	Des événements aux ensembles . . . . .	126
A.3	Propriétés de la probabilité . . . . .	126
A.4	Continuité des mesures de probabilité . . . . .	128
A.5	Construction d'espaces probabilisés . . . . .	129
A.6	Probabilités conditionnelles . . . . .	130
A.7	Théorème de Bayes . . . . .	131
A.8	Événements indépendants . . . . .	132
A.9	Espace produit. Expériences aléatoires répétées . . . . .	133
A.10	La notion de la variable aléatoire discrète . . . . .	134
A.11	Espérance mathématique et variance d'une v.a. . . . .	135
A.12	Moments . . . . .	138
A.13	La notion de la variable aléatoire continue . . . . .	139
A.14	Fonction de répartition . . . . .	139
A.15	Fonctions de densité . . . . .	141
A.16	Espérance mathématique - Variance . . . . .	142
A.17	Moments - Médiane - Quantiles . . . . .	143
A.18	Fonctions des variables aléatoires . . . . .	145
A.19	Extension au cas des v.a. discrètes . . . . .	145
A.20	Variables aléatoires à plusieurs dimensions . . . . .	146
A.21	Couples de variables aléatoires . . . . .	148
A.21.1	Lois marginales . . . . .	149
A.21.2	Lois conditionnelles . . . . .	150
A.21.3	Moments - Covariance . . . . .	150
A.21.4	Coefficient de corrélation . . . . .	150

---

L'objectif de cette annexe est de présenter un bref rappel des notions essentielles du calcul des probabilités. Comme plusieurs de ces notions nécessitent des connaissances empruntées à la théorie de la mesure, nous donnons aussi, d'une façon succincte, les définitions de ces concepts.

## LES NOTIONS FONDAMENTALES DES PROBABILITÉS

### A.1 Définition de la probabilité

Nous pouvons constater facilement que la notion de probabilité semble être assez intuitive. Par exemple les phrases :

- la probabilité d’avoir « pile » dans le jeu de « pile ou face » est de 50%
- la probabilité d’avoir un six quand on jette un dé est de 1/6

font partie de la vie courante et elles sont compréhensibles par tous. Il est intéressant d’examiner la démarche qui permet de comprendre ces phrases. Elle nous fournira la définition la plus ancienne, dite classique, de la probabilité.

#### A.1.1 Définition selon le modèle uniforme

Quatre étapes sont nécessaires pour arriver à la définition de la probabilité. Ce sont les suivantes :

- (1) **Détermination du phénomène à étudier et des conditions de son déroulement.**- On détermine, d’abord, l’objet de nos calculs, qu’on appelle en Théorie des Probabilités *expérience aléatoire* ou *épreuve* que l’on note  $\mathcal{E}$ . Dans notre cas, l’expérience aléatoire est soit le lancement d’une pièce de monnaie, soit le lancement d’un dé. En général une expérience aléatoire est toute expérience dont le résultat de sa réalisation dépend du hasard comme, par exemple, le lancement d’un dé.
- (2) **Détermination de tous les résultats du phénomène.**- On détermine, ensuite, le nombre de résultats possibles de l’expérience. Ces résultats sont appelés *événements élémentaires* ou *éventualités*. Ici, dans le cas du jeu de « pile ou face », nous avons deux événements élémentaires et six dans le cas du dé.  
Nous noterons par la suite un événement élémentaire par  $\omega$ . L’ensemble des événements élémentaires est appelé *espace des événements* et sera noté par  $\Omega$ . Dans un premier temps, on considère que cet espace a un nombre fini d’événements, e.g.  $N$ . Nous avons ainsi  $\Omega = \{ \omega_1, \dots, \omega_N \}$ .
- (3) **Quantification de la possibilité de réalisation de chaque résultat du phénomène.**- On fait l’hypothèse que tous les événements élémentaires ont, d’une part, la même vraisemblance (chance) de se réaliser lors d’une expérience et, d’autre part, s’excluent mutuellement.
- (4) **Définition de la probabilité d’un résultat du phénomène.**- Afin de déterminer la probabilité d’un événement élémentaire de se réaliser on utilisera la vraisemblance (chance) avec laquelle peut se réaliser cet événement lors d’une expérience. En fonction de l’hypothèse précédente nous avons :

$$P(\omega) = \frac{1}{\text{nombre d'éventualités}} = \frac{1}{N}; \quad \forall \omega \in \Omega$$

Cette définition est historiquement la première qui est apparue. On l'appelle *modèle uniforme de probabilité*, parce qu'elle attribue la même valeur de probabilité à tous les événements élémentaires. Elle est bien adaptée aux jeux du hasard mais, en général, n'est pas satisfaisante car pour son élaboration nous avons effectué à la 3e étape une hypothèse selon laquelle tous les événements élémentaires ont la même chance de se réaliser, c'est-à-dire ils sont équiprobables. Néanmoins il est possible d'envisager des expériences avec des éventualités non équiprobables, e.g. jeu de « pile ou face » avec une pièce de monnaie non équilibrée. Pour palier à cette faiblesse, nous allons utiliser la définition fréquentielle.

### A.1.2 Définition fréquentielle

Pour obtenir cette définition de la probabilité nous allons remplacer les deux dernières étapes ci-dessus par les suivantes :

#### 3' Quantification de la possibilité de réalisation de chaque résultat du phénomène.-

On introduit la notion de la fréquence de réalisation d'un événement élémentaire. Pour fixer les idées supposons qu'une expérience aléatoire a comme résultats  $N$  événements élémentaires  $\omega_1, \omega_2, \dots, \omega_N$ . Supposons aussi que nous avons  $n$  dispositifs, chacun permettant de réaliser cette expérience sous les mêmes conditions. (Par exemple pour le jeu de « pile ou face » nous avons  $n$  pièces de monnaie.) Nous procédons à ces  $n$  expériences et soit  $n_k$  le nombre de fois où l'événement élémentaire  $\omega_k$  se réalise, avec  $n_k \in \{0, 1, \dots, n\}$  et  $k = 1, 2, \dots, N$ . Nous appelons *fréquence d'apparition* de l'événement élémentaire  $\omega_k$  pour les  $n$  expériences, le rapport :

$$f_n(\omega_k) = \frac{n_k}{n}; \quad \forall \omega_k \in \Omega; \quad n \geq 1$$

#### 4' Définition de la probabilité d'un résultat du phénomène.-

On fait l'hypothèse que la limite  $\lim_{n \rightarrow \infty} f_n(\omega_k)$  existe pour tout  $k = 1, 2, \dots, N$ . Dans ce cas nous pouvons définir la probabilité d'un événement élémentaire  $\omega_k$  comme la fréquence d'apparition de cet événement, quand  $n$  devient grand, i.e.

$$P(\omega_k) = \lim_{n \rightarrow \infty} f_n(\omega_k); \quad \forall \omega_k \in \Omega$$

Cette définition, qui englobe la précédente, est la définition *fréquentielle* ou *statistique*. Elle a un caractère empirique car son fondement se trouve au fait que nous pouvons réaliser l'expérience plusieurs fois et sous des conditions identiques. De plus nous ne pouvons calculer la probabilité qu'à condition que la limite de la fréquence existe. Bien que plusieurs expériences remplissent ces conditions, nous pouvons imaginer des cas pour lesquels toutes les conditions ne sont pas remplies, comme, par exemple, quand la limite de la fréquence n'existe pas. La notion de la limite étant une notion mathématique, il devient évident qu'il faut, en ce qui concerne la définition de la probabilité, abandonner l'approche intuitive ou empirique au profit d'une définition mathématique.

### A.1.3 Définition axiomatique

L'objectif est de pouvoir traiter toutes les expériences aléatoires et ne pas s'enfermer à des définitions qui s'adaptent à une partie seulement de ces expériences. C'est-à-dire notre but est de fournir un modèle mathématique général qui peut s'appliquer à toute expérience aléatoire. Il faut, donc, revenir aux deux dernières étapes et les modifier de nouveau comme suit :

**3° Quantification de la possibilité de réalisation pour chaque résultat du phénomène.-**

Étant donné que toute quantification particulière, applicable à une classe de phénomènes, pourrait ne pas s'appliquer à la totalité des phénomènes, nous allons quantifier la possibilité de réalisation d'un résultat de façon très vague : Nous dirons que pour chaque événement élémentaire  $\omega_k \in \Omega$ , nous sommes capables de déterminer un poids  $w(\omega_k) \in \mathbb{R}_+$ . Ce poids représente la chance qu'il a l'événement  $\omega_k$  de se réaliser lors d'une expérience et nous faisons l'hypothèse que nous sommes capables, pour chaque phénomène particulier, de l'évaluer.

**4° Définition de la probabilité d'un résultat du phénomène.-** De façon évidente, nous pouvons avoir pour la probabilité :

$$P(\omega_k) = \frac{w(\omega_k)}{\sum_{\omega_l \in \Omega} w(\omega_l)} ; \forall \omega_k \in \Omega$$

Mais cette définition n'est pas complète, car elle ne nous permet pas d'évaluer la probabilité des événements qui nous intéressent et qui ne sont pas élémentaires, c'est-à-dire qui peuvent se décomposer en événements élémentaires et qu'on appelle *événements*. Il faut donc, créer d'abord une famille de sous-ensembles de  $\Omega$ , composée d'événements et d'événements élémentaires, et établir ensuite la probabilité comme une application de cette famille.

– *Famille de sous-ensembles de  $\Omega$ .*- Cette famille, que l'on notera par  $\mathcal{A}$ , doit être un sous-ensemble de l'ensemble des parties  $\mathcal{P}(\Omega)$  de  $\Omega$ . En règle générale, les événements élémentaires font partie de cette famille, mais non pas en tant qu'éléments de  $\Omega$ , mais en tant que sous-ensembles de  $\Omega$ ; i.e. si  $\omega_k \in \Omega$  alors  $\{\omega_k\} \in \mathcal{A}$ . Cette substitution de la notion de l'événement par celle de l'ensemble permet d'appliquer aux événements les opérations entre ensembles (union, intersection, complément, ...) et les relations d'ordre et d'égalité ( $\subset$ ,  $=$ ). Plus précisément la définition formelle de  $\mathcal{A}$  est la suivante :

**DÉFINITION A.1.1** Soit  $\Omega$  l'espace des événements associé à une expérience  $\mathcal{E}$ . On appelle algèbre (d'événements) sur  $\Omega$  toute famille  $\mathcal{A}$  de parties de  $\Omega$  telle que :

- $\Omega \in \mathcal{A}$
- Si  $A, B \in \mathcal{A}$  alors  $A \cup B \in \mathcal{A}$ ;
- Si  $A \in \mathcal{A}$  alors  $A^c \in \mathcal{A}$ .

Il est aisé de voir que :

- $\emptyset \in \mathcal{A}$

- Si  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots, n$  alors  $\bigcup_{i=1}^n A_i \in \mathcal{A}$
- Si  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots, n$  alors  $\bigcap_{i=1}^n A_i \in \mathcal{A}$

Si  $\Omega$  est un ensemble dénombrable, la notion de l'algèbre n'est pas suffisante pour les réunions ou les intersections infinies. On la remplace par celle de  $\sigma$ -algèbre :

**DÉFINITION A.1.2** Soit  $\Omega$  l'espace (dénombrable ou continu) des événements associé à une expérience  $\mathcal{E}$ . On appelle  $\sigma$ -algèbre ou tribu (d'événements) sur  $\Omega$  toute famille  $\mathcal{A}$  de parties de  $\Omega$  telle que :

- $\Omega \in \mathcal{A}$
- Si  $A_i \in \mathcal{A}$ ,  $i = 1, 2, \dots$  alors  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
- Si  $A \in \mathcal{A}$  alors  $A^c \in \mathcal{A}$ .

Dans la pratique, pour constituer la tribu d'événements on y place d'abord les événements, élémentaires ou non, qui nous intéressent. On complète ensuite la tribu par d'autres événements afin que les propriétés ci-dessus de l'algèbre (ou de la tribu) soient vérifiées.

Nous avons, donc, constitué un couple  $(\Omega, \mathcal{A})$  qui s'appelle *espace probabilisable*. Nous allons maintenant définir la probabilité sur cet espace.

- *Mesure de probabilité sur  $\mathcal{A}$* .- En utilisant l'équivalence entre événements et ensembles, nous définissons la probabilité sur  $\Omega$  comme une mesure sur  $\mathcal{A}$ . La mesure est une application ayant la propriété de l'additivité. Formellement, nous avons :

**DÉFINITION A.1.3** Soit  $\mathcal{A}$  tribu sur  $\Omega$ . On appelle mesure (positive) sur  $\mathcal{A}$  toute application  $\mu : \mathcal{A} \rightarrow \mathbb{R}_+$  qui a les propriétés suivantes :

- (additivité simple). Pour tout  $A_1, A_2 \in \mathcal{A}$  tels que  $A_1 \cap A_2 = \emptyset$  on a :

$$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$$

- ( $\sigma$ -additivité). Pour toute suite  $(A_i)_{i=1,2,\dots}$  d'éléments de  $\mathcal{A}$ , deux à deux disjoints, on a :

$$\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$$

La probabilité peut, à son tour, se définir comme suit :

**DÉFINITION A.1.4** On appelle probabilité sur l'espace  $(\Omega, \mathcal{A})$  toute mesure  $P : \mathcal{A} \rightarrow \mathbb{R}_+$  qui satisfait à l'axiome de normalisation :

$$P(\Omega) = 1$$

On voit ainsi que la probabilité, selon cette approche, est une mesure normée. Cette troisième définition de la probabilité est la définition axiomatique donnée par A. Kolmogorov en 1933. Elle est aujourd'hui le fondement de la théorie mathématique des probabilités et elle contient, comme cas particuliers, les deux définitions précédentes.

La donnée de la mesure de probabilité  $P$  permet de passer du couple  $(\Omega, \mathcal{A})$ , qui forme l'espace probabilisable, au triplet  $(\Omega, \mathcal{A}, P)$ , qui est l'espace *probabilisé*. D'autre part le couple  $(\mathcal{A}, P)$  s'appelle *champ de probabilité*. Il est évident que nous pouvons avoir, pour le même espace d'événements  $\Omega$ , plusieurs champs de probabilité, en fonction de nos objectifs de traitement des données.

Dans la suite nous noterons la mesure de la probabilité de l'événement  $A$ , considérée comme une application dans  $[0, 1]$ , par  $P(A)$ . Par contre une valeur calculée de la probabilité de l'événement  $A$  sera notée par  $p_A$ . Bien évidemment  $p_A \in [0, 1]$ .

## A.2 Des événements aux ensembles

Nous avons vu que pour pouvoir combiner les événements entre eux et calculer la probabilité d'un événement qui se décompose en événements élémentaires, nous avons utilisé le langage de la théorie des ensembles. Ainsi, si nous avons deux événements  $A, B$  relatifs à l'expérience aléatoire  $\mathcal{E}$ , nous avons ipso-facto deux ensembles  $A, B \in \mathcal{P}(\Omega)$  où  $\Omega$  est l'espace d'événements associé à  $\mathcal{E}$ . Pour que cette correspondance soit complète, il faut pouvoir nommer, en langage d'événements, les résultats des opérations ensemblistes, comme e.g.  $A \cup B$  ou  $A \cap B$ . Nous donnons à la table A.1 une liste avec les principales opérations ensemblistes et leurs équivalents en événements.

## A.3 Propriétés de la probabilité

Nous allons présenter dans ce paragraphe, quelques propriétés de la probabilité qui sont très utiles pour les calculs.

On considère un espace probabilisé  $(\Omega, \mathcal{A}, P)$  relatif à une expérience aléatoire  $\mathcal{E}$ .

PROPRIÉTÉ A.3.1 .-  $P$  est une fonction croissante, i.e.

$$\forall A, B \in \mathcal{A} : A \subset B \implies P(A) \leq P(B)$$

PROPRIÉTÉ A.3.2 .-  $\forall A \in \mathcal{A} : P(A) \leq 1$

PROPRIÉTÉ A.3.3 .- Probabilité de l'événement impossible :  $P(\emptyset) = 0$

PROPRIÉTÉ A.3.4 .- Probabilité de l'événement contraire

$$P(A^c) = 1 - P(A)$$

PROPRIÉTÉ A.3.5 .- Sous-additivité

- pour deux événements :

$$\forall A, B \in \mathcal{A} : P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- pour trois événements  $\forall A, B, C \in \mathcal{A} :$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

NOTATIONS	ENSEMBLES	ÉVÉNEMENTS
$A \in \mathcal{A}$	$A$ est un élément de $\mathcal{A}$	$A$ est un 'événement
$A = \emptyset$	$A$ ensemble vide	$A$ est l'événement impossible
$A = \Omega$	Espace entier	$A$ est l'événement certain
$A^c \in \mathcal{A}$	$A^c$ ensemble complémentaire de $A$	$A^c$ événement contraire de $A$
$A \subset B$	$A$ est inclu dans $B$	$A$ implique $B$
$C = A \cup B$ (noté aussi $C = A + B$ )	$C$ est la réunion de $A, B$	$C$ se réalise lorsqu'au moins un de $A, B$ est réalisé
$A \cap B = \emptyset$	$A, B$ ensembles disjoints	$A, B$ événements incompatibles
$C = A \cap B$ (noté aussi $C = A \cdot B$ )	$C$ est la partie commune de $A, B$	$C$ se réalise lorsque $A$ et $B$ se réalisent
$C = A - B = A \cap B^c$	Différence	$C$ se réalise lorsque $A$ se réalise et $B$ ne se réalise pas
$C = A \Delta B = (A - B) \cup (B - A)$	Différence symétrique	$C$ se réalise lorsque ou bien $A$ ou bien $B$ est réalisé (ou exclusif)
$\bigcup_{i=1}^n (A_i) = \Omega, A_i \cap A_j = \emptyset$	Partition de $\Omega$	Système exhaustif ou système complet d'événements

TAB. A.3 – Correspondance ensembles - événements

PROPRIÉTÉ A.3.6 .- Inégalité de Boole :

– pour deux événements :

$$\forall A, B \in \mathcal{A} : P(A \cup B) \leq P(A) + P(B)$$

– pour une suite d'événements :

$$\forall (A_i)_{1 \leq i \leq n} \text{ avec } A_i \in \mathcal{A} : P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

PROPRIÉTÉ A.3.7 Sous- $\sigma$ -additivité : Pour une suite d'événements  $(A_i)$  de  $\mathcal{A}$  nous avons :

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

PROPRIÉTÉ A.3.8 *Additivité des événements disjoints :*

– Pour deux événements

$$\forall A, B \in \mathcal{A} \text{ avec } A \cap B = \emptyset : P(A \cup B) = P(A) + P(B)$$

– Pour une suite d'événements  $(A_i)_{1 \leq i \leq n}$ ,  $A_i \in \mathcal{A}$ , deux à deux disjoints, nous avons :

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

PROPRIÉTÉ A.3.9 *Soit une partition de  $\Omega$  :  $A_1, \dots, A_n$  avec  $A_i \in \mathcal{A} \forall i = 1, \dots, n$ . Alors pour tout événement  $B \in \mathcal{A}$  nous avons :*

$$P(B) = \sum_{i=1}^n P(B \cap A_i)$$

PROPRIÉTÉ A.3.10 *Pour une suite d'événements  $(A_i)_{i=1, \dots}$  de  $\mathcal{A}$  nous avons :*

$$P\left(\bigcap_{i=1}^{\infty} A_i\right) \geq 1 - \sum_{i=1}^{\infty} P(A_i^c)$$

#### A.4 Continuité des mesures de probabilité

Considérons l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  et une suite dénombrable  $(A_n)_{n=1, \dots}$  d'événements  $A_n \in \mathcal{A} \forall i = 1, \dots, n$ . Il est parfois utile de pouvoir évaluer la probabilité à la limite, i.e.  $\lim_{n \rightarrow \infty} P(A_n)$ . Cette évaluation n'est possible que pour des suites soit décroissantes, soit croissantes.

DÉFINITION A.4.1 *La suite d'ensembles  $(A_n)_{n=1, \dots}$  est dite décroissante (resp. croissante) si  $A_n \supseteq A_{n+1}$  (resp.  $A_n \subseteq A_{n+1}$ ) pour tout  $n \geq 1$ .*

Pour une telle suite nous avons :

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n \quad (\text{resp. } \lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n)$$

C'est précisément la possibilité de définir  $\lim_{n \rightarrow \infty} A_n$  qui permet le calcul des probabilités à la limite pour ce type des suites. Nous avons :

PROPRIÉTÉ A.4.1 *Limite d'une suite croissante.- Pour une suite croissante d'événements  $(A_n)_{n=1, \dots}$  de  $\mathcal{A}$  nous avons :*

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcup_n A_n\right) = P\left(\lim_{n \rightarrow \infty} A_n\right)$$

PROPRIÉTÉ A.4.2 Continuité monotone d'une mesure de probabilité.- *Pour une suite décroissante d'événements  $(A_n)_{n=1,\dots}$  de  $\mathcal{A}$  nous avons :*

$$\lim_{n \rightarrow \infty} P(A_n) = P\left(\bigcap_n A_n\right) = P\left(\lim_{n \rightarrow \infty} A_n\right)$$

La notion de la continuité monotone, qui est, pour une mesure de probabilité, l'équivalent de la notion de continuité pour une fonction, permet de définir d'une autre façon, équivalente, la probabilité. En effet nous avons le :

THÉORÈME A.4.1 *Soit l'espace probabilisable  $(\Omega, \mathcal{A})$ . Considérons une mesure  $P : \mathcal{A} \rightarrow \mathbb{R}_+$ , telle que  $P(\Omega) = 1$ . Alors les deux propositions suivantes sont équivalentes :*

- (i)  *$P$  est une mesure de probabilité sur  $(\Omega, \mathcal{A})$ .*
- (ii)  *$P$  a la propriété de l'additivité finie et de la continuité monotone.*

## A.5 Construction d'espaces probabilisés

Dans un espace proabilisé  $(\Omega, \mathcal{A}, P)$ , l'espace d'événements  $\Omega$  peut être

- discret, si  $\Omega$  est un ensemble fini ou infini dénombrable ;
- continu, si  $\Omega$  est un ensemble infini non dénombrable.

Dans le cas d'un espace  $\Omega$  discret pour construire l'espace probabilisé, il faut :

- (i) Attribuer une valeur de probabilité  $P(\{\omega\}) \geq 0$  à tout événement élémentaire  $\omega \in \Omega$  et telle que  $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$ .
- (ii) Identifier la tribu d'événements  $\mathcal{A}$  à l'ensemble des parties  $\mathcal{P}(\Omega)$  de  $\Omega$ .
- (iii) Pour tout événement  $A \in \mathcal{A}$  évaluer la valeur de probabilité selon la formule :

$$P(A) = \sum_{\omega \in \Omega} P(\{\omega\})$$

De plus dans le cas d'un espace d'événements infini dénombrable, il faut s'assurer que la famille des événements sur lesquels nous allons travailler est soit croissante, soit décroissante, afin de pouvoir lui appliquer les résultats du paragraphe précédent.

Si  $\Omega$  est continu, alors on ne peut pas se contenter de donner une valeur de probabilité aux événements élémentaires car il y a des événements  $A \in \mathcal{A}$  qui sont continus et de ce fait ne peuvent pas être représentés par une réunion dénombrable d'événements élémentaires. Il faut donc pouvoir attribuer une valeur de probabilité à des sous-ensembles de  $\Omega$ , qui sont aussi continus. Bien qu'en principe  $\Omega$  peut être n'importe quel ensemble continu, en pratique nous ne travaillons qu'avec la droite des réels  $\mathbb{R}$ . Ce qui signifie que nous travaillerons uniquement avec des expériences aléatoires dont les éventualités sont des intervalles de  $\mathbb{R}$ . Si ce n'est pas le cas, nous pouvons, par l'intermédiaire d'une application, obtenir une correspondance entre les éventualités de l'expérience aléatoire et les intervalles de  $\mathbb{R}$ . Dans la suite nous prenons donc  $\Omega = \mathbb{R}$ . La tribu sur  $\mathbb{R}$  sera engendrée par l'ensemble de demi-droites  $]-\infty, x]$ ,  $x \in \mathbb{R}$ , qui, comme nous pouvons le démontrer, est la plus petite

tribu qui contient tous les points et tous les intervalles de  $\mathbb{R}$ . Cette tribu est notée  $\mathcal{B}_{\mathbb{R}}$  et est appelée *tribu des boréliens* sur  $\mathbb{R}$ . Si on considère un ensemble continu  $\Omega' \subset \mathbb{R}$ , on peut construire sa tribu des boréliens  $\mathcal{B}_{\mathbb{R}}(\Omega')$  qui est la plus petite tribu qui contient tous les points et tous les intervalles de  $\Omega'$ .

Afin de définir une probabilité pour tout événement  $A = ]-\infty, x] \in \mathcal{B}_{\mathbb{R}}$  nous allons passer par l'intermédiaire d'une autre fonction  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  telle que :

- (i)  $f$  est discontinue à un nombre fini de points pour chaque intervalle fini de  $\mathbb{R}$ .
- (ii)  $f$  est intégrable.
- (iii)  $\int_{\mathbb{R}} f(x) dx = 1$ .

Alors on peut montrer que l'application  $P : \mathcal{B}_{\mathbb{R}} \rightarrow \mathbb{R}_+$  définie par

$$P(B) = \int_{\mathbb{R}} f(x) dx; \quad B \in \mathcal{B}_{\mathbb{R}}$$

est une mesure de probabilité sur l'espace probablisable  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ .

L'intégrale est considérée au sens de Lebesgue, afin de tenir compte de n'importe quelle type de tribu. Néanmoins pour les calculs on peut utiliser l'intégrale au sens de Riemann, car, comme nous verrons plus tard, pour  $\mathcal{B}_{\mathbb{R}}$  l'intégrale de  $f(\cdot)$  sur  $B$  définie au sens de Lebesgue a la même valeur avec l'intégrale de  $f(\cdot)$  sur  $B$  définie au sens de Riemann, à condition bien sûr que les deux intégrales existent.

Notons aussi qu'une fonction qui a ces propriétés s'appelle *densité de probabilité*.

## CONDITIONNEMENT ET INDÉPENDANCE

Les notions de conditionnement et de l'indépendance des événements sont des notions très importantes en probabilité. La première renvoie à la possibilité de modifier la probabilité d'un événement après avoir obtenu des informations sur le résultat de l'expérience aléatoire. La seconde facilite les calculs concernant la probabilité de la réalisation simultanée des plusieurs événements.

### A.6 Probabilités conditionnelles

Étant donné que  $B$  s'est réalisé, la probabilité pour que  $A$  se réalise est égale à  $P(A \cap B)$ . Cette définition est exacte si on considère l'espace des événements  $\Omega' = B$ . Mais comme on cherche  $P(A)$  sur  $\Omega$ , il faut normaliser  $P(A \cap B)$  par la probabilité de  $B$  donnée par l'espace probablisé  $(\Omega, \mathcal{A}, P)$ . Nous avons ainsi :

**DÉFINITION A.6.1** Soit l'espace probablisé  $(\Omega, \mathcal{A}, P)$  et un événement  $B \in \mathcal{A}$ , avec  $P(B) \neq 0$ . On appelle *probabilité conditionnelle* de  $A \in \mathcal{A}$  relative à  $B$ , l'application  $P_B$  :

$\mathcal{A} \rightarrow \mathbb{R}_+$ , définie par

$$\mathcal{A} \ni A \mapsto P_B(A) = \frac{P(A \cap B)}{P(B)} \in \mathbb{R}_+$$

On note souvent par  $P(A/B)$  la probabilité conditionnelle  $P_B(A)$  et on l'appelle aussi *probabilité de A si B*.

Il faut bien entendu s'assurer que la définition de l'application  $P_B$  constitue bien une probabilité au sens de la définition A.1.4

La probabilité conditionnelle a les propriétés suivantes :

PROPRIÉTÉ A.6.1 Si  $A \cap B = \emptyset$  alors  $P(A/B) = 0$ .

PROPRIÉTÉ A.6.2 Si  $B \subset A$  alors  $P(A/B) = 1$ .

PROPRIÉTÉ A.6.3 *Probabilités composées (Forme multiplicative de la formule des probabilités conditionnelles) :*

$$\forall A, B \in \mathcal{A} : P(A \cap B) = P(B)P(A/B) = P(A)P(B/A)$$

*Extension de la formule.- Si  $A_1, A_2, \dots, A_n$  événements de l'algèbre  $\mathcal{A}$ , alors :*

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2/A_1) \dots P(A_n/A_1 \cap \dots \cap A_{n-1})$$

## A.7 Théorème de Bayes

Soit un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et  $(B_i)_{i=1, \dots}$  un système complet d'événements, i.e.  $(B_i)_{i=1, \dots}$  est une partition de  $\Omega$  :

$$B_i \cap B_j = \emptyset \quad \forall i \neq j \quad \text{et} \quad \sum_{i=1}^{\infty} P(B_i) = 1$$

On considère que  $P(B_i) > 0 \quad \forall i = 1, 2, \dots$  Nous avons la propriété suivante :

PROPRIÉTÉ A.7.1 (Probabilités totales).- *Pout tout événement  $A \in \mathcal{A}$  nous avons :*

$$P(A) = \sum_{i=1}^{\infty} P(B_i) P(A/B_i)$$

*et le théorème :*

THÉORÈME A.7.1 Formule de Bayes :

$$P(B_i/A) = \frac{P(B_i) P(A/B_i)}{P(A)} = \frac{P(B_i) P(A/B_i)}{\sum_{j=1}^{\infty} P(B_j) P(A/B_j)}$$

Cette formule fournit la probabilité de réalisation de l'événement  $B_i$  ;  $i = 1, 2, \dots$  si  $A$  est déjà réalisé. La formule de Bayes est aussi appelée formule de probabilités des causes car elle peut être vue comme une formule qui fournit la probabilité de la réalisation de l'événement  $A$  est due à la cause  $B_i$ . Cette formule permet donc le calcul de la probabilité des causes ou hypothèses  $B_i$  qui est une *probabilité a posteriori*, c'est-à-dire après la réalisation de l'événement  $A$  et trouver ainsi celle qui - le plus probablement - a conduit à la réalisation de  $A$ . La difficulté ici est de connaître les probabilités  $P(B_i)$  que nous pouvons, dans ce contexte, appeler *probabilités a priori*. Une autre difficulté provient du fait qu'en général on néglige de préciser bien l'espace  $\Omega$ . Une telle négligence peut parfois conduire à des résultats extravagants.

## A.8 Événements indépendants

On se place dans le cadre d'un espace probabilisé  $(\Omega, \mathcal{A}, P)$ .

**DÉFINITION A.8.1** Deux événements  $A, B \in \mathcal{A}$  avec  $P(B) > 0$  sont (stochastiquement) indépendants si la réalisation de l'un n'influe pas sur la réalisation de l'autre, i.e. si

$$P(A/B) = P(A)$$

Nous avons comme conséquence la formule symétrique :

$$P(B/A) = P(B)$$

De cette définition on obtient aussi la propriété suivante :

**PROPRIÉTÉ A.8.1** Soient  $A, B \in \mathcal{A}$  deux événements. Alors les deux propositions suivantes :

- (i)  $A$  et  $B$  sont indépendants ;
- (ii)  $P(A \cap B) = P(A) \cdot P(B)$

sont équivalentes.

Nous pouvons aussi utiliser cette propriété comme définition de l'indépendance entre deux événements. A fortiori nous dirons que deux événements  $A, B \in \mathcal{A}$  ne sont pas indépendants, c'est-à-dire ils sont liés, si  $P(A \cap B) \neq P(A) \cdot P(B)$ .

Il convient ici de faire la différence entre la notion d'événements indépendants et celle d'événements incompatibles. En effet deux événements  $A, B \in \mathcal{A}$  incompatibles, i.e. tels que  $A \cap B = \emptyset$ , sont indépendants si  $P(A) \cdot P(B) = 0$ , c'est-à-dire si l'un au moins des événements  $A, B$  a une probabilité nulle ou, encore, si au moins l'un des  $A, B$  est l'événement impossible. Dans tous les autres cas deux événements incompatibles ne sont pas indépendants.

Une autre propriété utile est la suivante :

**PROPRIÉTÉ A.8.2** Si deux événements  $A, B \in \mathcal{A}$  sont indépendants, alors les couples  $(A, B^c)$ ,  $(A^c, B)$  et  $(A^c, B^c)$  sont constitués d'événements indépendants.

Cette notion de l'indépendance peut être appliquée à  $n$  événements  $A_1, A_2, \dots, A_n$  avec  $A_i \in \mathcal{A}$ ,  $i = 1, \dots, n$ . Nous disons que ces événements sont mutuellement indépendants si

$$P\left(\bigcap_{i \in I_k} A_i\right) = \prod_{i \in I_k} P(A_i) \quad ; \quad I_k \subset \{1, 2, \dots, n\}$$

PROPRIÉTÉ A.8.3 Soit la famille  $(A_i)_{i=1, \dots, n}$  d'événements mutuellement indépendants, avec  $A_i \in \mathcal{A}$ ,  $i = 1, \dots, n$ . Alors la famille  $(A_i^*)_{i=1, \dots, n}$  est aussi une famille d'événements mutuellement indépendants, où par  $A_i^*$  on a désigné soit l'événement  $A_i$ , soit son contraire.

Nous pouvons, enfin, considérer le cas des tribus d'événements indépendantes.

DÉFINITION A.8.2 Soient  $\mathcal{A}$  et  $\mathcal{B}$  deux tribus d'événements sur le même espace  $\Omega$  d'événements élémentaires. On dit qu'elles sont indépendantes, si  $\forall A \in \mathcal{A}$  et  $\forall B \in \mathcal{B}$ ,  $A$  et  $B$  sont indépendants.

Plus généralement, soient  $n$  tribus  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$ . On dit qu'elles sont indépendantes si tout ensemble de  $n$  événements  $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2, \dots, A_n \in \mathcal{A}_n$  est indépendant, i.e. satisfait à la relation :  $P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$ .

Nous avons le :

THÉORÈME A.8.1 Soit l'espace probabilisé  $(\Omega, \mathcal{A}, P)$ . La famille des événements  $(A_i)_{i=1, \dots, n}$  avec  $A_i \in \mathcal{A}$ ,  $i = 1, \dots, n$  est indépendante si et seulement si la famille  $(\mathcal{A}_i = \{A_i, A_i^c, \Omega, \emptyset\})_{i=1, \dots, n}$  est une famille des tribus indépendante, c'est-à-dire si et seulement si

$$P(B_1 \cap B_2 \cap \dots \cap B_n) = P(B_1) \cdot P(B_2) \cdot \dots \cdot P(B_n)$$

avec  $B_i \in \mathcal{A}_i = \{A_i, A_i^c, \Omega, \emptyset\}$ ,  $i = 1, \dots, n$ .

## A.9 Espace produit. Expériences aléatoires répétées

Considérons  $n$  espaces probabilisés discrets  $(\Omega_i, \mathcal{A}_i, P_i)$ ,  $i = 1, \dots, n$ . Formons l'espace probabilisable produit  $(\Omega, \mathcal{A})$  défini par

- $\Omega = \prod_{i=1}^n \Omega_i$ ,
  - $\mathcal{A}$  tribu des parties de  $\Omega$ .
- Si on pose

$$P(\omega_1, \dots, \omega_n) = \prod_{i=1}^n P_i(\omega_i) \quad ; \quad \omega_i \in \Omega_i, \quad i = 1, \dots, n$$

alors  $P$  est une mesure de probabilité sur  $(\Omega, \mathcal{A})$ , car

$$P(\omega_1, \dots, \omega_n) \geq 0$$

et

$$\sum_{(\omega_1, \dots, \omega_n) \in \Omega} P(\omega_1, \dots, \omega_n) = \sum_{(\omega_1, \dots, \omega_n) \in \Omega} P_1(\omega_1) \cdots P_n(\omega_n) = \sum_{\omega_1 \in \Omega_1} P_1(\omega_1) \cdots \sum_{\omega_n \in \Omega_n} P_n(\omega_n) = 1$$

Ainsi nous avons formé l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  qui est *le produit de la famille*  $(\Omega_i, \mathcal{A}_i, P_i)_{i=1, \dots, n}$  d'espaces probabilisés discrets.

Une application importante du produit d'une famille d'espaces probabilisés ce sont les expériences indépendantes répétées.

Considérons une expérience aléatoire et soit  $(\Omega, \mathcal{A}, P')$  l'espace probabilisé associé. Si on répète, dans des conditions identiques,  $n$  fois cette expérience, nous obtiendrons comme résultat un  $n$ -tuple  $(\omega_1, \dots, \omega_n) \in \Omega^n$  qui décrit les résultats successifs des  $n$  expériences aléatoires. Dans la mesure où ces expériences répétées sont indépendantes, c'est-à-dire le résultat d'une quelconque expérience n'a pas d'influence sur les résultats des autres, nous avons :

$$P(\omega_1, \dots, \omega_n) = \prod_{i=1}^n P'(\omega_i)$$

Par conséquent l'espace probabilisé associé à une expérience aléatoire répétée  $n$  fois, de façon indépendante et dans des conditions identiques, est  $(\Omega^n, \mathcal{P}(\Omega^n), P)$  avec

$$P(\omega_1, \dots, \omega_n) = \prod_{i=1}^n P'(\omega_i), \omega_i \in \Omega, i = 1, \dots, n.$$

## VARIABLES ALÉATOIRES

Jusqu'à maintenant la probabilité d'un événement a été examinée en rapport avec une algèbre d'événements. Mais dans la plupart des applications nous avons besoin d'effectuer des calculs et l'algèbre des événements ne s'y prête pas toujours. La notion de la variable aléatoire est liée à ce besoin de pouvoir attribuer une valeur numérique aux événements d'une expérience aléatoire. Ce paragraphe est consacré à la présentation de la notion de la variable aléatoire et aux techniques utilisées pour son traitement.

### A.10 La notion de la variable aléatoire discrète

Les événements d'une expérience aléatoire étant de nature quelconque, ne prêtent pas forcément à un calcul numérique. Une solution consiste à associer à chaque événement une valeur numérique, le calcul sur des événements devenant ainsi un calcul sur des nombres. Nous avons les définitions suivantes :

**DÉFINITION A.10.1** Soient  $(\Omega, \mathcal{A})$  espace probabilisable et  $E \subseteq \mathbb{R}$  ensemble. L'application  $X : \Omega \rightarrow E$  est mesurable si  $\forall x \in E : X^{-1}(x) \in \mathcal{A}$ .

DÉFINITION A.10.2 Soient  $(\Omega, \mathcal{A}, P)$  espace probabilisé et  $E \subset \mathbb{R}$  ensemble discret. On appelle variable aléatoire (v.a.) discrète définie sur  $(\Omega, \mathcal{A})$ , toute application mesurable  $X : \Omega \rightarrow E$ .

Nous voyons ainsi qu'en partant de l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  et de l'ensemble  $E$ , nous avons pu, à l'aide de la v.a.  $X$ , construire un autre espace probabilisé  $(E, \mathcal{B}_X, P_X)$ , engendré par  $X$ .

DÉFINITION A.10.3 Soient  $(\Omega, \mathcal{A}, P)$  espace probabilisé,  $E \subset \mathbb{R}$  ensemble discret et  $X$  v.a. sur  $(\Omega, \mathcal{A})$ . La probabilité  $P_X$  induite par  $X$  est appelée loi de probabilité de la v.a.  $X$ .

Nous avons la propriété suivante :

PROPRIÉTÉ A.10.1 . Si  $X, Y$  deux v.a. sur  $(\Omega, \mathcal{A}, P)$ , alors  $X + Y$  et  $X \cdot Y$  sont aussi des v.a..

PROPRIÉTÉ A.10.2 Soient  $(\Omega, \mathcal{A}, P)$  espace probabilisé,  $E \subset \mathbb{R}$  ensemble discret et  $X : \Omega \rightarrow E$  variable aléatoire. Si  $h$  est une fonction de  $X$ , i.e. si  $h : E \rightarrow E' \subset \mathbb{R}$ , alors  $h(X)$  est aussi une variable aléatoire.

Dans la suite nous noterons par des lettres majuscules  $X, Y, \dots$  les variables aléatoires et par des lettres minuscules  $x, y, \dots$  leurs valeurs.

### A.11 Espérance mathématique et variance d'une v.a.

Considérons une expérience aléatoire  $\mathcal{E}$  et son espace probabilisé discret  $(\Omega, \mathcal{A}, P)$  et soit  $X$  une v.a. sur  $(\Omega, \mathcal{A})$ . Pour fixer les idées, supposons que  $\Omega$  est fini  $\Omega = \{1, 2, \dots, n\}$ . Notons par  $p_k$  la probabilité  $P(\omega_k)$ ;  $k = 1, \dots, n$ , avec  $\sum_{k=1}^n p_k = 1$  et par  $x_k$  la valeur de la v.a.  $X(\omega_k)$ . Supposons maintenant qu'on répète  $N$  fois l'expérience  $\mathcal{E}$ . Si l'événement  $\omega_1$  apparaît  $n_1$  fois pendant les  $N$  répétitions, l'événement  $\omega_2$ ,  $n_2$  fois, ..., l'événement  $\omega_n$ ,  $n_n$  fois, alors la v.a.  $X$  prendra  $n_1$  fois la valeur  $x_1$ ,  $n_2$  fois la valeur  $x_2$ , ...,  $n_n$  fois la valeur  $x_n$  et la somme de ses valeurs sera  $n_1x_1 + n_2x_2 + \dots + n_nx_n$ . La valeur moyenne de  $n$  valeurs est :

$$m_N = \frac{n_1}{N}x_1 + \frac{n_2}{N}x_2 + \dots + \frac{n_n}{N}x_n$$

Quand  $N$  devient assez grand les fréquences d'apparition  $\frac{n_k}{N}$  des événements  $\omega_k$  approchent les probabilités  $p_k$  de réalisation de  $\omega_k$  et par conséquent :

$$\lim_{N \rightarrow \infty} m_N = p_1x_1 + p_2x_2 + \dots + p_nx_n$$

DÉFINITION A.11.1 La limite de  $m_N$  quand  $N \rightarrow \infty$  s'appelle moyenne ou espérance mathématique de la v.a.  $X$  et sera notée par  $E(X)$ . Nous avons donc :

$$E(X) = \sum_{k=1}^n p_kx_k \tag{A.11.1}$$

Supposons maintenant que  $\Omega$  est un ensemble dénombrable. Dans ce cas la détermination de  $E(X)$  pose deux problèmes :

- (1) Il faut que la série  $m_N = \sum_{k=1}^{\infty} p_k x_k$  soit convergente quand  $N \rightarrow \infty$ .
- (2) Même dans le cas où la série est convergente, on ne peut pas démontrer de façon simple que  $\lim_{N \rightarrow \infty} m_N$  exprime la moyenne des valeurs qu'elle prend la v.a.  $X$ . Il faut pour cela utiliser le théorème de Khintchine que nous verrons plus loin.

Pour l'instant on admet donc que, si  $\Omega$  dénombrable, l'espérance mathématique d'une v.a.  $X$  est donnée par

$$E(X) = \sum_{k=1}^{\infty} p_k x_k \quad (\text{A.11.2})$$

à condition que la série converge.

L'espérance mathématique a les propriétés suivantes :

PROPRIÉTÉ A.11.1 . Si  $\lambda, \mu \in \mathbb{R}$  deux constantes, alors :

$$E(\lambda X + \mu) = \lambda E(X) + \mu \quad (\text{A.11.3})$$

PROPRIÉTÉ A.11.2 (Linéarité de l'opérateur  $E$ ). Soient  $n$  v.a.  $X_1, \dots, X_n$  telles que  $E(X_k)$  existe pour tout  $k = 1, \dots, n$ . Alors  $E(\sum_{k=i}^n X_k)$  existe aussi et nous avons

$$E\left(\sum_{k=i}^n X_k\right) = \sum_{k=i}^n E(X_k) \quad (\text{A.11.4})$$

En plus, en vertu de la propriété précédente, nous avons aussi :

$$\forall \lambda_1, \dots, \lambda_n \in \mathbb{R} : E\left(\sum_{k=i}^n \lambda_k X_k\right) = \sum_{k=i}^n \lambda_k E(X_k) \quad (\text{A.11.5})$$

PROPRIÉTÉ A.11.3 . Étant donné la v.a.  $X$ , nous pouvons définir une v.a., dérivée de  $X$ ,  $X_c = X - E(X)$  qui est la v.a. centrée, car

$$E(X_c) = E(X - E(X)) = 0 \quad (\text{A.11.6})$$

qui exprime le fait que la valeur moyenne des écarts des valeurs de la v.a.  $X$  de sa valeur moyenne est nulle quand le nombre de répétitions devient assez grand.

Cette dernière propriété nous conduit à la notion de la variance. Si une v.a.  $X$  prend comme valeurs  $x_1, x_2, \dots, x_n$  lorsqu'on répète une expérience aléatoire et si  $E(X)$  est la moyenne de la v.a., alors on dit que la v.a. présente une grande dispersion quand elle prend souvent des valeurs qui sont éloignées de sa moyenne. Au contraire quand elle prend souvent des valeurs qui sont proches de la moyenne, on dit qu'elle a une petite dispersion.

DÉFINITION A.11.2 Soit  $X$  v.a. sur  $(\Omega, \mathcal{A})$ . Une mesure de la dispersion de la v.a.  $X$  par rapport à son espérance mathématique est la variance, qui est la moyenne des carrés des écarts entre les différentes valeurs de la v.a. et sa valeur moyenne :

$$V(X) = (x_1 - E(X))^2 p_1 + (x_2 - E(X))^2 p_2 + \dots + (x_n - E(X))^2 p_n \quad (\text{A.11.7})$$

ou, de façon condensée :

$$V(X) = E((X - E(X))^2) \quad (\text{A.11.8})$$

La racine carrée  $\sqrt{V(X)}$  de la variance s'appelle l'écart-type.

La variance a les propriétés suivantes :

PROPRIÉTÉ A.11.4 . La variance s'écrit aussi :

$$V(X) = E(X^2) - E(X)^2 \quad (\text{A.11.9})$$

PROPRIÉTÉ A.11.5 (Invariance par rapport à la translation). Si  $X$  est une v.a. et  $\lambda \in \mathbb{R}$  une constante, nous avons :

$$V(X + \lambda) = V(X)$$

Par conséquent la variance d'une v.a.  $X$  caractérise la loi de  $X$  et non ses valeurs.

PROPRIÉTÉ A.11.6 (Variance d'une somme de v.a.). Considérons les v.a.  $X_1, \dots, X_n$  deux à deux indépendantes. Nous avons la relation :

$$V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k) \quad (\text{A.11.10})$$

Comme conséquence nous avons pour  $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ , constantes :

$$V\left(\sum_{k=1}^n \lambda_k X_k\right) = \sum_{k=1}^n \lambda_k^2 V(X_k) \quad (\text{A.11.11})$$

et si  $X, Y$  v.a. indépendantes, alors :

$$V(X + Y) = V(X) + V(Y) \quad (\text{A.11.12})$$

PROPRIÉTÉ A.11.7 . Pour une constante  $\lambda \in \mathbb{R}$  ;  $\lambda \neq E(X)$ , nous pouvons exprimer la dispersion des valeurs de la v.a.  $X$  autour de  $\lambda$  par la formule :

$$E((X - \lambda)^2) = V(X) + (E(X) - \lambda)^2 \quad (\text{A.11.13})$$

Soit la v.a.  $X$  définie sur  $(\Omega, \mathcal{A})$ . On peut définir une nouvelle v.a. :

$$Z_X = \frac{X - E(X)}{\sqrt{V(X)}} \quad (\text{A.11.14})$$

qui est de moyenne nulle et de variance unité. Cette v.a., qui joue un grand rôle en Statistique, s'appelle v.a. centrée réduite (v.a.c.r.). Elle fait partie d'une catégorie plus générale de v.a., à savoir les v.a. de même type.

DÉFINITION A.11.3 Deux v.a.  $X$  et  $Y$  sur  $(\Omega, \mathcal{A})$  sont de même type s'il existe deux constantes  $\lambda, \mu \in \mathbb{R}$  telles que la v.a.  $\lambda Y + \mu$  admette la même loi de probabilité que  $X$ .

Nous pouvons vérifier que la v.a.c.r.  $Z_X$  est de même type que  $X$ .

N.B.- Dans la suite nous utiliserons la notation  $E(X)$  pour indiquer l'opérateur espérance mathématique appliqué à la v.a.  $X$ . Nous utiliserons la notation  $\mu_X$  (ou  $\mu$ ) pour indiquer la valeur de cette espérance mathématique calculée sur les valeurs prises par la v.a.  $X$ . De même  $V(X)$  indiquera l'opérateur de variance appliqué à la v.a.  $X$ , tandis que la notation  $\sigma_X^2$  (ou  $\sigma^2$ ) indiquera la valeur de la variance pour un ensemble de valeurs de la v.a.  $X$ .

## A.12 Moments

Soit une v.a.  $X$  sur  $(\Omega, \mathcal{A}, P)$  pouvant prendre les valeurs  $x_1, x_2, \dots, x_n$  avec probabilités  $p_1, p_2, \dots, p_n$  respectivement et  $\sum_{i=1}^n p_i = 1$ . On considère, comme précédemment, un système de masses ponctuelles (nuage des points)  $p_1, p_2, \dots, p_n$  placées aux points  $x_1, x_2, \dots, x_n$ . Nous avons vu que  $E((X - \lambda)^2)$ , où  $\lambda \in \mathbb{R}$  constante, est le moment d'inertie  $I_\Delta$  par rapport à un axe  $\Delta$ , qui est distant de  $\lambda$  du centre de gravité  $E(X)$  du nuage des points. Nous pouvons généraliser et avoir

DÉFINITION A.12.1 Soient  $k \in \mathbb{N}$  nombre naturel et une constante  $\lambda \in \mathbb{R}$ . On appelle moment d'ordre  $k$  autour de  $\lambda$  la quantité

$$\mu_k(\lambda) = E((X - \lambda)^k) = \sum_{i=1}^n (x_i - \lambda)^k p_i \quad (\text{A.12.1})$$

si la série converge.

Si  $\lambda = E(X)$ , alors nous avons le moment centré d'ordre  $k$  :

$$\mu_k = E((X - E(X))^k) \quad (\text{A.12.2})$$

Si  $\lambda = 0$  alors nous avons le moment d'ordre  $k$  non centré (initial) :

$$\mu'_k = E(X^k) \quad (\text{A.12.3})$$

On peut aussi avoir les moments absolus :

$$M_k(\lambda) = E(|X - \lambda|^k) \quad (\text{A.12.4})$$

$$M_k = E(|X - E(X)|^k) \quad (\text{A.12.5})$$

$$M'_k = E(|X|^k) \quad (\text{A.12.6})$$

On voit que l'espérance mathématique et la variance sont des moments, car  $E(X) = \mu'_1$  et  $V(X) = \mu_1$ . En dehors de ces moments, on définit aussi, afin de faciliter les calculs, le moment factoriel.

DÉFINITION A.12.2 . On appelle moment factoriel d'ordre  $k$  la quantité

$$\mu_{[k]} = E(X \cdot (X - 1) \cdot \dots \cdot (X - k + 1)) \tag{A.12.7}$$

Nous avons les relations :

$$\begin{aligned} \mu_{[1]} &= \mu'_1 \\ \mu_{[2]} &= \mu'_2 - \mu_1 \\ \mu_{[3]} &= \mu'_3 - 3\mu'_2 + 2\mu_1 \\ \dots & \dots \dots \end{aligned} \tag{A.12.8}$$

### A.13 La notion de la variable aléatoire continue

Les variables aléatoires continues sont utilisées en tant que telles dans les cas des probabilités géométriques. Dans la pratique nous n'avons pas des variables continues car les mesures sont toujours discrètes, à cause de leur précision finie. Néanmoins les v.a. continues sont utiles pour les calculs car elles peuvent être considérées comme des limites des v.a. discrètes.

Nous allons travailler uniquement avec des v.a. continues réelles. Donc  $X$  sera une application dans  $\mathbb{R}$  (ou  $\mathbb{R}^n$ ). Nous avons déjà vu que dans ce cas la tribu utilisée c'est la tribu  $\mathcal{B}_{\mathbb{R}}$  des boréliens, c'est-à-dire la plus petite tribu qui contient tous les points et aussi tous les intervalles de  $\mathbb{R}$ . Ce dernier point est important car s'agissant d'une v.a.  $X$  continue on ne s'intéresse pas à calculer la probabilité pour que  $X$  prenne une valeur donnée  $x \in \mathbb{R}$ , mais plutôt à connaître la probabilité que  $X$  se trouve dans un intervalle  $[a, b] \in \mathbb{R}$  donné.

DÉFINITION A.13.1 On appelle variable aléatoire réelle, continue sur l'espace probabilisé  $(\Omega, \mathcal{A}, P)$  toute application mesurable  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ .

Du fait que  $X$  est mesurable nous avons pour tout intervalle  $B \subset \mathbb{R} (\Rightarrow B \in \mathcal{B}_{\mathbb{R}})$  :

$X^{-1}(B) \in \mathcal{A}$ . Par conséquent  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  est dotée de la probabilité  $P_X$  induite par la v.a.  $X$ . Nous avons ainsi

DÉFINITION A.13.2 On appelle loi de probabilité de la v.a. continue  $X$  l'application  $P_X : \mathcal{B}_{\mathbb{R}} \rightarrow [0, 1]$ , définie par

$$\forall B \in \mathcal{B}_{\mathbb{R}} : P_X(B) = P(X^{-1}(B)) \tag{A.13.1}$$

Notons que  $P(X^{-1}(B))$  existe car on a  $X^{-1}(B) \in \mathcal{A}$ .

REMARQUE.- Nous avons  $X^{-1}(B) = \{\omega \in \Omega / X(\omega) \in (-\infty, b]\}$  si  $B = (-\infty, b]$ . Donc en toute rigueur, on doit noter  $P(\{\omega \in \Omega / X(\omega) \in (-\infty, b]\})$  ce qui est lourd. Pour alléger la notation, on écrit  $X^{-1}(B) = \{X \leq b\}$  et, par conséquent :  $P(X^{-1}(B)) = P(\{X \leq b\})$ .

## A.14 Fonction de répartition

Comme on vient de l'indiquer dans le cas où la v.a.  $X$  est continue, ce qui nous intéresse est la probabilité que la valeur de  $X$  est dans un intervalle  $]-\infty, b[ \in \mathcal{B}_{\mathbb{R}}$ .

**DÉFINITION A.14.1** On appelle fonction de répartition de la v.a.  $X$  la fonction  $F_X : \mathbb{R} \rightarrow \mathbb{R}$  définie par  $\mathbb{R} \ni b \mapsto F_X(b) = P_X(]-\infty, b]) = P(\{X \leq b\})$

Nous pouvons démontrer le

**THÉORÈME A.14.1** Soient  $(\Omega, \mathcal{A}, P)$  un espace probabilisé et  $X$  v.a. sur  $(\Omega, \mathcal{A})$ . La fonction  $P_X : \mathcal{B}_{\mathbb{R}} \rightarrow [0, 1]$  définie par  $P_X(B) = P(X^{-1}(B))$  est une mesure de probabilité sur  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ .

Nous avons les propriétés suivantes pour la fonction de répartition :

**PROPRIÉTÉ A.14.1**  $F_X(-\infty) = 0$  ;  $F_X(+\infty) = 1$ .

**PROPRIÉTÉ A.14.2** Probabilité d'un intervalle :

$$P(a \leq X \leq b) = F_X(b) - F_X(a)$$

$$P(X \geq b) = 1 - F_X(b)$$

**PROPRIÉTÉ A.14.3**  $F_X$  est une fonction non décroissante, i.e.

$$\forall a, b \in \mathbb{R} \text{ t.q. } a \leq b : F_X(a) \leq F_X(b)$$

En vertu des propriétés 5.2.1 et 5.2.3 nous avons que  $F_X(x) \in [0, 1] \forall x \in \mathbb{R}$ .

**PROPRIÉTÉ A.14.4**  $F_X$  est continue à gauche, i.e. si  $(x_n)_{n=1, \dots}$  une suite de points  $x_n \in \mathbb{R}$  avec  $\lim_{n \rightarrow +\infty} x_n = x$  et  $x_n < x \forall n = 1, \dots$  alors  $\lim_{n \rightarrow +\infty} F_X(x_n) = F_X(x)$ .

**PROPRIÉTÉ A.14.5** Pour tout  $a \in \mathbb{R}$  tel que  $F_X$  est continue, nous avons  $P(X = a) = 0$ . Si, par contre,  $F_X$  est discontinue pour  $a \in \mathbb{R}$  alors  $P(X = a) > 0$ . De plus le nombre de points  $a \in \mathbb{R}$  pour lesquels  $F_X$  est discontinue est au plus infini dénombrable.

**PROPRIÉTÉ A.14.6** Calcul des probabilités induites à l'aide de la fonction de répartition :

$$P_X(a \leq X < b) = F_X(b) - F_X(a)$$

$$P_X(a \leq X \leq b) = F_X(b) - F_X(a) + P_X(X = b)$$

$$P_X(a < X < b) = F_X(b) - F_X(a) - P_X(X = a)$$

$$P_X(a < X \leq b) = F_X(b) - F_X(a) - P_X(X = a) + P_X(X = b) - P_X(a \leq X < +\infty) = 1 - F_X(a)$$

$$P_X(a < X < +\infty) = 1 - F_X(a) - P_X(X = a)$$

$$P_X(-\infty < X < b) = F_X(b)$$

$$P_X(-\infty < X \leq b) = F_X(b) + P_X(X = b)$$

Cette propriété montre que la fonction de répartition  $F_X$  caractérise complètement la loi  $P_X$  de la v.a.  $X$ .

PROPRIÉTÉ A.14.7 Si  $(x_n)_{n=1, \dots}$  une suite de points  $x_n \in \mathbb{R}$  avec  $\lim_{n \rightarrow +\infty} x_n = x$  et  $x_n > x \forall n = 1, \dots$  alors  $\lim_{n \rightarrow +\infty} F_X(x_n) = F_X(x) + P_X(X = x)$ .

Parfois il est utile de pouvoir tester si une application  $F_X : \mathbb{R} \rightarrow \mathbb{R}$  est une fonction de répartition. Le théorème suivant précise les conditions.

THÉORÈME A.14.2 Soit la fonction  $F_X : \mathbb{R} \rightarrow \mathbb{R}$  telle que :

- (1)  $F$  n'est pas décroissante,
- (2)  $F$  est continue à gauche,
- (3)  $\lim_{x \rightarrow -\infty} F(x) = 0$  et  $\lim_{x \rightarrow +\infty} F(x) = 1$ .

Alors  $F$  peut être considérée comme une fonction de répartition d'une v.a.  $X$  dont la loi de probabilité  $P_X$  sur  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$  est définie par la relation  $P_X(]-\infty, x]) = F(x)$ .

EXERCICE A.1 Soit la fonction  $F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$ . On remarque que  $F$  est non décroissante, continue à gauche avec  $\lim_{x \rightarrow -\infty} F(x) = 0$  et  $\lim_{x \rightarrow +\infty} F(x) = 1$ . Donc on peut la considérer comme une fonction de répartition de la v.a.  $X$  dont sa loi est donnée par

$$P_X(]-\infty, x]) = F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } 0 \leq x \leq 1 \\ 1 & \text{si } x > 1 \end{cases}$$

### A.15 Fonctions de densité

Ayant défini la v.a. continue, il est normal d'essayer de calculer ses grandeurs statistiques importantes, à savoir l'espérance mathématique et la variance.

Considérons, donc, un espace probabilisé  $(\Omega, \mathcal{A}, P)$  et soit  $X$  une v.a. continue de  $(\Omega, \mathcal{A})$  dans  $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ . L'espérance mathématique de  $X$  s'exprime, en s'inspirant de la formule correspondante du cas discret, par l'intégrale

$$E(X) = \int_{\Omega} X(\omega) dP_X(B_{\omega}) = \int_{\Omega} X(\omega) dP(\omega) \tag{A.15.1}$$

où on a noté  $B_{\omega} = \{B \in \mathcal{B}_{\mathbb{R}} \mid X^{-1}(B) = \omega\}$ . Cette intégrale pose un problème car c'est une intégrale au sens de Lebesgue. Or en règle générale on ne peut faire des calculs qu'avec des intégrales au sens de Riemann. Pour contourner cette difficulté nous allons renoncer à utiliser la mesure de probabilité  $P(\omega)$  et nous allons utiliser, à sa place, la fonction de répartition qui, comme nous l'avons vu, caractérise complètement la loi  $P_X$ . Rappelons que  $F_X$  est une fonction de  $\mathbb{R}$  dans  $\mathbb{R}$ . Nous pouvons donc définir l'intégrale au sens de Stieltjes :

$\int_{\mathbb{R}} x dfn(x)$ . Ce n'est pas encore une intégrale au sens de Riemann mais on verra qu'avec une contorsion supplémentaire on peut y arriver. Pour le moment essayons plutôt de voir où nous en sommes. Le théorème suivant le précise.

**THÉORÈME A.15.1** Soit  $F_X$  fonction de répartition d'une v.a.  $X$  sur  $(\Omega, \mathcal{A})$ . Alors si l'espérance mathématique de  $X$  existe, elle est égale numériquement à l'intégrale de Stieltjes

$$E(X) = \int_{\mathbb{R}} x dfn_X(x) \quad (\text{A.15.2})$$

Donc, bien que nous avons changé de mode de calcul (et, accessoirement, de manière de voir le monde), l'espérance mathématique est toujours là ! Il reste maintenant à passer à une intégrale de Riemann. Nous pouvons le faire, pratiquement sans bouger, si on s'assure que la fonction de répartition  $F_X$  admet une dérivée continue. Dans ce cas l'intégrale de Stieltjes peut se transformer à une intégrale de Riemann et on aura pour l'espérance mathématique :

$$E(X) = \int_{\mathbb{R}} x F'_X(x) dx \quad (\text{A.15.3})$$

où  $F'_X$  est la dérivée de  $F_X$ .

Dans la suite on notera par  $f_X$  la dérivée  $F'_X$  de  $F$  et on l'appellera fonction de densité. Plus précisément :

**DÉFINITION A.15.1** Une fonction  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  est appelée fonction de densité si elle satisfait aux conditions suivantes :

- (1)  $f_X(x) \geq 0 \forall x \in \mathbb{R}$ ,
- (2)  $f_X$  admet au plus un nombre fini de discontinuités dans chaque intervalle fini de  $\mathbb{R}$ ,
- (3)  $\int_{\mathbb{R}} f_X(x) dx = 1$ .

La relation entre fonction de répartition et fonction de densité est évidemment :

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi \quad (\text{A.15.4})$$

**EXERCICE A.2** Soit la fonction de répartition  $F_X(x) = \begin{cases} 1 - e^{-kx} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$ . La fonction

de densité est  $f_X(x) = \begin{cases} ke^{-kx} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$ .

## A.16 Espérance mathématique - Variance

Nous sommes maintenant en mesure de donner les définitions concernant l'espérance mathématique et la variance d'une v.a. continue  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ .

DÉFINITION A.16.1 L'espérance mathématique de la v.a. continue  $X$  sur  $(\Omega, \mathcal{A})$  est donnée par

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx \quad (\text{A.16.1})$$

où  $f_x$  fonction de densité de la v.a.  $X$ .

L'intégration étant un opérateur linéaire, les propriétés de l'espérance mathématique qu'on avait déjà examiné au chapitre précédent, sont valables ici aussi. En particulier si  $h(X)$  est une fonction de la v.a.  $X$ , alors  $h(X)$  est aussi une v.a. avec espérance mathématique

$$E[h(X)] = \int_{\mathbb{R}} h(x) f_X(x) dx \quad (\text{A.16.2})$$

DÉFINITION A.16.2 La variance de la v.a. continue  $X$  est donnée par

$$V(X) = \int_{\mathbb{R}} (x - E(X))^2 f_X(x) dx \quad (\text{A.16.3})$$

De même si  $h(X)$  fonction de  $X$ , alors sa variance est

$$V[h(X)] = \int_{\mathbb{R}} [h(x) - E[h(X)]]^2 f_X(x) dx \quad (\text{A.16.4})$$

La variance par rapport à une constante  $\lambda \in \mathbb{R}$  est donnée par la formule

$$\int_{\mathbb{R}} [x - \lambda]^2 f_X(x) dx = V(X) + (E(X) - \lambda)^2 \quad (\text{A.16.5})$$

qui est l'équivalent de la formule (4.13) pour le cas discret.

## A.17 Moments - Médiane - Quantiles

Nous pouvons, pour les moments, que nous avons déjà examiné au chapitre précédent, établir facilement leur équivalent en v.a. continues. Nous avons donc pour les moments d'ordre  $k$ .

– moment d'ordre  $k$  autour de  $\lambda \in \mathbb{R}$  :

$$\mu_k(\lambda) = \int_{\mathbb{R}} [x - \lambda]^k df_{n_X}(x) = \int_{\mathbb{R}} [x - \lambda]^k f_X(x) dx \quad (\text{A.17.1})$$

– moment centré d'ordre  $k$  :

$$\mu_k = \int_{\mathbb{R}} [x - E(X)]^k df_{n_X}(x) = \int_{\mathbb{R}} [x - E(X)]^k f_X(x) dx \quad (\text{A.17.2})$$

– moment d'ordre  $k$  non centré (initial) :

$$\mu'_k = \int_{\mathbb{R}} x^k df_{n_X}(x) = \int_{\mathbb{R}} x^k f_X(x) dx \quad (\text{A.17.3})$$

Pour des études statistiques on utilise aussi les trois notions suivantes :

DÉFINITION A.17.1 On appelle médiane d'une v.a.  $X$  tout nombre  $m \in \mathbb{R}$  tel que

$$P_X(X \leq m) \geq 0.5 \quad \text{et} \quad P_X(X \geq m) \geq 0.5.$$

La médiane est donc un nombre  $m \in \mathbb{R}$  qui a autant de chances d'être dépassé par une valeur de la v.a.  $X$ , que de ne pas être dépassé.

Nous avons, concernant la médiane, les propriétés suivantes :

PROPRIÉTÉ A.17.1  $F_X(m) = 0.5$

PROPRIÉTÉ A.17.2  $\int_{-\infty}^m f_X(x) dx = 1 - \int_{-\infty}^m f_X(x) dx = 0.5$

PROPRIÉTÉ A.17.3  $\sigma(X) \geq |E(X) - m|$

Étant donnée une v.a.  $X$  et sa loi de probabilité  $P_X$ , la valeur médiane existe toujours et parfois il y en a plusieurs. En effet considérons le jeu « pile ou face » et associons à ce jeu la v.a.  $X$  de sorte que  $X$  prend la valeur 0 si le résultat est « pile » et la valeur 1 si le résultat est « face ». Alors pour tout nombre  $m \in [0, 1]$  nous avons  $P_X(X \leq m) \geq 0.5$  et  $P_X(X \geq m) \geq 0.5$ . Donc tout nombre dans l'intervalle  $[0, 1]$  convient pour médiane de la v.a.  $X$ .

DÉFINITION A.17.2 Soit une v.a.  $X$  avec fonction de répartition  $F_X$ . On appelle  $p$ -ième quantile de  $X$ , avec  $p \in ]0, 1[$ , un nombre noté  $x_p$  et tel que  $P_X(X \leq x_p) \geq p$  et  $P_X(X \geq x_p) \geq 1 - p$ .

Cette définition conduit à la définition équivalente :

DÉFINITION A.17.3 On appelle quantiles d'ordre  $k$ , les  $k - 1$  nombres  $q_1, q_2, \dots, q_{k-1}$  tels que :

$$P_X(X < q_l) \geq \frac{l}{k} \quad \text{et} \quad P_X(X \geq q_l) \geq \frac{k-l}{k} ; \quad l = 1, 2, \dots, k-1 \quad (\text{A.17.4})$$

La relation des quantiles avec la fonction de répartition est donnée par la propriété suivante :

PROPRIÉTÉ A.17.4 . Si la fonction de répartition  $F_X$  est continue et strictement croissante, alors

$$F_X(q_l) = \frac{l}{k} ; \quad l = 1, 2, \dots, k-1 \quad (\text{A.17.5})$$

D'habitude, parmi les quantiles, on utilise

- les quartiles, qu'on obtient des quantiles pour  $k = 4$ , i.e.  $P_X(X < q_l) \geq \frac{l}{4}$  et  $P_X(X \geq q_l) \geq \frac{4-l}{4}$  ;  $l = 1, 2, 3$  ;
- les centiles, qu'on obtient des quantiles pour  $k = 100$ , i.e.  $P_X(X < q_l) \geq \frac{l}{100}$  et  $P_X(X \geq q_l) \geq \frac{100-l}{100}$  ;  $l = 1, 2, \dots, 99$ .

DÉFINITION A.17.4 On appelle mode d'une v.a.  $X$  avec fonction de densité  $f_X$  la valeur  $x_M \in \mathbb{R}$  pour laquelle  $f_X$  est maximale, c'est-à-dire :

$$x_M = \underset{x \in \mathbb{R}}{\text{arg max}} f_X(x) \tag{A.17.6}$$

Pour une v.a. discrète  $X$  qui prend les valeurs  $x_1, x_2, \dots$  avec probabilités  $p_1, p_2, \dots$  respectivement, le mode est défini de façon analogue :

$$x_M = \underset{k=1,2,\dots}{\text{arg max}} P_X(X = x_k) = \underset{k=1,2,\dots}{\text{arg max}} p_k \tag{A.17.7}$$

REMARQUES.- (i) Le mode n'est pas forcément unique. Dans ce cas la fonction de densité de  $X$  est dite *multimodale*.

(ii) Si la fonction de répartition admet un seul axe de symétrie et cet axe passe par son maximum, alors ce maximum représente à la fois l'espérance mathématique, la médiane et le mode.

### A.18 Fonctions des variables aléatoires

Considérons une v.a. continue  $X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ . La propriété suivante établit les conditions pour qu'une fonction de  $X$  soit aussi une v.a.

PROPRIÉTÉ A.18.1 Soient  $F_X$  la fonction de répartition et  $f_X$  la fonction de densité de la v.a. continue  $X$ . Considérons l'application mesurable  $h : \mathbb{R} \rightarrow \mathbb{R}$ . Notons  $Y$  la fonction  $h(X)$ . Alors  $Y$  est une v.a. si  $h$  est inversible et continûment différentiable. Dans ce cas la fonction de répartition de  $Y$  est

$$F_Y(y) = P_X(X < h^{-1}(y)) = F_X(h^{-1}(y))$$

Pour la fonction de densité nous avons bien sûr  $f_Y(y) = \frac{d}{dy} F_Y(y)$ , mais aussi

$$f_Y(y) = f_X(h^{-1}(y)) \cdot |(h^{-1}(y))'|$$

### A.19 Extension au cas des v.a. discrètes

Nous allons, avant de terminer ce paragraphe, définir les notions des fonctions de répartition et de densité pour des v.a. discrètes.

Considérons une v.a. discrète  $X$  sur  $(\Omega, \mathcal{A}, P)$  qui prend les valeurs  $x_1, x_2, \dots$  avec probabilités  $p_1, p_2, \dots$ . On note par  $E$  l'ensemble des valeurs  $x_k$ .

DÉFINITION A.19.1 La fonction de répartition  $F_X$  de la v.a. discrète  $X$  est une application de  $E$  définie pour tout  $x \in E$  par  $F_X(x) = P_X(X < x)$ , avec  $P_X(X < x) = P(X^{-1}(\{x_k \in E \mid x_k < x\})) = P(\{\omega_k \in \Omega \mid X(\omega_k) = x_k < x\})$

La fonction de répartition pour une v.a. discrète est une fonction en escalier et elle présente des discontinuités aux points  $x_1, x_2, \dots$ .

DÉFINITION A.19.2 La fonction de densité de la v.a. discrète  $X$  est une application de  $E$  définie par  $f_X(x_k) = P_X(X = x_k)$  pour tout  $x_k \in E$

Pour la fonction de densité nous avons

$$\sum_{x_k \in E} f_X(x_k) = 1$$

La relation entre fonction de répartition et fonction de densité est donnée par la

PROPRIÉTÉ A.19.1 Si  $F_X$  et  $f_X$  sont fonctions de répartition et de densité respectivement, de la v.a.  $X$ , alors

$$F(x) = P_X(X < x) = \sum_{\substack{x_k \in E \\ x_k < x}} f_X(x_k)$$

Si  $h$  est une fonction mesurable de la v.a.  $X$  alors  $Y = h(X)$  est aussi une v.a. avec fonction de densité

$$f_Y(y) = \sum_{\substack{x_k \in E \\ h(x_k) = y}} f_X(x_k)$$

Dans le cas particulier où  $h$  est bijective, nous avons :  $f_Y(y) = f_X(h^{-1}(y))$ .

## VECTEURS ALÉATOIRES

L'étude d'une expérience aléatoire à l'aide d'une variable aléatoire n'est pas toujours suffisante. Lorsque e.g. nous voulons examiner le taux de change franc/marc il est manifeste qu'une seule variable ne suffit pas pour expliquer son évolution. Nous devons prendre en considération le différentiel de l'inflation de deux pays, le différentiel des taux d'intérêt, les résultats du commerce extérieur, etc. Nous pouvons donc envisager pour étudier une expérience aléatoire, l'utilisation simultanée des plusieurs variables aléatoires.

### A.20 Variables aléatoires à plusieurs dimensions

Soit une expérience aléatoire dont le résultat est composé d'un système de  $q$  valeurs  $x_1, x_2, \dots, x_q$  et où  $x_1$  est une valeur de la v.a.  $X_1$ ,  $x_2$  est une valeur de la v.a.  $X_2$ , ...,  $x_q$  est une valeur de la v.a.  $X_q$ . Afin de faciliter les calculs, nous pouvons considérer les valeurs  $x_1, x_2, \dots, x_q$ , qui apparaissent à chaque répétition de l'expérience aléatoire, comme les coordonnées d'un vecteur à  $q$  dimensions que l'on notera  $\mathbf{x}$ . Sous ces conditions la famille des v.a.  $X_1, X_2, \dots, X_q$  - qui, rappelons-le, sont des fonctions - peut être vue comme un vecteur-colonne de  $q$  lignes, appelé vecteur aléatoire et noté  $\mathbf{X}$ . Plus précisément :

DÉFINITION A.20.1 Soient  $(\Omega, \mathcal{A}, P)$  espace probabilisé et  $(\mathbb{R}^q, \mathcal{B}_q)$  espace probabilisable. On appelle vecteur aléatoire réel à  $q$  dimensions toute application mesurable  $\mathbf{X}$  de  $(\Omega, \mathcal{A})$  dans  $(\mathbb{R}^q, \mathcal{B}_q)$ .

Du fait que l'application  $\mathbf{X}$  est mesurable, nous avons

$$\forall \mathbf{x} \in \mathbb{R}^q : \{ \omega \in \Omega \mid \mathbf{X}(\omega) \leq \mathbf{x} \} \in \mathcal{A}$$

Nous pouvons, comme dans le cas de la v.a., définir la loi de probabilité  $P_{\mathbf{X}}$  du vecteur aléatoire  $\mathbf{X}$  à valeurs dans  $\mathbb{R}^q$ , comme étant la mesure image de  $P$  par l'application  $\mathbf{X}$ .

DÉFINITION A.20.2 On appelle loi de probabilité conjointe  $P_{\mathbf{X}}$  la probabilité induite par le vecteur aléatoire  $\mathbf{X}$  :

$$\forall B \in \mathcal{B}_q : P_{\mathbf{X}}(B) = P[\mathbf{X}^{-1}(B)]$$

Nous avons aussi la définition suivante :

DÉFINITION A.20.3 On appelle probabilité marginale  $P_{X_i}$  en  $X_i$  du vecteur aléatoire  $\mathbf{X} = [X_1, X_2, \dots, X_q]^T$  la loi de probabilité de la v.a.  $X_i$  :

$$\forall B \in \mathcal{B}_q : P_{X_i}(B) = P_{\mathbf{X}}(X_1 \in \mathbb{R}, \dots, X_i \in B, \dots, X_q \in \mathbb{R}) ; i = 1, \dots, q$$

Si les v.a.  $X_1, X_2, \dots, X_q$  sont continues et la probabilité que le point  $\mathbf{x} \in \mathbb{R}^q$  se trouve dans l'hyper-réctangle  $\{[x_1, x_1 + dx_1[ , \dots, [x_q, x_q + dx_q[ \}$  est de la forme  $f_{\mathbf{X}}(x_1, \dots, x_q) dx_1 \cdots dx_q$ , alors  $f_{\mathbf{X}}$  est la densité de probabilité du vecteur aléatoire  $\mathbf{X} = [X_1, X_2, \dots, X_q]^T$ . Nous avons évidemment

$$\int \cdots \int_{\mathbb{R}^q} f_{\mathbf{X}}(x_1, \dots, x_q) dx_1 \cdots dx_q = 1$$

L'extension au cas discret est immédiate. Si les v.a.  $X_1, X_2, \dots, X_q$  sont discrètes, alors la probabilité que  $\mathbf{X} = \mathbf{x} \in \mathbb{R}^q$  est donnée par  $P_{\mathbf{X}}(\mathbf{X} = \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{X}}(x_1, \dots, x_q)$ , avec  $\sum_{\mathbf{x}} f_{\mathbf{X}}(\mathbf{x}) = 1$ . La somme ici est considérée par rapport à tous les points  $\mathbf{x} \in \mathbb{R}^q$  pour lesquels  $\mathbf{X}^{-1}(\mathbf{x}) \in \mathcal{A}$ .

Pour la fonction de répartition d'un vecteur aléatoire, nous avons la relation

$$F_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} \leq \mathbf{x}) = \int \cdots \int_{-\infty}^{\mathbf{x}} f_{\mathbf{X}}(\xi_1, \dots, \xi_q) d\xi_1 \cdots d\xi_q ; \mathbf{x} \in \mathbb{R}^q$$

Dans le cas de la probabilité marginale, nous pouvons définir la fonction de répartition marginale en  $X_i$  par la relation

$$F_{X_i}(x_i) = P_{X_i}(X_i < x_i) = F_{\mathbf{X}}(+\infty, \dots, +\infty, x_i, +\infty, \dots, +\infty)$$

La fonction de densité marginale en  $X_i$  est donnée par la relation :

$$f_{X_i}(x_i) = \int \cdots \int_{\mathbb{R}^{q-1}} f_{\mathbf{X}}(x_1, \dots, x_q) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_q$$

L'espérance mathématique du vecteur aléatoire  $\mathbf{X}$  est le vecteur

$$E(\mathbf{X}) = [E(X_1), \dots, E(X_q)]$$

Pour une v.a.  $X$  la variance est.  $V(X) = E(X - E(X))^2$ . Si on passe à un vecteur aléatoire nous obtenons, à cause du fait que la variance est le carré de l'espérance mathématique du scalaire  $[X - E(X)]$ , la matrice  $E[(\mathbf{X} - E(\mathbf{X})) \cdot (\mathbf{X} - E(\mathbf{X}))^\top]$ , de dimension  $(q \times q)$ , qui peut être vu comme le carré de l'espérance mathématique du vecteur  $[\mathbf{X} - E(\mathbf{X})]$ .

DÉFINITION A.20.4 On appelle matrice de variance-covariance du vecteur aléatoire  $\mathbf{X}$ , la matrice

$$C(\mathbf{X}) = E[(\mathbf{X} - E(\mathbf{X})) \cdot (\mathbf{X} - E(\mathbf{X}))^\top]$$

de terme général

$$C_{ij} = C(X_i, X_j) = E[(X_i - E(X_i)) \cdot (X_j - E(X_j))] ; i, j = 1, \dots, q$$

La quantité  $C_{ij}$  s'appelle la covariance des v.a.  $X_i$  et  $X_j$ .

Remarquons que

- la matrice  $C(\mathbf{X})$  est symétrique
- $C_{ii} = C(X_i, X_i) = E[(X_i - E(X_i)) \cdot (X_i - E(X_i))] = E(X_i - E(X_i))^2 = V(X_i)$  ;  $i = 1, \dots, q$  c'est-à-dire que les termes diagonaux de la matrice de variance-covariance sont les variances  $V(X_i)$  des coordonnées  $X_i$  du vecteur aléatoire  $\mathbf{X}$ .

Il est possible aussi, comme dans le cas des v.a., de définir une transformation du vecteur aléatoire  $\mathbf{X}$ . Nous allons nous limiter aux transformations linéaires, c'est-à-dire aux transformations qui peuvent s'exprimer à l'aide d'une matrice. Nous avons ainsi que si  $\mathbf{X}$  est un vecteur aléatoire composé de  $q$  v.a.  $X_1, X_2, \dots, X_q$ , alors  $\mathbf{Y} = A \cdot \mathbf{X}$  est aussi un vecteur aléatoire, avec  $A$  matrice carrée  $(q \times q)$  de terme général  $(a_{ij})$ . L'espérance mathématique de  $\mathbf{Y}$  est :

$$E(\mathbf{Y}) = A \cdot E(\mathbf{X})$$

La matrice de variance-covariance  $C(\mathbf{Y})$  a comme terme général :

$$\begin{aligned} C(Y_i, Y_j) &= E[(Y_i - E(Y_i)) \cdot (Y_j - E(Y_j))] \\ &= \sum_{k,l} a_{ik} a_{jl} E[(X_k - E(X_k)) \cdot (X_l - E(X_l))] \\ &= \sum_{k,l} a_{ik} a_{jl} C(X_k, X_l) \end{aligned}$$

ou, sous forme matricielle :

$$C(\mathbf{Y}) = A \cdot C(\mathbf{X}) \cdot A^\top$$

### A.21 Couples de variables aléatoires

Soient deux v.a.  $X$  et  $Y$  définies sur deux espaces probabilisés (pas nécessairement différents)  $(\Omega, \mathcal{A})$  et  $(\Omega', \mathcal{B})$ . Supposons que  $X$  prend les valeurs  $x_1, x_2, \dots$  avec probabilités  $p_1, p_2, \dots$  respectivement et  $\sum_i p_i = 1$  et  $Y$  prend les valeurs  $y_1, y_2, \dots$  avec probabilités  $q_1, q_2, \dots$  respectivement et  $\sum_i q_i = 1$ . En utilisant ces deux v.a. on peut former un couple aléatoire  $Z = (X, Y)$  qui est aussi une v.a. qui prend les valeurs  $(x_i, y_j); i, j = 1, 2, \dots$  (On peut aussi, à propos de  $Z$ , parler de vecteur aléatoire, mais cette terminologie et son extension à plus de deux dimensions sera étudiée à un chapitre ultérieur.) La probabilité associée à la valeur  $(x_i, y_j)$  sera notée  $p_{ij}$  et sa valeur est :

$$p_{ij} = P \left[ (X = x_i) \cap (Y = y_j) \right] \tag{A.21.1}$$

Cette probabilité s'appelle *loi conjointe* du couple de v.a.  $(X, Y)$ . Nous pouvons voir que  $p_{ij}$  constitue bien une mesure de probabilité, au sens du chapitre 2, car  $p_{ij} \geq 0; i, j = 1, 2, \dots$  et  $\sum_i \sum_j p_{ij} = 1$ .

Remarquons que si  $X$  et  $Y$  sont indépendantes, alors  $p_{ij} = p_i \cdot p_j$ .

#### A.21.1 Lois marginales

DÉFINITION A.21.1 . Étant donné un couple de v.a.  $(X, Y)$ , nous appelons probabilité marginale relative à une v.a. du couple la loi de probabilité de cette v.a., i.e.

- probabilité marginale par rapport à  $X : p_{i.} = \sum_j p_{ij}$
- probabilité marginale par rapport à  $Y : p_{.j} = \sum_i p_{ij}$

En statistique et en analyse de données on construit à partir d'un couple aléatoire le tableau suivant, qu'on appelle *tableau de contingence* :

	Y	$y_1$	...	$y_j$	...	
X		$q_1$	...	$q_j$	...	
$x_1$	$p_1$	$p_{11}$	...	$p_{1j}$	...	$p_{1.}$
.	.	...	...	...	...	.
$x_i$	$p_i$	$p_{i1}$	...	$p_{ij}$	...	$p_{i.}$
.	.	...	...	...	...	.
		$p_{.1}$	...	$p_{.j}$	...	$\sum_i \sum_j p_{ij} = 1$

Le plus souvent à la place des probabilités  $p_{ij}$  que le couple  $(X, Y)$  prenne les valeurs  $(x_i, y_j)$ , qui sont inconnues, on utilise les fréquences d'apparition des valeurs  $(x_i, y_j)$  dans une série de  $N$  répétitions de l'expérience aléatoire.

### A.21.2 Lois conditionnelles

Étant donné un couple de v.a.  $(X, Y)$  et la loi conjointe  $p_{ij}$ , on appelle *loi conditionnelle de X sachant que Y = y<sub>j</sub>* la suite  $(p_{i|j})_{i=1, \dots}$  des probabilités que la v.a.  $X$  prenne la valeur  $x_i$  sachant que la v.a.  $Y$  a pris la valeur  $y_j$ .

Par application du théorème des probabilités composées nous avons :

$$p_{i|j} = \frac{p_{ij}}{p_{.j}}; i, j = 1, 2, \dots \quad (\text{A.21.2})$$

### A.21.3 Moments - Covariance

Pour le couple de v.a.  $(X, Y)$  nous pouvons définir les moments d'ordre  $(k, l)$  autour de l'origine, selon la formule :

$$\mu'_{kl} = E(X^k Y^l) \quad (\text{A.21.3})$$

Si l'un de nombres  $k, l$  est nul, le moment  $\mu'_{kl}$  se réduit au moment d'une seule v.a., i.e. nous obtenons le moment de la loi marginale.

Pour les moments centrés nous avons :

$$\mu_{kl} = E[(X - E(X))^k (Y - E(Y))^l] \quad (\text{A.21.4})$$

Le moment centré

$$\mu_{11} = \text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] \quad (\text{A.21.5})$$

s'appelle *covariance*. Nous avons la relation :

$$\text{cov}(X, Y) = E(XY) - E(X) \cdot E(Y) \quad (\text{A.21.6})$$

Il s'ensuit que si les v.a. sont indépendantes ( $\Rightarrow E(XY) = E(X) \cdot E(Y)$ ), alors  $\text{cov}(X, Y) = 0$ , mais la réciproque n'est pas, en général, vraie.

### A.21.4 Coefficient de corrélation

DÉFINITION A.21.2 . Soit un couple de v.a.  $(X, Y)$  La quantité  $\sigma_{X_1}^2$

$$\rho_{XY} = \frac{E[(X - E(X))(Y - E(Y))]}{\sigma(X) \cdot \sigma(Y)} \quad (\text{A.21.7})$$

s'appelle *coefficient de corrélation*.

S'il n'y a pas d'ambiguïté, on note simplement par  $\rho$  le coefficient de corrélation. Nous avons aussi la relation

$$\rho = \frac{\text{cov}(X, Y)}{\sigma(X) \cdot \sigma(Y)} \quad (\text{A.21.8})$$

Si  $X, Y$  sont indépendantes, alors  $\rho = 0$  et on appelle les v.a.  $X, Y$  non corrélées. Le contraire, en général, n'est pas vrai, i.e.  $\rho = 0$  n'implique pas que  $X, Y$  soient indépendantes.



## Bibliographie

- [1] M. Abramowitz, I. A. Stegun : *Handbook of Mathematical Functions*, Dover, 1965
- [2] T. W. Anderson : *An Introduction to Multivariate Statistical Analysis*, Wiley, 1958
- [3] T. A. Atchison, H. F. Martz (ed.) : *Proceedings of the Symposium on Empirical Bayes Estimation and Computing in Statistics*, Texas Tech.Un., 1969
- [4] C. Baskiotis, J. Raymond, A. Rault : Parameter identification and discriminant analysis for jet engine mechanical state diagnosis, IEEE Conf. in Dec & Control, 1979, p.648
- [5] D. A. Belsley, E. Kuh, R. E. Welsch : *Regression Diagnostics*, Wiley, 1980
- [6] J.-P. Benzécri et col. : *L'analyse des données*, vol 1 et 2, Dunod 1973
- [7] J.-P. Benzécri : Analyse discriminante et analyse factorielle, Cah. Ana. Don., vol II, no 4,1977, p.369
- [8] J.-P. et F. Benzécri : *Pratique de l'analyse des données*, Dunod, 1984
- [9] J. O. Berger, R. L. Wolpert : *The Likelihood Principle*, Inst.Math.Stat., 1984
- [10] F. Cailiez, J.P. Pages : *Introduction à l'analyse des données*, 1976
- [11] CEA : *Statistique appliquée à l'exploitation des mesures*, Tomes I et II, Masson,1978
- [12] L. L. Chao : *Statistics : Methods and Analyses*, McGraw Hill, 1969
- [13] S. Chatterjee, B. Price : *Regression Analysis by Example*, Wiley, 1977
- [14] H.T. Clifford, W. Stephenson : *An Introduction to Numerical Classification*, Academic Press, 1975
- [15] C. H. Coombs, R. C. Kao : *Nometric factor Analysis*, Michigan Un., 1955
- [16] J. Coursol : *Technique statistique des modèles linéaires*, CIMPA, 1980
- [17] H. Cramér : *Mathematical Methods of Statistics*, Princeton Un. Press, 1946
- [18] P. Dagnelie : *Analyse statistique à plusieurs variables*, Vander-Oyez, 1975
- [19] E. Diday et col. : *Optimisation en classification automatique*, vol. 1 et 2, INRIA,1979
- [20] N. R. Draper, N. Smith : *Applied Regression Analysis*, Wiley, 1981
- [21] M. Duflo, D. Florens-Zmirou : *Décisions statistiques pas à pas*, CIMPA, 1981
- [22] D. Dugué : *Traité de statistique théorique et appliquée*, Masson, 1958
- [23] B. S. Duran, P. L. Odell : *Cluster Analysis, A Survey*, Springer-Verlag, 1974
- [24] J. Durbin : *Distribution Theory for Tests Based on the Sample Distribution Function*, Siam, 1973
- [25] J. D. Finn : *A General Model for Multivariate Analysis*, Holr,Rinehartand Winston, 1974
- [26] F. Gendre : *L'analyse statistique multivariée*, Librairie Droz, 1976
- [27] M. Goldstein, W. R. Dillon : *Discrete Discriminant Analysis*, Wiley, 1978
- [28] L. A. Goodman, W. H. Kruskal : *Measures of Association for Cross Classifications*, Springer-Verlag, 1979

- [29] W. C. Guenther : *Concepts of Statistical Inference*, McGraw hill, 1965
- [30] S. J. Haberman : *The Analysis of Frequency Data*, Chicago Un.Press, 1974
- [31] D. J. Hand : *Discrimination and Classification*, Wiley, 1981
- [32] H. H. Harman : *Modern Factor Analysis*, Chicago Un.Press, 1967
- [33] J. A. Hartigan : *Clustering Algorithms*, Wiley, 1975
- [34] P. J. Huber : *Robust Statistics*, Wiley, 1981
- [35] J. G. Kalbfleisch : *Probability and Statistical Inference*, I et II, Springer-Verlag, 1979
- [36] M. G. Kendall, A. Stuart : *The Advanced Theory of Statistics*, 3 vol., Griffin and Co, 1961
- [37] M. G. Kendall : *A Course in Multivariate Analysis*, Griffin, 1972
- [38] L. Lebart, J.-P. Fénelon : *Statistique et informatique appliquées*, Dunod 1975
- [39] J. Lefebvre : *Introduction aux analyses statistiques multidimensionnelles*, Masson, 1976
- [40] C. Lipson, N. J. Sheth : *Statistical Design and Analysis of Engineering Experiments*, McGraw Hill, 1973
- [41] J. S. Maritz : *Distribution-Free Statistical Methods*, Chapman and Hall, 1981
- [42] M. Masson : *Méthodologies générales de traitement statistique de l'information de masse*, Cedric-Nathan, 1980
- [43] R. G. Miller Jr. : *Simultaneous Statistical Inference*, Springer-Verlag, 1980
- [44] A. M. Mood, F. A. Graybill : *Introduction to the Theory of Statistics*, McGraw-Hill, 1963
- [45] J. E. Overall, C. J. Klett : *Applied Multivariate Analysis*, McGraw-Hill, 1972
- [46] J. H. Pollard : *A Handbook of Numerical and Statistical Techniques*, Cambridge U.P., 1977
- [47] H. Raiffa, R. Schlaifer : *Applied Statistical Decision Theory*, MIT Press, 1972
- [48] W. J. J. Rey : *Robust Statistical Methods*, Springer-Verlag, 1978
- [49] G. G. Roussas : *A First Course in Mathematical Statistics*, Adderson-Wesley, 1973
- [50] Séminaire de Statistique Orsay : *Théorie de la robustesse et estimation d'un paramètre*, Soc.Math.Fr. 1977
- [51] R. N. Shepard, A. K. Romney, S. B. Nerlove (ed.) : *Multidimensional Scaling*, Vol. I et II, Seminar Press, 1972
- [52] S. D. Silvey : *Statistical Inference*, Chapman and Hall, 1975
- [53] K. Takeuchi, H. Yanai, B. N. Mukherjee : *The Foundations of Multivariate Analysis*, Wiley, 1982
- [54] Tao Gu, B. Dubuisson : A lose-pattern process approach to clustering fuzzy data sets, IEEE Tr. Pat. An. ,v. PAMI-7, no 3, 1985, p.756
- [55] P. Tassi : *Méthodes Statistiques*, Economica, 1992
- [56] J. Torrens-Ibern : *Modèles et méthodes de l'analyse factorielle*, Dunod, 1972

- [57] G. J. C. Upton, B. Fingleton : *Spatial Data Analysis by Example*, Wiley, 1985
- [58] S. B. Vardeman : *Statistics for Engineering Problem Solving*, IEEE Press, 1994
- [59] S. Weisberg : *Applied Linear Regression*, Wiley, 1980



# Table des matières

<b>AVANT-PROPOS</b>	<b>1</b>
<b>1 INTRODUCTION</b>	<b>3</b>
<b>2 LES THÉORÈMES LIMITES</b>	<b>7</b>
2.1 Convergence et mesure de probabilité . . . . .	8
2.1.1 Convergence presque sûre . . . . .	8
2.1.2 Convergence en probabilité . . . . .	8
2.1.3 Convergence en moyenne quadratique . . . . .	9
2.1.4 Convergence en loi . . . . .	10
2.1.5 Relations entre les différentes convergences . . . . .	10
2.2 Quelques inégalités en probabilités . . . . .	10
2.3 Lois des grands nombres . . . . .	11
2.3.1 Loi forte des grands nombres . . . . .	11
2.3.2 Loi faible des grands nombres . . . . .	12
2.3.3 Conséquences des lois des grands nombres . . . . .	12
2.4 Théorème central limite (TCL) . . . . .	13
<b>3 LOIS DE PROBABILITÉS</b>	<b>15</b>
3.1 Loi de Bernoulli . . . . .	16
3.2 Loi binomiale . . . . .	17
3.3 Loi multinomiale . . . . .	19
3.4 Loi binomiale négative . . . . .	20
3.5 Loi hypergéométrique . . . . .	22
3.6 Loi de Poisson . . . . .	23
3.7 Loi uniforme discrète . . . . .	25
3.8 Loi uniforme continue . . . . .	26
3.9 Loi normale centrée, réduite (standardisée) . . . . .	27
3.10 Loi normale (loi de Laplace - Gauss) . . . . .	29
3.11 Loi log-normale . . . . .	31
3.12 Loi exponentielle . . . . .	32
3.13 Loi gamma . . . . .	34
3.14 La loi du $\chi^2$ (chi deux) . . . . .	36
3.15 Loi de Student (loi de t) . . . . .	37
3.16 Loi de Fisher (loi f) . . . . .	39
3.17 Loi de Weibull . . . . .	40

<b>4</b>	<b>APPROXIMATIONS DES LOIS</b>	<b>43</b>
4.1	Approximation des lois en utilisant la loi uniforme . . . . .	43
4.2	Approximation de $P(a < S_n \leq b)$ . . . . .	44
4.3	Approximation d'une loi discrète . . . . .	45
4.3.1	Approximation de la loi binomiale . . . . .	45
4.3.2	Approximation de la loi de Poisson . . . . .	46
<b>5</b>	<b>ÉCHANTILLONNAGE</b>	<b>47</b>
5.1	Les relations entre population et échantillon . . . . .	47
5.2	Échantillon avec des v.a. normales . . . . .	48
<b>6</b>	<b>INFÉRENCE</b>	<b>51</b>
6.1	Statistique exhaustive . . . . .	52
6.2	Famille complète de densités . . . . .	53
6.3	Estimateurs . . . . .	53
6.3.1	Biais d'un estimateur . . . . .	53
6.3.2	Convergence d'un estimateur . . . . .	54
6.3.3	Efficacité d'un estimateur . . . . .	54
6.3.4	Risque quadratique . . . . .	54
6.4	Amélioration d'un estimateur . . . . .	55
6.4.1	Conclusions . . . . .	55
6.5	Éléments pour un choix . . . . .	56
<b>7</b>	<b>ESTIMATION</b>	<b>57</b>
7.1	Estimation avec statistique complète et exhaustive . . . . .	58
7.2	Estimation en absence de statistique complète et exhaustive . . . . .	58
7.3	Estimateur du maximum de vraisemblance . . . . .	59
7.3.1	Cas où $\theta$ est scalaire . . . . .	60
<b>8</b>	<b>INTERVALLE DE CONFIANCE</b>	<b>61</b>
8.1	Définition d'un intervalle de confiance . . . . .	62
8.2	Construction des intervalles de confiance . . . . .	62
8.2.1	Intervalles de confiance pour les moyennes . . . . .	64
8.2.1.1	Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec variance connue . . . . .	64
8.2.1.2	Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec variance inconnue . . . . .	64
8.2.2	Intervalle de confiance pour les variances . . . . .	65
8.2.2.1	Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec moyenne connue . . . . .	65
8.2.2.2	Loi normale $\mathcal{N}(\mu, \sigma^2)$ avec moyenne inconnue . . . . .	66
<b>9</b>	<b>TESTS D'HYPOTHÈSES</b>	<b>67</b>
9.1	Choix d'un test . . . . .	69
9.2	Test de la moyenne avec une valeur de référence . . . . .	71
9.2.1	Variance $\sigma^2$ de la population connue . . . . .	72
9.2.1.1	Test simple . . . . .	72
9.2.1.2	Test double . . . . .	72

9.2.2	Variance $\sigma^2$ inconnue . . . . .	73
9.2.2.1	Test simple . . . . .	73
9.2.2.2	Test double . . . . .	74
9.3	Test de la variance avec une valeur de référence . . . . .	75
9.3.0.3	Test simple . . . . .	75
9.3.0.4	Test double . . . . .	76
9.4	Comparaison entre deux moyennes . . . . .	76
9.4.1	Variances $\sigma_X^2, \sigma_Y^2$ des populations connues . . . . .	77
9.4.1.1	Test simple . . . . .	77
9.4.1.2	Test double . . . . .	77
9.4.2	Variances des populations inconnues mais égales . . . . .	78
9.4.2.1	Test simple . . . . .	78
9.4.2.2	Test double . . . . .	79
9.4.3	Variances des populations inconnues et non égales . . . . .	80
9.4.3.1	Test simple . . . . .	80
9.4.3.2	Test double . . . . .	80
9.5	Comparaison entre deux variances . . . . .	81
9.5.0.3	Test simple . . . . .	81
9.5.0.4	Test double . . . . .	82
<b>10</b>	<b>CONTRÔLE DE RÉCEPTION</b>	<b>83</b>
10.1	Comparaison d'une proportion à une valeur de référence . . . . .	84
10.1.1	Loi binomiale . . . . .	84
10.1.2	Loi de Poisson . . . . .	85
10.1.3	Loi normale . . . . .	85
10.2	Comparaison de deux proportions . . . . .	86
10.2.1	Loi hypergéométrique . . . . .	86
10.2.2	Loi normale . . . . .	87
<b>11</b>	<b>ANALYSE DE VARIANCE</b>	<b>89</b>
11.1	Un facteur contrôlé . . . . .	90
11.2	Deux facteurs contrôlés . . . . .	92
<b>12</b>	<b>MODÈLE LINÉAIRE À m VARIABLES EXPLICATIVES</b>	<b>95</b>
12.1	Notations . . . . .	95
12.2	Modèle linéaire . . . . .	96
12.3	Propriétés de l'estimation . . . . .	97
12.4	Coefficient de corrélation multiple . . . . .	97
12.5	Tests de signification . . . . .	98
12.5.1	Tests sur $R^2$ . . . . .	98
12.5.2	Comparaison d'un coefficient à une valeur de référence . . . . .	98
12.5.3	Test sur un ensemble des variables explicatives . . . . .	98
<b>13</b>	<b>ANALYSE DE DONNÉES</b>	<b>99</b>
13.1	Calcul des facteurs . . . . .	99
13.2	Visualisation du nuage des observations . . . . .	102
13.3	Relations entre $\mathbb{R}^n$ et $\mathbb{R}^m$ . . . . .	103

<b>14 LES PRINCIPALES MÉTHODES DE L'ANALYSE DE DONNÉES</b>	<b>107</b>
14.1 Analyse en composantes principales . . . . .	107
14.2 Analyse factorielle des correspondances . . . . .	108
<b>15 STRUCTURATION DES DONNÉES</b>	<b>113</b>
15.1 Analyse factorielle discriminante . . . . .	114
15.2 Classification . . . . .	115
15.2.1 Algorithme de K-moyennes (K-Means) . . . . .	116
15.2.2 Nuées dynamiques . . . . .	116
15.3 Classement . . . . .	118
15.3.1 N plus proches voisins . . . . .	118
15.3.2 Classement flou . . . . .	119
<b>A RAPPELS DES PROBABILITÉS</b>	<b>121</b>
A.1 Définition de la probabilité . . . . .	122
A.1.1 Définition selon le modèle uniforme . . . . .	122
A.1.2 Définition fréquentielle . . . . .	123
A.1.3 Définition axiomatique . . . . .	124
A.2 Des événements aux ensembles . . . . .	126
A.3 Propriétés de la probabilité . . . . .	126
A.4 Continuité des mesures de probabilité . . . . .	128
A.5 Construction d'espaces probabilisés . . . . .	129
A.6 Probabilités conditionnelles . . . . .	130
A.7 Théorème de Bayes . . . . .	131
A.8 Événements indépendants . . . . .	132
A.9 Espace produit. Expériences aléatoires répétées . . . . .	133
A.10 La notion de la variable aléatoire discrète . . . . .	134
A.11 Espérance mathématique et variance d'une v.a. . . . .	135
A.12 Moments . . . . .	138
A.13 La notion de la variable aléatoire continue . . . . .	139
A.14 Fonction de répartition . . . . .	139
A.15 Fonctions de densité . . . . .	141
A.16 Espérance mathématique - Variance . . . . .	142
A.17 Moments - Médiane - Quantiles . . . . .	143
A.18 Fonctions des variables aléatoires . . . . .	145
A.19 Extension au cas des v.a. discrètes . . . . .	145
A.20 Variables aléatoires à plusieurs dimensions . . . . .	146
A.21 Couples de variables aléatoires . . . . .	148
A.21.1 Lois marginales . . . . .	149
A.21.2 Lois conditionnelles . . . . .	150
A.21.3 Moments - Covariance . . . . .	150
A.21.4 Coefficient de corrélation . . . . .	150
Bibliographie	