

# Introduction aux Probabilités et aux Statistiques

par Benoît Mselati et Florent Benaych-Georges ([benaych@dma.ens.fr](mailto:benaych@dma.ens.fr))

24 septembre 2006

# Table des matières

<b>1</b>	<b>Notions de base</b>	<b>3</b>
1.1	Définitions . . . . .	3
1.2	Lois usuelles . . . . .	7
1.2.1	Lois usuelles discrètes . . . . .	7
1.2.2	Lois à densité usuelles . . . . .	9
1.2.3	Loi normale . . . . .	10
<b>2</b>	<b>Les théorèmes limites fondamentaux</b>	<b>12</b>
2.1	Deux types de convergence . . . . .	12
2.2	Loi des grands nombres . . . . .	13
2.3	Le théorème central limite . . . . .	14
2.4	L'approximation de Poisson . . . . .	16
<b>3</b>	<b>Vecteurs gaussiens</b>	<b>18</b>
3.1	Variables aléatoires gaussiennes . . . . .	18
3.2	Les vecteurs gaussiens . . . . .	19
3.2.1	Définition . . . . .	19
3.2.2	Loi d'un vecteur aléatoire . . . . .	20
3.2.3	Transformée de Fourier d'un vecteur gaussien . . . . .	20
3.2.4	Variables aléatoires gaussiennes indépendantes . . . . .	22
3.2.5	Image d'un vecteur gaussien par une application linéaire . . . . .	24
3.3	Le théorème central limite vectoriel . . . . .	24
3.4	Carrés de gaussiennes . . . . .	24
3.4.1	Les lois gamma . . . . .	25
3.4.2	Carrés de gaussiennes - Lois du chi-deux . . . . .	26
3.4.3	Lois de Student . . . . .	27
<b>4</b>	<b>Statistiques</b>	<b>28</b>
4.1	Modèles à paramètres . . . . .	28
4.1.1	Définition . . . . .	28
4.1.2	Exemples de modèles paramétriques . . . . .	29
4.2	Estimateurs . . . . .	29
4.2.1	Estimateur du maximum de vraisemblance . . . . .	29
4.2.2	Estimateurs sans biais, estimateurs convergents . . . . .	30
4.2.3	Estimateurs de l'espérance et de la variance . . . . .	32
4.3	Intervalles de confiance non asymptotiques . . . . .	33
4.3.1	Définitions . . . . .	33
4.3.2	Exemple : le modèle de Bernoulli . . . . .	34
4.3.3	Exemple : le modèle gaussien . . . . .	35
4.4	Intervalles de confiance asymptotiques . . . . .	40
4.4.1	Définition . . . . .	40
4.4.2	Intervalles de confiance asymptotiques pour l'espérance - Cas général . . . . .	40

4.4.3	Intervalles de confiance asymptotiques pour l'espérance - Modèle de Bernoulli	41
4.5	Tests . . . . .	43
4.5.1	Formulation . . . . .	43
4.5.2	Lien avec les intervalles de confiance . . . . .	44
4.5.3	L'exemple des modèles gaussiens . . . . .	45
4.6	Modèle de regression simple . . . . .	48
4.6.1	Introduction . . . . .	48
4.6.2	Théorie . . . . .	48
4.6.3	Test d'hypothèses . . . . .	52

# Chapitre 1

## Notions de base

### Introduction

La théorie des probabilités est la modélisation que les mathématiques font des phénomènes dans lesquels intervient ce que l'on appelle le hasard. Elle repose sur un certain nombre de notions : événement, probabilité d'un événement, variable aléatoire, loi, espérance d'une variable aléatoire. Nous ne donnerons de ces notions que des définitions intuitives. Elles sont suffisantes pour l'usage simple que l'on en fera, mais des fondements plus rigoureux sont nécessaires à une utilisation un peu plus approfondie de la théorie des probabilités.

### 1.1 Définitions

La meilleure définition intuitive du concept d'*événement* est celle donnée par son sens courant. On peut dire par exemple que c'est un phénomène  $E$  qui peut se produire, et aussi ne pas se produire, et qui se produit avec une certaine probabilité  $P(E)$ .

Une *variable aléatoire* (en abrégé *v.a.*) est une donnée dont la valeur précise dépend du hasard. Ce peut être un nombre réel (on parle alors de *v.a. réelle*) ou complexe (*v.a. complexe*), un vecteur de  $\mathbb{R}^n$  ou  $\mathbb{C}^n$  (on parle alors de *vecteur aléatoire de dimension  $n$* ), ou une donnée non numérique, comme une couleur (par exemple quand on prend une boule au hasard dans une urne) ou une carte (quand on tire une carte au hasard).

Pour  $X$  v.a. à valeurs dans  $K$  ( $K = \mathbb{R}, \mathbb{C}$ , ou  $\mathbb{R}^n$ ), la *loi* de  $X$  est l'application de l'ensemble  $\mathcal{P}(K)$  des parties de  $K$  (le plus souvent, on ne considère pas toutes les parties de  $K$ , mais seulement celles qui sont dites boréliennes, mais nous passerons sur cette subtilité) vers  $[0, 1]$  qui à une partie  $A$  de  $K$  associe la probabilité  $P(\{X \in A\})$  (notée plus simplement  $P(X \in A)$ ) de l'événement  $\{X \in A\}$ . Cette définition de la loi comme une application est assez abstraite et ne correspond pas à l'usage qu'on en fera. Le plus souvent, on parlera de la loi d'une v.a.  $X$  en disant " $X$  suit une loi binomiale de paramètres  $n, p$ ", ou " $X$  a une loi gaussienne de paramètres  $(m, \sigma^2)$ ", etc... ce qui signifie respectivement

$$\begin{aligned} \forall k \in \{0, \dots, n\}, P(X = k) &= C_n^k p^k (1-p)^{n-k}, \\ \forall a < b, P(a \leq X \leq b) &= \frac{1}{\sqrt{2\pi}\sigma} \int_a^b \exp(-(x-m)^2/(2\sigma^2)) dx, \text{ etc...} \end{aligned}$$

**Exemple :** Les *lois à densité* constituent un exemple particulièrement intéressant. Une v.a. réelle  $X$  est dite avoir une loi à densité si il existe une fonction continue par morceaux  $p : \mathbb{R} \rightarrow \mathbb{R}^+$  telle que l'intégrale pour tout  $a < b \in \mathbb{R}$ ,

$$P(a \leq X \leq b) = \int_a^b p(x) dx. \tag{1.1}$$

On remarque que dans ce cas, l'intégrale  $\int_{-\infty}^{+\infty} p(x)dx$  converge. En effet,  $p \geq 0$  et pour tout  $a < b \in \mathbb{R}$ ,  $\int_a^b p(x)dx \leq 1$ . On admet que l'équation (1.1) reste vraie quand  $a = -\infty$  et/ou  $b = +\infty$ . La fonction  $p$  est appelée *densité* de  $X$ , à ne pas confondre avec la fonction  $F_X : x \mapsto P(X \leq x)$ , dite *fonction de répartition* de  $X$  (c'est la primitive de  $p$  qui a pour limite zéro en  $-\infty$ ).

Des variables aléatoires sont dites *identiquement distribuées* (en abrégé i.i.d.) lorsqu'elles ont la même loi. Attention, ça ne signifie pas qu'elles sont égales. Par exemple, si on jette un dé deux fois, la loi du résultat obtenu au premier lancé est la même que celle du résultat obtenu au deuxième lancé, ces résultats sont donc des v.a.i.d., mais il arrive (souvent) qu'elles soient différentes. Mieux : lors du lancé d'un dé, la loi de la face tournée vers le bas et la loi de la face tournée vers le haut sont les mêmes, alors que ces deux v.a. ne prennent jamais la même valeur.

Soient  $E, F$  des événements. Considérer que  $E$  et  $F$  sont indépendants (au sens courant) revient à dire que savoir que  $E$  se produit ne donne aucune information sur la réalisation de  $F$ . Ceci se traduit par

$$P(F) = P(F \text{ sachant } E) \text{ (noté } P(F|E)\text{)}.$$

Il est assez naturel de considérer que  $P(F|E) = P(E \cap F)/P(E)$ , donc l'indépendance de  $E$  et  $F$  reviendrait à l'équation  $P(F) = P(E \cap F)/P(E)$ . Nous adoptons cette définition, et donc deux événements  $E, F$  sont dits *indépendants* si

$$P(E \cap F) = P(E)P(F).$$

Deux variables aléatoires  $X, Y$  à valeurs respectives dans  $K, L$  sont dites *indépendantes* si pour tous  $A \subset K, B \subset L$ , les événements  $\{X \in A\}, \{Y \in B\}$  sont indépendants, autrement dit si

$$\forall A \subset K, B \subset L, P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

De façon plus générale, des v.a.  $X_1, \dots, X_n$  à valeurs respectives dans  $K_1, \dots, K_n$  sont dites *indépendantes* si  $\forall A_1 \subset K_1, A_2 \subset K_2, \dots, A_n \subset K_n$ ,

$$P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n),$$

et une famille infinie (par exemple indexée par  $\mathbb{N}$ , comme on le verra pour le TCL et la LGN) de v.a. est dite *indépendante* si toutes ses sous familles finies sont indépendantes.

On admettra que l'indépendance de v.a. réelles  $X_1, \dots, X_n$  est équivalente au fait que

$$\forall a_1, \dots, a_n \in \mathbb{R}, P(X_1 \leq a_1, \dots, X_n \leq a_n) = P(X_1 \leq a_1) \cdots P(X_n \leq a_n).$$

Autrement dit, dans la définition, il suffit de prendre des  $A_i$  de la forme  $] - \infty, a_i]$ .

**Exercice 1.1.1** Montrer que si des v.a.  $X_1, \dots, X_n$  sont indépendantes, alors pour tout  $p \in \{1, \dots, n\}$ , pour tous  $1 \leq i_1 < i_2 < \dots < i_p \leq n$ , les v.a.  $X_{i_1}, \dots, X_{i_p}$  sont indépendantes.

**Exercice 1.1.2** Montrer que si des v.a. complexes  $X_1, \dots, X_n$  sont indépendantes, alors pour toutes fonctions  $F_1, F_2, \dots, F_n$  définies sur  $\mathbb{C}$  à valeurs dans  $\mathbb{C}$ , les v.a.  $F_1(X_1), \dots, F_n(X_n)$  sont indépendantes.

On admet le théorème suivant (sa démonstration nécessite des définitions plus rigoureuses et le recours aux tribus) :

**Théorème 1.1.1 (Théorème des coalitions)** Soit  $m \geq 1$ , soient  $X_1, \dots, X_n$  des variables aléatoires réelles indépendantes.

Soient  $p \in \{1, \dots, n\}$  et  $1 \leq i_1 < i_2 < \dots < i_p \leq n$

Alors les  $p$  vecteurs aléatoires concaténés

$$(X_1, \dots, X_{i_1}), (X_{i_1+1}, \dots, X_{i_2}), \dots, (X_{i_{p-1}+1}, \dots, X_{i_p})$$

sont indépendants. De plus, pour toutes fonctions  $F_1$  de  $i_1$  variables,  $F_2$  de  $i_2 - i_1$  variables, ...,  $F_p$  de  $i_p - i_{p-1}$  variables, les v.a.

$$F_1(X_1, \dots, X_{i_1}), F_2(X_{i_1+1}, \dots, X_{i_2}), \dots, F_p(X_{i_{p-1}+1}, \dots, X_{i_p})$$

sont indépendantes.

Intuitivement, l'espérance de gain associée à un revenu aléatoire est la moyenne des valeurs que peut prendre ce gain, pondérées par les probabilités que ces valeurs soient obtenues. On définit donc l'*espérance*  $\mathbb{E}(X)$  d'une v.a. réelle ou complexe  $X$  ne prenant qu'un nombre fini  $x_1, \dots, x_k$  de valeurs comme la moyenne des nombres  $x_1, \dots, x_k$  pondérés par les probabilités respectives  $P(X = x_1), \dots, P(X = x_n)$ . Autrement dit

$$\mathbb{E}(X) = \sum_{i=1}^k P(X = x_i)x_i.$$

On admet que la notion d'espérance s'étend à certaines variables aléatoires réelle ou complexes ne prenant pas forcément un nombre fini de valeurs. Ces v.a. sont dites *intégrables*. Par exemple, lorsque  $X$  est une v.a. réelle positive ( $X \geq 0$ ), on a

$$\mathbb{E}(X) = \lim_{n \rightarrow +\infty} \sum_{k=0}^{+\infty} P\left(\frac{k}{n} \leq X < \frac{k+1}{n}\right) \frac{k}{n}$$

(et si la limite n'est pas définies, c'est que  $X$  n'est pas une variable aléatoire intégrable).

La différence entre les v.a. intégrables et celles qui ne le sont pas dépasse le niveau de ce cours, et nous nous contenteront de donner un critère d'intégrabilité pour les v.a. à densité.

N'ayant pas défini rigoureusement l'espérance  $\mathbb{E}(X)$  d'une v.a.  $X$  (réelle ou complexe, et intégrable, mais nous omettrons dorenavant ces précisions), nous ne pouvons qu'admettre les propriétés suivantes :

- si  $X$  est positive,  $\mathbb{E}(X) \in \mathbb{R}^+$  et on a **l'inégalité de Chebichev** :

$$\forall a > 0, P(X \geq a) \leq \frac{\mathbb{E}(X)}{a};$$

- pour  $X, Y$  v.a.,  $\alpha, \beta \in \mathbb{R}$  (ou  $\mathbb{C}$ ),  $\mathbb{E}(\alpha X + \beta Y) = \alpha \mathbb{E}(X) + \beta \mathbb{E}(Y)$  ;
- pour  $X, Y$  v.a. complexes (donc éventuellement réelles),

$$X, Y \text{ indépendantes} \Rightarrow \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y).$$

De plus, si  $X$  est une v.a. à densité  $p$ ,  $X$  est intégrable si et seulement si la fonction  $x \mapsto xp(x)$  est intégrable sur  $\mathbb{R}$ , et dans ce cas,

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xp(x)dx.$$

De façon plus générale, pour toute fonction  $f : \mathbb{R} \rightarrow \mathbb{C}$  continue par morceaux,  $f(X)$  est intégrable si et seulement si la fonction  $x \mapsto |f(x)|p(x)$  est intégrable sur  $\mathbb{R}$ , et dans ce cas,

$$\mathbb{E}(f(X)) = \int_{-\infty}^{+\infty} f(x)p(x)dx.$$

Dans le cas où  $f$  est à valeurs complexes,  $\int_{-\infty}^{+\infty} f(x)p(x)dx$  n'a pas été défini en classe prépa, mais on le définit comme étant

$$\int_{-\infty}^{+\infty} \Re(f(x))p(x)dx + i \int_{-\infty}^{+\infty} \Im(f(x))p(x)dx.$$

On définit la *variance*  $\text{Var}(X)$  d'une v.a. réelle  $X$  par la formule

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

et la covariance de deux v.a. réelles  $X, Y$  par la formule

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

**Exercice 1.1.3** 1. Exprimer, pour  $X_1, \dots, X_n$  v.a. réelles, la variance d'une combinaison linéaire  $\lambda_1 X_1 + \dots + \lambda_n X_n$  à partir des covariances (et des  $\lambda_i$ ).

2. Deux v.a.  $X, Y$  sont dites *non corellées* si leur covariance est nulle. Montrer que

$$\begin{aligned} X, Y \text{ indépendantes} &\Rightarrow X, Y \text{ non corellées, et} \\ X_1, \dots, X_n \text{ 2 à 2 non corellées} &\Rightarrow \text{Var}(X_1 + \dots + X_n) \\ &= \text{Var}(X_1) + \dots + \text{Var}(X_n). \end{aligned}$$

**Exercice 1.1.4** Par récurrence sur  $n \geq 2$  et en utilisant le théorème des coalissions, montrer que si  $X_1, \dots, X_n$  sont des v.a. indépendantes,

$$\mathbb{E}(X_1 \cdots X_n) = \mathbb{E}(X_1) \cdots \mathbb{E}(X_n). \quad (1.2)$$

La dernière définition de cette section est celle de la *transformée de Fourier* (parfois appelée *fonction caractéristique*, sans lien avec la fonction caractéristique  $\Xi_A$  d'un ensemble  $A$ ),  $\varphi_X$  d'une v.a. réelle  $X$ .  $\varphi_X$  est l'application de  $\mathbb{R}$  vers  $\mathbb{C}$  qui à  $t \in \mathbb{R}$  associe l'espérance

$$\varphi_X(t) = \mathbb{E}(e^{itX})$$

de la v.a. complexe  $e^{itX}$ . On définit de la même façon la transformée de Fourier (ou fonction caractéristique)  $\varphi_X$  d'un vecteur aléatoire  $X = (X_1, \dots, X_n)$  à valeurs dans  $\mathbb{R}^n$  : c'est l'application de  $\mathbb{R}^n$  vers  $\mathbb{C}$  qui à un vecteur  $t = (t_1, \dots, t_n)$  de  $\mathbb{R}^n$  associe l'espérance

$$\varphi_X(t) = \mathbb{E}[\exp(i \sum_{k=1}^n t_k X_k)]$$

de la v.a. complexe  $\exp(i \sum_{k=1}^n t_k X_k)$ .

**Exercice 1.1.5** Donner  $\varphi_X$  lorsque  $X$  est une v.a. ne prenant que les valeurs  $x_1, \dots, x_n$ , et ce avec probabilités respectives  $p_1, \dots, p_n$ .

On admettra le théorème fondamental suivant :

**Théorème 1.1.2** *Deux v.a. réelles (et, plus généralement, deux vecteurs aléatoires de  $\mathbb{R}^n$ ) ont même loi si et seulement si leurs transformées de Fourier sont les mêmes en tout point  $t$  de  $\mathbb{R}$  (ou de  $\mathbb{R}^n$ ).*

Ainsi, la transformée de Fourier d'une v.a. caractérise sa loi. Aussi surprenant que cela puisse paraître, ce résultat nous sera très utile. On s'en servira de la façon suivante. On va calculer une fois pour toutes les transformées de Fourier des v.a. de lois usuelles (gaussienne, exponentielle, etc...), et pour déterminer la loi d'une v.a.  $X$ , on calculera sa transformée de Fourier. Si on tombe sur une de celles des lois usuelles, on n'aura plus à chercher plus loin, on connaîtra la loi de  $X$ .

**Fonctions caractéristiques et indépendance** Si les coordonnées  $X_1, \dots, X_n$  du vecteur aléatoire réel  $X = (X_1, \dots, X_n)$  sont des v.a. réelles indépendantes, alors, pour tout vecteur  $t = (t_1, \dots, t_n)$  de  $\mathbb{R}^n$ , les v.a. complexes  $e^{it_1 X_1}, \dots, e^{it_n X_n}$  sont indépendantes (par l'exercice 1.1.2), donc, par l'équation (1.2), on a

$$\varphi_X(t) = \mathbb{E}[\prod_{k=1}^n \exp(it_k X_k)] = \prod_{k=1}^n \mathbb{E}[\exp(it_k X_k)] = \prod_{k=1}^n \varphi_{X_k}(t_k).$$

En fait, la réciproque est vraie aussi :

**Théorème 1.1.3** *Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire de  $\mathbb{R}^n$ . Alors les v.a. réelles  $X_1, \dots, X_n$  sont indépendantes si et seulement si pour tout vecteur  $t = (t_1, \dots, t_n)$  de  $\mathbb{R}^n$ ,*

$$\varphi_X(t) = \prod_{k=1}^n \varphi_{X_k}(t_k).$$

Ainsi, la transformée de Fourier caractérise aussi l'indépendance.

## 1.2 Lois usuelles

### 1.2.1 Lois usuelles discrètes

Les lois discrètes sont les lois sur  $\mathbb{R}$  concentrées sur un ensemble dénombrable (i.e. fini ou indexé par  $\mathbb{N}$ ).

#### Loi de Bernoulli

Une loi de Bernoulli est en toute généralité la loi d'une variable aléatoire ne prenant que 2 valeurs. Typiquement, on considère une variable aléatoire  $X$  ne prenant que les valeurs 0 et 1. La loi de  $X$  est déterminée par la donnée d'un réel  $p$  compris entre 0 et 1 tel que

$$P(X = 0) = 1 - p, \quad P(X = 1) = p.$$

Une telle loi est notée  $\mathcal{B}(p)$  et appelée "Bernoulli de paramètre  $p$ ". On a :

- $\mathbb{E}(X) = p$ .
- $\text{Var}(X) = p(1 - p)$ .

#### Loi binômiale

Une loi binômiale est la loi de la somme de  $n$  variables indépendantes identiquement distribuées (en abrégé *i.i.d.*) de Bernoulli de paramètre  $p$ . Autrement dit

$$X = X_1 + \dots + X_n$$

où  $X_1, \dots, X_n$  sont des variables aléatoires indépendantes vérifiant pour tout  $i$

$$P(X_i = 0) = 1 - p, \quad P(X_i = 1) = p.$$

La variable aléatoire  $X$  prend ses valeurs dans l'ensemble  $\{1, \dots, n\}$ . Pour que  $X$  soit égal à  $k$ , il faut que  $k$  parmi les  $n$  v.a.  $X_1, \dots, X_n$  soient égales à 1 et que les  $n - k$  autres soient égales à 0. Des arguments combinatoires simples nous donnent

$$P(X = k) = C_n^k p^k (1 - p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k} \text{ pour tout } 1 \leq k \leq n.$$

Nous disons que  $X$  suit la loi  $\mathcal{B}(n, p)$ . On vérifie les identités suivantes :

- $\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X_i) = np$ .
- $\text{Var}(X) = \sum_{i=1}^n \text{Var}(X_i) = np(1 - p)$ .

#### Loi géométrique

Soit  $0 \leq p \leq 1$ . La loi géométrique de paramètre  $p$  est la loi sur  $\mathbb{N}^*$  d'une variable aléatoire  $X$  telle que

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

Notons que, pour tout  $n \geq 1$ ,

$$P(X \geq n) = \sum_{k \geq n} (-p)p^k = (1 - p)p^n \sum_{k \geq 0} p^k = (1 - p)p^n / (1 - p) = p^n.$$

La loi géométrique possède ainsi la propriété remarquable suivante : si  $n \geq 1$ , et  $m \geq 0$ ,

$$P(X \geq n + m | X \geq n) = P(X \geq m).$$

Cette propriété est appelée propriété de non-vieillesse et elle justifie l'utilisation de lois géométriques pour modéliser des durées de vie d'objets ne s'usant pas : sachant que  $X$  est au moins égal à  $n$ , la probabilité que  $X$  soit plus grand que  $n + m$ , est égal à la probabilité initiale d'atteindre  $m$ .



## Lois de Poisson

Rappelons que pour tout  $z \in \mathbb{C}$ ,

$$\exp(z) = \sum_{n \geq 0} \frac{z^n}{n!}.$$

La loi de Poisson de paramètre  $\theta \geq 0$  notée  $\mathcal{P}(\theta)$  est la loi d'une variable aléatoire  $X$  à valeurs dans  $\mathbb{N}$  telle que

$$P(X = n) = \exp(-\theta) \frac{\theta^n}{n!}.$$

On laisse au lecteur le soin de vérifier les propriétés suivantes.

**Proposition 1.2.1** Soit  $X$  de loi  $\mathcal{P}(\theta)$ .

- La fonction caractéristique de  $X$  est

$$\varphi_X(t) = \exp(\theta(\exp(it) - 1)).$$

- L'espérance de  $X$  et la variance de  $X$  sont données par

$$\mathbb{E}(X) = \theta, \quad \text{Var}(X) = \theta.$$

**Proposition 1.2.2** Supposons que  $X_1, \dots, X_n$  soient des variables aléatoires indépendantes suivant des lois de Poisson de paramètres respectifs  $\theta_1, \dots, \theta_n$ . Alors la loi de  $X_1 + \dots + X_n$  est  $\mathcal{P}(\theta_1 + \dots + \theta_n)$ .

*Preuve :* Posons

$$X = X_1 + \dots + X_n.$$

Soit  $t \geq 0$ . Par indépendance,

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}(\exp(itX)) \\ &= \mathbb{E}(\exp(itX_1) \dots \exp(itX_n)) \\ &= \mathbb{E}(\exp(itX_1)) \dots \mathbb{E}(\exp(itX_n)) \\ &= \varphi_{X_1}(t) \dots \varphi_{X_n}(t). \end{aligned}$$

Grâce au calcul de la fonction caractéristique plus haut, on a

$$\varphi_X(t) = \exp(\theta_1(\exp(it) - 1)) \dots \exp(\theta_n(\exp(it) - 1)) = \exp(\theta(\exp(it) - 1)),$$

où  $\theta = \theta_1 + \dots + \theta_n$ . On reconnaît la fonction caractéristique d'une loi de Poisson de paramètre  $\theta$ . Donc  $X$  suit bien une loi exponentielle de paramètre  $\theta$ .  $\square$

**Exercice 1.2.1** Soit  $X$  une v.a. de loi  $\mathcal{P}(\lambda)$ ,  $\lambda > 0$ . Déterminez l'événement  $X = n$  le plus probable.

**Indication :**  $\forall n \in \mathbb{N}, P(X = n) = \exp(-\lambda) \frac{\lambda^n}{n!}$ , et la fonction  $n \mapsto \frac{\lambda^n}{n!}$  ne se dérive pas aisément sur  $\mathbb{R}^+$  (à cause du  $n!$ , qui n'a, à priori, pas de sens si  $n \notin \mathbb{N}$ ). La bonne méthode est donc de calculer  $\frac{X=n+1}{X=n}$ .

### Exercice 1.2.2 Lois multinômiales

On considère  $n$  variables aléatoires  $X_1, \dots, X_n$  i.i.d. prenant pour simplifier leurs valeurs dans l'ensemble fini  $\{1, \dots, k\}$ . Il existe donc des réels  $p_1, \dots, p_k \geq 0$  tels que  $p_1 + \dots + p_k = 1$  et

$$P(X_j = i) = p_i$$

pour tout  $1 \leq j \leq n$ . On note alors pour tout  $1 \leq i \leq k$ ,  $N(i)$  le nombre de  $X_j$  égaux à  $i$  parmi les  $n$  nombres  $X_1, \dots, X_n$ . En particulier, on a toujours

$$\sum_{i=1}^k N(i) = n.$$

On note  $N$  le vecteur  $(N(1), \dots, N(k))$ . La loi du vecteur aléatoire  $N$  est appelée loi multinômiale de paramètres  $n$  et  $p_1, \dots, p_k$ .

L'objet de cet exercice est calculer la loi de  $N$  sans utiliser d'argument combinatoire.

Pour cela, on définit  $F$  comme la fonction de  $k$  variables réelles  $x_1, \dots, x_k$  définie par

$$F(x) = F(x_1, \dots, x_k) = \mathbb{E} \left( \prod_{i=1}^k x_i^{N(i)} \right).$$

1. En décomposant selon les valeurs possibles prises par  $N$ , montrer que

$$F(x) = \sum_{n_1 + \dots + n_k = n} \mathbb{P}(N(1) = n_1, \dots, N(k) = n_k) x_1^{n_1} \dots x_k^{n_k}, \quad (1.3)$$

où la somme est prise sur tous les  $k$ -uplets  $(n_1, \dots, n_k)$  de  $\mathbb{N}^k$  tels que  $n_1 + \dots + n_k = n$ .

2. Calculer les dérivées  $n$ -ièmes de  $F$ .
3. En utilisant le caractère i.i.d. des  $X_j$ , montrer que

$$F(x) = \left( \sum_{i=1}^k p_i x_i \right)^n.$$

4. En déduire que

$$\mathbb{P}(N(1) = n_1, \dots, N(k) = n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

si  $n_1 + \dots + n_k = n$ .

**Exercice 1.2.3 Processus de Poisson à temps discret** Un processus de comptage est par définition une suite de variables aléatoires  $(N_n)_{n \geq 0}$  telle que

$$N_0 = 0 \text{ et } N_{n+1} - N_n = 0 \text{ ou } 1, \text{ pour tout } n \geq 0.$$

Nous définissons alors les temps de saut  $(S_n)_{n \geq 1}$  comme la suite de variables aléatoires à valeurs dans  $\mathbb{N}^*$  définie par

$$S_1 = \min\{n \geq 1 ; N_n - N_{n-1} = 1\}$$

et, pour tout  $n \geq 2$ ,

$$S_n = \min\{n \geq 1 ; N_{S_{n-1}+n} - N_{S_{n-1}} = 1\}.$$

Montrer l'équivalence des deux conditions suivantes :

- Les accroissements  $N_{n+1} - N_n, n \geq 0$  sont des variables i.i.d. de Bernoulli de paramètre  $p$ .
- Les v.a.  $S_n, n \geq 0$  sont des variables i.i.d. de loi géométrique de paramètre  $p$ .

Si les conditions précédentes sont vérifiées, on dit que  $(N_n)_{n \geq 0}$  est un processus de Poisson à temps discret.

## 1.2.2 Lois à densité usuelles

**Loi uniforme sur  $[0, 1]$**

Soit  $U \sim 1_{[0,1]}(x)dx$ . On entend par là que  $U$  a pour densité  $1_{[0,1]}$  (fonction indicatrice de l'intervalle  $[0, 1]$ ). La loi de  $U$  est appelée loi uniforme sur  $[0, 1]$ . La fonction de répartition de  $U$  est donnée par

$$\mathbb{P}(U \leq x) = \begin{cases} 0 & \text{si } x < 0. \\ x & \text{si } 0 \leq x \leq 1. \\ 1 & \text{si } x > 1. \end{cases}$$

On peut "construire" toute loi à partir de la loi uniforme au sens de la proposition suivante.

**Proposition 1.2.3** Soit  $X$  une variable aléatoire de fonction de répartition  $F$ . Notons  $F^{-1}$  l'inverse de  $F$  à gauche définie par

$$\forall x \in ]0, 1[, F^{-1}(x) = \inf\{y ; F(y) \geq x\}.$$

Alors  $X$  a même loi que  $F^{-1}(U)$ .

**Remarques :**

- Si  $F$  est strictement croissante,  $F$  est inversible et l'inverse à gauche coïncide avec l'inverse au sens usuel.

- On parle d'inverse à gauche car  $F^{-1}$  est toujours continue à gauche.

- Si on peut calculer explicitement  $F^{-1}$ , pour simuler  $X$ , il suffit donc de savoir simuler  $U$ , ce qui est classiquement programmé sur ordinateur.

*Preuve :* Montrons que pour tout  $x \in [0, 1]$  et tout  $y \in \mathbb{R}$ ,

$$F^{-1}(x) \leq y \iff x \leq F(y).$$

Supposons d'abord que  $x \leq F(y)$ . Cela implique que

$$y \geq \inf\{z ; F(z) \geq x\} = F^{-1}(x).$$

Inversement, supposons que

$$x > F(y).$$

Cela signifie que

$$P(X \leq y) < x.$$

La fonction de répartition étant toujours continue à droite et étant croissante, il existe  $\varepsilon > 0$  tel que

$$P(X \leq z) < x$$

pour tout  $z \leq y + \varepsilon$ . Il est donc clair que

$$\inf\{z ; F(z) \geq x\} \geq z + \varepsilon > y.$$

Nous avons montré l'équivalence voulue.

Soit maintenant  $y \in \mathbb{R}$ . Par le résultat précédent, comme  $U$  suit une loi uniforme sur  $[0, 1]$ , on a

$$P(F^{-1}(U) \leq y) = P(U \leq F(y)) = F(y).$$

Par conséquent,  $F^{-1}(U)$  et  $X$  ont même fonction de répartition et donc même loi.  $\square$

### 1.2.3 Loi normale

On dit que  $X$  suit une loi gaussienne (ou normale) d'espérance  $m$  et de variance  $\sigma^2 > 0$  si et seulement si  $X$  a pour densité

$$\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \frac{1}{\sigma\sqrt{2\pi}}.$$

On note alors  $X \sim \mathcal{N}(m, \sigma^2)$ . La variable aléatoire  $X$  vérifie alors

$$\begin{cases} \mathbb{E}(X) = m. \\ \text{Var}(X) = \sigma^2. \\ \varphi_X(t) = \exp\left(itm - \frac{\sigma^2 t^2}{2}\right). \end{cases}$$

Plus généralement, pour tout  $\lambda$  dans  $\mathbb{C}$ ,

$$\mathbb{E}(\exp(\lambda X)) = \exp\left(\lambda m + \frac{\lambda^2 \sigma^2}{2}\right).$$

Si  $m = 0$  et  $\sigma^2 = 1$ , on dit que la loi est normale centrée et réduite. Si  $X \sim \mathcal{N}(m, \sigma^2)$  avec  $\sigma > 0$ ,  $\frac{X-m}{\sigma}$  est normale centrée réduite.

Il est à noter que la fonction caractéristique de  $X$  ne dépend que de l'espérance et de la variance de  $X$ . Nous reviendrons plus en détail sur les variables gaussiennes.

## Loi exponentielle

Soit  $\theta > 0$  et

$$X \sim \theta \exp(-\theta x) 1_{x>0} dx.$$

La loi de  $X$  est par définition la **loi exponentielle de paramètre  $\theta$**  : elle vérifie

$$\begin{cases} \mathbb{E}(X) = \frac{1}{\theta}. \\ \varphi_X(t) = \frac{\theta}{\theta - it}. \end{cases}$$

La fonction de répartition  $F_\theta$  d'une loi exponentielle de paramètre  $\theta$  se calcule aisément. On a

$$F_\theta(x) = 1_{x \geq 0} \int_0^x \theta \exp(-\theta u) du = 1_{x \geq 0} (1 - \exp(-\theta x)).$$

On voit qu'alors pour tout  $y \in [0, 1)$ ,

$$F^{-1}(y) = -\frac{1}{\theta} \log(1 - y).$$

D'après la proposition 1.2.3 si  $U$  suit une loi uniforme sur  $[0, 1]$ ,  $-\frac{1}{\theta} \log(1 - U)$  suit une loi exponentielle de paramètre  $\theta$ . Comme  $U$  et  $1 - U$  ont même loi,

$$-\frac{1}{\theta} \log(U)$$

suit également une loi exponentielle de paramètre  $\theta$ .

On peut remarquer que l'on a aussi

$$P(X \geq x) = 1 - F_\theta(x) = \exp(-\theta x),$$

pour tout  $x \geq 0$ . Si  $x$  et  $y$  sont positifs, on en déduit aisément que

$$P(X \geq x + y | X \geq x) = P(X \geq y).$$

Comme dans le cas discret pour les lois géométriques, on parle de propriété de non vieillissement et on utilise donc parfois les lois exponentielles pour modéliser des durées de vie.

# Chapitre 2

## Les théorèmes limites fondamentaux

### Introduction

Après avoir introduit dans un premier temps les deux principaux types de convergence de suites de variables aléatoires, nous présenterons des théorèmes fondamentaux de la théorie des probabilités dont les conséquences seront développées dans les chapitres suivants.

La **loi des grands nombres** nous dit que lorsque l'on répète une même expérience aléatoire indépendamment un grand nombre de fois, la **fréquence empirique** d'un événement converge vers la **probabilité théorique** de cet événement.

De manière plus précise, le **théorème central limite** nous montre que l'écart entre la **moyenne empirique** et la **moyenne théorique** suit asymptotiquement une **loi gaussienne** dont l'écart-type diminue comme l'inverse de la racine carrée du nombre d'expériences.

L'**approximation de Poisson** nous montre que la loi de Poisson est une bonne approximation de la loi de la somme d'un grand nombre de variables aléatoires indépendantes de Bernoulli de paramètre petit.

### 2.1 Deux types de convergence

Soit  $(X_n)$  une suite de variables aléatoires réelles et  $X$  une variable aléatoire réelle.

On dit que  $X_n$  converge vers  $X$  et on écrit

$$X_n \xrightarrow[n \rightarrow \infty]{} X,$$

si la probabilité de l'événement

$\{ \text{la suite de nombres réels } (X_n)_{n \in \mathbb{N}} \text{ converge et } \lim_{n \rightarrow \infty} X_n = X \}$

est 1.

**Remarque :** Dans de nombreux ouvrages de mathématiques, on parle, dans ce cas, de *convergence presque sûre*. Il est légitime de se demander pourquoi on parle de convergence "presque sûre", et pas de convergence "sûre" ? C'est dû au fait que l'événement

$\{ \text{la suite de nombres réels } (X_n)_{n \in \mathbb{N}} \text{ ne converge pas vers } X \}$

a une probabilité nulle, mais n'est pas vide. Cet événement n'arrive jamais, mais il est envisageable, formulable. On parle de convergence "presque sûre" plutôt que de convergence "sûre" par "égard" pour cet événement, qui, bien que de probabilité nulle, existe. De la même façon, si  $U$  est une v.a ; uniforme sur  $[0, 1]$ , l'événement

$$\{U = 1/2\}$$

a une probabilité nulle, mais existe quand même. Ces considérations ne sont pas dénuées d'intérêt lorsque l'on utilise la théorie des probabilités de manière plus approfondie, mais elles n'apportent rien à ce cours.

La *convergence en loi* est assez différente, du point de vue conceptuel, de la convergence précédente. Considérons une suite de variables aléatoires  $(X_n)_{n \geq 0}$ . On dit que  $X_n$  converge en loi vers  $X$ , et on note

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X$$

si et seulement si pour tout  $x \in \mathbb{R}$  tel que  $P(X = x) = 0$ ,

$$P(X_n \leq x) \xrightarrow[n \rightarrow \infty]{} P(X \leq x).$$

On peut noter que si

$$X_n \xrightarrow[n \rightarrow \infty]{\text{loi}} X,$$

et si  $a \leq b$  vérifient  $P(X = a) = P(X = b) = 0$ , alors

$$P(a \leq X_n \leq b) \xrightarrow[n \rightarrow \infty]{} P(a \leq X \leq b).$$

Nous admettrons que la convergence implique la convergence en loi. La réciproque est fautive. Essayons de comprendre pourquoi. Dire que  $X_n$  converge vers  $X$  revient à dire que quand  $n$  est grand, les deux nombres  $X_n$  et  $X$ , qui dépendent du hasard, seront proches. On aura  $X_n \simeq X$ . Il est donc naturel que

$$P(a \leq X_n \leq b) \simeq P(a \leq X \leq b).$$

Par contre, de la même manière que le fait que deux v.a. aient la même loi n'implique pas qu'elles soient égales, le fait que

$$P(a \leq X_n \leq b) \simeq P(a \leq X \leq b)$$

n'implique pas que  $X_n \simeq X$ . Reprenons l'exemple d'une v.a.  $X$  donnée par la face tournée vers le haut d'un dé qu'on jette. Considérons la v.a.  $Y$  de la face tournée vers le bas du même dé lors du même lancé.  $Y$  et  $X$  ont la même loi, donc en posant : pour tout  $n$ ,  $X_n = Y$ , on a

$$\forall n \in \mathbb{N}, P(a \leq X_n \leq b) = P(a \leq X \leq b).$$

Pourtant,  $X_n$  et  $X$  sont toujours deux nombres distants de au moins 1.

Le théorème suivant montre que la convergence en loi se caractérise aussi avec les transformées de Fourier.

**Théorème 2.1.1 (Théorème de Paul Lévy)**  $X_n$  converge en loi vers  $X$  si et seulement si pour tout  $t \in \mathbb{R}$ ,

$$\varphi_{X_n}(t) \xrightarrow[n \rightarrow \infty]{} \varphi_X(t).$$

## 2.2 Loi des grands nombres

Commençons en énonçant la loi des grands nombres (LGN) sous sa forme la plus générale.

**Théorème 2.2.1 (Loi forte des grands nombres)** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires i.i.d.. Posons

$$S_n = X_1 + \dots + X_n,$$

et  $m = \mathbb{E}(X_1)$  ( $= \mathbb{E}(X_n)$  pour tout  $n$ , puisque les  $X_n$  sont identiquement distribuées), Alors

$$\frac{S_n}{n} \xrightarrow[n \rightarrow \infty]{} m.$$

Ce théorème affirme que si on veut calculer l'espérance d'une v.a.  $X$ , il suffit de prendre un grand nombre de v.a.  $X_1, \dots, X_n$  indépendantes et ayant toutes la loi de  $X$ , et de faire la moyenne

$$\frac{X_1 + \dots + X_n}{n}$$

des valeurs données par ces v.a.. On a alors

$$\mathbb{E}(X) \simeq \frac{X_1 + \dots + X_n}{n}.$$

Nous allons maintenant donner le corollaire de ce théorème qui relie la fréquence de réalisation d'un événement et sa probabilité.

**Corollaire 2.2.2 (Loi empirique des grands nombres)** Soit  $(A_n)_{n \geq 1}$  une suite d'événements indép. de même probabilité  $p$ . Alors en définissant, pour  $n \geq 1$ , la v.a.

$$N_n = \text{Card}(\{k \in \{1, \dots, n\} / A_k \text{ est réalisé}\}),$$

on a  $\lim_{n \rightarrow \infty} \frac{N_n}{n} = p$ .

Ce corollaire affirme que si on veut calculer la probabilité d'un événement  $A$ , il suffit de considérer un grand nombre  $A_1, \dots, A_n$  d'événements indépendants et ayant tous la même probabilité que  $A$ , et de compter le nombre  $N_n$  de ces événements qui sont en fait réalisés. On a alors

$$P(A) \simeq \frac{N_n}{n}.$$

*Preuve :* Il suffit de définir, pour tout  $n$ , la v.a.  $X_n$ , qui vaut 1 si  $A_n$  est réalisé, et 0 sinon. Les  $A_n$  étant indép., les  $X_n$  le sont aussi, de plus, pour tout  $n$ , la loi de  $X_n$  est une loi de Bernoulli de paramètre  $p$ , donc on a

$$\frac{N_n}{n} = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(X_1) = p.$$

□

**Application 2.2.3** Considérons un dé pipé, qui donne aux faces  $1, \dots, 6$  les probabilités respectives  $p_1, \dots, p_6$  de tomber. Alors en lançant le dé un grand nombre  $n$  de fois, en notant respectivement  $C_1, \dots, C_6$  le nombre de fois qu'il est tombé sur  $1, \dots, 6$ , on a  $(p_1, \dots, p_6) \simeq (\frac{C_1}{n}, \dots, \frac{C_6}{n})$ .

**Remarque :** Le Th. Central Limite nous dira que si on le lance  $n$  fois, la précision de l'estimation  $(p_1, \dots, p_6) \simeq (\frac{C_1}{n}, \dots, \frac{C_6}{n})$  sera de l'ordre de  $\frac{1}{\sqrt{n}}$ . Par exemple, si on veut estimer  $p_1, \dots, p_6$  avec une précision de variance 0, 1, il faudra le lancer 100 fois.

## 2.3 Le théorème central limite

La loi des grands nombres nous dit qu'en présence d'une suite  $(X_n)_{n \geq 1}$  de variables aléatoires i.i.d., la **moyenne empirique**  $\frac{S_n}{n}$  converge vers la **moyenne théorique**  $\mathbb{E}(X_1)$ . Cependant, nous ne savons rien des fluctuations autour de cette moyenne théorique, autrement dit de la vitesse de convergence de la moyenne empirique en fonction de  $n$ .

**Théorème 2.3.1 (Théorème central limite)** Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires i.i.d. Notons  $m = \mathbb{E}(X_1)$  et  $\sigma^2 = \text{Var}(X_1)$ .

Alors,

$$\sqrt{n} \left( \frac{S_n}{n} - m \right) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, \sigma^2)$$

en loi.  $\mathcal{N}(0, \sigma^2)$  désigne une gaussienne centrée de variance  $\sigma^2$ .

### Remarques :

- On écrit parfois le théorème central limite sous la forme

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1).$$

- Le théorème central limite nous donne une sorte de développement limité de la moyenne empirique lorsque  $n$  est grand. Autrement dit, pour  $n$  grand, on a

$$\frac{S_n}{n} \stackrel{\text{loi}}{\simeq} m + \frac{\sigma}{\sqrt{n}} G$$

où  $G$  est une gaussienne centrée réduite. La convergence dans la loi des grands nombres est donc pour ainsi dire en  $\frac{\sigma}{\sqrt{n}}$  : Les fluctuations autour de la moyenne théorique au bout de  $n$  expériences sont de l'ordre de  $\frac{\sigma}{\sqrt{n}}$ .

- Il est assez remarquable de noter que les gaussiennes apparaissent à la limite quelle que soit la loi commune des  $X_i$ , pourvu que les  $X_i$  soient de carré intégrable. Ceci nous montre le caractère **universel** des lois normales.

- En pratique, lorsque  $n \geq 30$ , l'approximation “ $\frac{\sqrt{n}}{\sigma} \left( \frac{S_n}{n} - m \right)$  est gaussienne centrée réduite” est bonne. Nous admettons ce lemme, à partir duquel nous démontrerons le théorème.

**Lemme 2.3.2** Soit  $X$  une variable aléatoire de carré intégrable de fonction caractéristique  $\varphi$ . Lorsque  $t \rightarrow 0$ , nous avons

$$\varphi(t) = 1 + it\mathbb{E}(X) - \frac{t^2}{2}\mathbb{E}(X^2) + o(t^2),$$

*c'est-à-dire*

$$\frac{\varphi(t) - 1 - it\mathbb{E}(X) + \frac{t^2}{2}\mathbb{E}(X^2)}{t^2} \xrightarrow[t \rightarrow 0, t \neq 0]{} 0.$$

Donnons maintenant une preuve du TCL :

*Preuve :* Quitte à remplacer  $X_i$  par  $\frac{X_i - m}{\sigma}$ , nous supposons que  $m = 0$  et  $\sigma = 1$ . Il s'agit de considérer la fonction caractéristique  $\varphi_n$  de  $\frac{S_n}{\sqrt{n}}$ . On a

$$\varphi_n(t) = \mathbb{E} \left( \exp \left( it \frac{S_n}{\sqrt{n}} \right) \right) = \mathbb{E} \left( \exp \left( \frac{itX_1}{\sqrt{n}} + \dots + \frac{itX_n}{\sqrt{n}} \right) \right).$$

Comme les  $X_i$  sont indépendantes et de même loi,

$$\varphi_n(t) = \mathbb{E} \left( \exp \left( \frac{itX_1}{\sqrt{n}} \right) \right) \dots \mathbb{E} \left( \exp \left( \frac{itX_n}{\sqrt{n}} \right) \right) = \mathbb{E} \left( \exp \left( \frac{itX_1}{\sqrt{n}} \right) \right)^n.$$

Soit  $\varphi$  la fonction caractéristique de  $X_1$ . Nous avons donc obtenu

$$\varphi_n(t) = \varphi \left( \frac{t}{\sqrt{n}} \right)^n.$$

L'idée est de remarquer que pour des grandes valeurs de  $n$ , d'après le théorème précédent,

$$\varphi \left( \frac{t}{\sqrt{n}} \right) = 1 + i \frac{t}{\sqrt{n}} \mathbb{E}(X) - \frac{t^2}{2n} \mathbb{E}(X^2) + o\left(\frac{1}{n}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right).$$

Donnons une idée de la fin de la démonstration. On remarque que lorsque  $n$  tend vers l'infini,

$$\varphi_n(t) = \varphi \left( \frac{t}{\sqrt{n}} \right)^{\frac{1}{n}} = \exp \left( n \log \varphi \left( \frac{t}{\sqrt{n}} \right) \right).$$

Par conséquent,

$$\varphi_n(t) = \exp \left( n \log \left( 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right) \right) = \exp \left( -n \left( \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right) \right) = \exp \left( -\frac{t^2}{2} + o(1) \right).$$



Ainsi,

$$\varphi_n(t) \xrightarrow{n \rightarrow \infty} \exp\left(-\frac{t^2}{2}\right).$$

On a vu que

$$t \mapsto \exp\left(-\frac{t^2}{2}\right)$$

est la fonction caractéristique d'une loi normale centrée réduite. D'après le théorème de Paul Lévy sur la convergence en loi, on a obtenu

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1).$$

□

**Exercice 2.3.1** On lance un dé non pipé 100 fois, de façon bien entendu indépendante. Quelle est la probabilité que la somme totale des points obtenus soit entre 300 et 400 ?

*Solution* : La somme totale est

$$S = \sum_{i=1}^{100} X_i,$$

où  $X_i$  représente le nombre de points obtenus. L'espérance de  $X_i$  est 3,5, et sa variance est 35/12. Par le TCL,  $\frac{S-350}{10\sqrt{35/12}}$  suit approximativement une loi  $\mathcal{N}(0, 1)$ .

$$\begin{aligned} \text{P}(300 \leq S \leq 400) &= \text{P}\left(\frac{-5}{\sqrt{35/12}} \leq \frac{S-350}{10\sqrt{35/12}} \leq \frac{5}{\sqrt{35/12}}\right) \\ &= \text{P}(-2,93 \leq \mathcal{N}(0, 1) \leq 2,93) \\ &= 0,9966 \end{aligned}$$

## 2.4 L'approximation de Poisson

Nous terminons ce chapitre avec un théorème limite assez fondamental. Nous rappelons que la loi binômiale  $\mathcal{B}(n, p)$  est la loi de la somme de  $n$  variables aléatoires indépendantes de Bernoulli de paramètre  $p$ . Autrement dit, pour  $0 \leq k \leq n$ ,

$$\mathcal{B}(n, p)(\{k\}) = C_n^k p^k (1-p)^{n-k}.$$

La loi de Poisson  $\mathcal{P}(\lambda)$  est la loi sur l'ensemble des entiers  $\mathbb{N}$  telle que pour tout  $k \geq 0$ ,

$$\mathcal{P}(\lambda)(\{k\}) = \exp(-\lambda) \frac{\lambda^k}{k!}.$$

**Théorème 2.4.1 (Approximation de Poisson)** *Supposons que  $(p_n)$  soit une suite de réels compris entre 0 et 1 et  $\lambda$  un réel tels que*

$$np_n \xrightarrow{n \rightarrow \infty} \lambda,$$

*alors, on a la convergence en loi*

$$\mathcal{B}(n, p_n) \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{P}(\lambda).$$

**Remarque** : Lorsqu'on est en présence d'une loi binômiale  $\mathcal{B}(n, p)$  avec  $p$  petit et  $n$  grand et si le produit  $np$  est de l'ordre de 1, il est raisonnable d'approcher  $\mathcal{B}(n, p)$  par une loi de Poisson de paramètre  $np$ , ce qui simplifie parfois considérablement les calculs.

*Preuve* : Comme d'habitude, on regarde les fonctions caractéristiques. Soit  $X_{1,n}, \dots, X_{n,n}$   $n$  variables aléatoires de Bernoulli indépendantes de paramètre  $p_n$ . Notons

$$X_n = X_{1,n} + \dots + X_{n,n}.$$

La fonction caractéristique de  $X_n$  est par indépendance

$$\begin{aligned} \varphi_{X_n}(t) &= \mathbb{E}(\exp(itX_n)) = \mathbb{E}(\exp(itX_{1,n})) \dots \mathbb{E}(\exp(itX_{n,n})) \\ &= ((1 - p_n) + p_n \exp(it))^n = (1 + p_n(\exp(it) - 1))^n. \end{aligned}$$

Comme  $np_n \xrightarrow[n \rightarrow \infty]{} \lambda$ , on a pour  $n$  grand,

$$\mathbb{E}(\exp(itX_n)) \simeq \left(1 + \frac{\lambda}{n}(\exp(it) - 1)\right)^n.$$

Un lemme classique d'analyse nous dit que pour tout nombre complexe  $z$ , pour toute suite  $u_n$  telle que  $\lim nu_n = z$ ,

$$(1 + u_n)^n \xrightarrow[n \rightarrow \infty]{} \exp(z).$$

Par conséquent,

$$\mathbb{E}(\exp(itX_n)) \xrightarrow[n \rightarrow \infty]{} \exp(\lambda(\exp(it) - 1)).$$

Nous avons vu que le membre de droite de la dernière égalité était égal à la valeur en  $t$  de la fonction caractéristique d'une loi de Poisson de paramètre  $\lambda$ . D'où le résultat escompté.  $\square$

Voici quelques exemples d'événements dont on peut approcher la loi par une loi de Poisson d'après le théorème précédent :

- le nombre de défauts de production dans un processus industriel.
- le nombre de fautes d'impression par page.
- le nombre de clients à une caisse par tranche de 5 minutes.
- le nombre de sinistres par jour dans une compagnie d'assurance.

...

# Chapitre 3

## Vecteurs gaussiens

### Introduction

Les variables gaussiennes sont des lois universelles qui sont d'une importance particulière. Il sera très utile en particulier en statistique de pouvoir passer du cas de la dimension 1 à des dimensions plus grandes.

Ce passage nécessite de développer un formalisme qui sera l'objet du début du chapitre. Nous calculerons la fonction caractéristique des **vecteurs gaussiens** que nous aurons préalablement définis. Nous montrerons aussi comment nous pouvons alors aborder de manière très efficace et très simple les problèmes d'**indépendance de variables gaussiennes**.

Nous énoncerons ensuite le **théorème central limite vectoriel** qui généralise le théorème central limite classique.

Nous terminerons par l'introduction des **lois du chi-deux** qui apparaissent en présence de carrés de variables gaussiennes.

### 3.1 Variables aléatoires gaussiennes

Nous avons déjà introduit les lois gaussiennes. Les v.a. gaussiennes sont les v.a. dont la loi est une loi gaussienne.

**Définition 3.1.1 (Variables gaussiennes - Lois normales)** Soit  $m \in \mathbb{R}$  et  $\sigma > 0$ . On dit qu'une variable aléatoire  $X$  est une gaussienne d'espérance  $m$  et de variance  $\sigma^2$  ou encore que  $X$  suit une loi normale d'espérance  $m$  et de variance  $\sigma^2$  si et seulement si  $X$  a pour densité

$$\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \frac{dx}{\sigma\sqrt{2\pi}},$$

autrement dit, si et seulement si pour tout  $a \in \mathbb{R}$ ,

$$P(X \leq a) = \int_{-\infty}^a \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \frac{dx}{\sigma\sqrt{2\pi}}.$$

On note  $\mathcal{N}(m, \sigma^2)$  cette loi. On note  $\mathcal{N}(m, 0)$  la loi d'une variable aléatoire constante égale à  $m$ . Lorsque  $X$  suit une loi  $\mathcal{N}(m, \sigma^2)$ ,  $\sigma \geq 0$ , on a

$$\mathbb{E}(X) = m, \quad \text{Var}(X) = \sigma^2.$$

Lorsque  $m = 0$ , on dit que  $X$  est centrée. Lorsque  $m = 0$  et  $\sigma = 1$ , on dit que  $X$  est une gaussienne centrée réduite (ou suit une loi normale centrée réduite).

**Remarques :**

- On peut noter que l'identité

$$\int_{-\infty}^{+\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) \frac{dx}{\sigma\sqrt{2\pi}} = 1$$

n'est pas évidente a priori. Nous l'admettons.

- Le calcul de l'espérance et de la variance à partir de la densité se font par intégrations par parties.

- Un point très important est que la loi d'une variable gaussienne ne dépend que de son espérance et de sa variance.

- Lorsque  $X \sim \mathcal{N}(m, \sigma^2)$  avec  $\sigma > 0$ , la variable aléatoire

$$\frac{X - m}{\sigma}$$

est centrée réduite.

Voici quelques propriétés des variables aléatoires gaussiennes.

**Proposition 3.1.2** Soit  $m \in \mathbb{R}$ ,  $\sigma \geq 0$  et soit une variable aléatoire

$$X \sim \mathcal{N}(m, \sigma^2).$$

(1) Pour tout  $\lambda \in \mathbb{C}$ ,

$$\mathbb{E}(\exp(\lambda X)) = \exp\left(\lambda m + \frac{\sigma^2 \lambda^2}{2}\right).$$

(2) Pour tout  $t \in \mathbb{R}$ ,

$$\mathbb{E}(\exp(itX)) = \exp\left(itm - \frac{\sigma^2 t^2}{2}\right).$$

(3) Si  $X$  est centrée réduite, alors pour tout  $n \geq 1$ ,

$$\mathbb{E}(X^{2n-1}) = 0, \quad \mathbb{E}(X^{2n}) = \frac{(2n)!}{2^n (n!)^2}.$$

**Théorème 3.1.3 (Combinaisons linéaires de v.a. gaussiennes indépendantes)** Soient  $X_1, \dots, X_n$  des v.a. gaussiennes d'espérances respectives  $m_1, \dots, m_n$  et de variances respectives  $\sigma_1^2, \dots, \sigma_n^2$ . On suppose ces v.a. indépendantes. Soient  $a_1, \dots, a_n \in \mathbb{R}$ .

Alors la v.a.  $a_1 X_1 + \dots + a_n X_n$  est gaussienne de moyenne et variance

$$a_1 m_1 + \dots + a_n m_n, \quad a_1^2 \sigma_1^2 + \dots + a_n^2 \sigma_n^2.$$

**Exercice 3.1.1** 1. Montrer le théorème précédent en utilisant la transformée de Fourier.

2. Trouver un contre-exemple dans le cas où les  $X_i$  ne sont pas indépendants. **Indication :** on pourra prendre  $n = 2$ ,  $X_1$  de loi normale centrée réduite, et  $X_2 = -X_1$ .

## 3.2 Les vecteurs gaussiens

### 3.2.1 Définition

Les v.a. gaussiennes apparaissent fréquemment dans les modélisations de phénomènes aléatoires. Il arrive que l'on ait à travailler avec une famille finie de v.a. (qui constituent donc un *vecteur aléatoire*) qui soient toutes gaussiennes et telles que leurs combinaisons linéaires soient aussi gaussiennes. C'est ce qui nous amène à donner la définition suivante :

**Définition 3.2.1 (Vecteur gaussien)** Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ . On dit que  $X$  est un vecteur gaussien si et seulement si pour tout  $(u_1, \dots, u_n) \in \mathbb{R}^n$ ,

$$\langle u, X \rangle := u_1 X_1 + \dots + u_n X_n$$

est une variable aléatoire gaussienne.

**Remarque :** Si  $X$  est un vecteur gaussien, alors pour tout  $0 \leq i \leq n$ , la variable aléatoire  $X_i$  correspondant à la  $i$ -ème coordonnée de  $X$  est une variable aléatoire gaussienne. Mais, il faut bien faire attention au fait qu'inversement, pour qu'un vecteur soit gaussien, il ne suffit pas que chacune des coordonnées soit une variable gaussienne. Il faut et il suffit que **toute combinaison linéaire des coordonnées** de  $X$  suive une loi normale.

**Exercice 3.2.1** Soit  $r \in \mathbb{N}^*$ ,  $n_1, \dots, n_r \in \mathbb{N}^*$ . Soient  $Y_1, \dots, Y_r$  des vecteurs aléatoires gaussiens **indépendants** à valeurs dans respectivement  $\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_r}$ .

Montrer que le vecteur aléatoire  $(Y_1, \dots, Y_r)$ , à valeurs dans  $\mathbb{R}^{n_1 + \dots + n_r}$ , est gaussien.

### 3.2.2 Loi d'un vecteur aléatoire

On rappelle la définition :

**Définition 3.2.2** Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ . On appelle loi de  $X$  la fonction

$$\begin{aligned} Q : \mathcal{P}(\mathbb{R}^n) &\rightarrow [0, 1] \\ A &\mapsto P(X \in A) \end{aligned}$$

On note souvent  $\mathcal{L}(X)$  pour  $Q$ .

**Exercice 3.2.2** Soient  $X, Y$  des vecteurs aléatoires à valeurs dans  $\mathbb{R}^n$ , de même loi, et soit  $f$  une fonction de  $\mathbb{R}^n$  vers  $\mathbb{R}$ .

Montrer que  $f(X)$  et  $f(Y)$  ont même loi.

Le résultat suivant est une sorte de réciproque de cet exercice :

**Théorème 3.2.3** Soient  $X, Y$  des vecteurs aléatoires à valeurs dans  $\mathbb{R}^n$  tels que pour tout  $v \in \mathbb{R}^n$ , les v.a.  $\langle v, X \rangle$  et  $\langle v, Y \rangle$  ont même loi. Alors  $X$  et  $Y$  ont même loi.

**Corollaire 3.2.4 (Transfo. de Fourier et loi des vecteurs aléatoires)** Soient  $X, Y$  des vecteurs aléatoires à valeurs dans  $\mathbb{R}^n$  tels que pour tout  $v \in \mathbb{R}^n$ ,

$$\mathbb{E}(\exp(i \langle v, X \rangle)) = \mathbb{E}(\exp(i \langle v, Y \rangle)).$$

Alors  $X$  et  $Y$  ont même loi.

### 3.2.3 Transformée de Fourier d'un vecteur gaussien

Le corollaire précédent montre l'intérêt du calcul de la tr. de Fourier pour déterminer la loi d'un vecteur aléatoire.

**Définition 3.2.5** Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$  dont toutes les coordonnées sont intégrables. On appelle **espérance** de  $X$  le vecteur de  $\mathbb{R}^n$   $\mathbb{E}(X) = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))$ .

Rappelons la définition et la propriété principale de la matrice de covariance :

**Définition 3.2.6** Supposons que  $X = (X_1, \dots, X_n)$  soit un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ . On appelle **matrice de covariance** de  $X$  la matrice carrée  $(n, n)$  notée

$$\text{Cov}(X) = \Sigma(X) = (\Sigma_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

telle que pour tout  $1 \leq i, j \leq n$ ,

$$\Sigma_{i,j} = \text{Cov}(X_i, X_j).$$

**Définition 3.2.7** Soit

$$A = (a_{i,j})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

une matrice symétrique. On appelle forme quadratique associée à  $A$  l'application  $q_A$ , de  $\mathbb{R}^n$  vers  $\mathbb{R}$ , définie par

$$q_A(x) = \sum_{i,j=1}^n a_{i,j} x_i x_j.$$

**Remarque :** Une matrice de covariance  $\Sigma$  est toujours symétrique et positive : ça signifie que la forme quadratique associée  $q_\Sigma$  est une fonction positive sur  $\mathbb{R}^n$ .

**Proposition 3.2.8** Soit  $X = (X_1, \dots, X_n)$  soit un vecteur aléatoire à valeurs dans  $\mathbb{R}^n$ , de carré intégrable, et de matrice de covariance  $A$ . Alors pour tout vecteur (non aléatoire)  $v \in \mathbb{R}^n$ , la variance de la v.a. réelle  $\langle v, X \rangle$  est  $q_A(v)$ , où  $q_A$  est la forme quadratique associée à  $A$ .

**Théorème 3.2.9 (Transformée de Fourier d'un vecteur gaussien)** Soit  $X = (X_1, \dots, X_n)$  un vecteur aléatoire gaussien à valeurs dans  $\mathbb{R}^n$ . On pose  $m = \mathbb{E}(X) \in \mathbb{R}^n$  et  $A = \text{Cov}(X)$ . Alors pour tout  $v \in \mathbb{R}^n$ , on a :

$$\mathbb{E}(\exp(i \langle v, X \rangle)) = \exp(i \langle v, m \rangle - \frac{1}{2} q_A(v)).$$

On avait défini un vecteur aléatoire gaussien en disant que c'était un vecteur aléatoire  $X = (X_1, \dots, X_n)$  tel que pour tout  $v \in \mathbb{R}^n$ , il existe  $m_v \in \mathbb{R}$ ,  $\sigma_v^2 \geq 0$  tels que la v.a.  $v_1 X_1 + \dots + v_n X_n = \langle v, X \rangle$  a pour loi  $\mathcal{N}(m_v, \sigma_v^2)$ . On sait maintenant exprimer  $m_v$  et  $\sigma_v^2$  :

**Corollaire 3.2.10** Pour tout  $v \in \mathbb{R}^n$ , la loi de  $v_1 X_1 + \dots + v_n X_n$  est la loi normale d'espérance  $v_1 \mathbb{E}(X_1) + \dots + v_n \mathbb{E}(X_n)$  et de variance  $q_A(v)$ .

**Corollaire 3.2.11** La loi d'un vecteur aléatoire gaussien est entièrement déterminée par son espérance et sa matrice de covariance.

**Remarque :** La loi d'un vecteur gaussien n'est pas déterminée par les lois de ses coordonnées :  
 - soit  $N$  une v.a. de loi  $\mathcal{N}(0, 1)$ , et soit  $Y$  le vecteur aléatoire à valeurs dans  $\mathbb{R}^n$  def. par

$$Y = (\underbrace{N, N, \dots, N}_{n \text{ fois}})$$

- soient  $X_1, \dots, X_n$  des v.a.i.i.d. de loi  $\mathcal{N}(0, 1)$  et soit  $Z$  le vecteur aléatoire à valeurs dans  $\mathbb{R}^n$  def. par

$$Z = (X_1, X_2, \dots, X_n)$$

La loi de toute coordonnée de  $Y$  ou de  $Z$  est  $\mathcal{N}(0, 1)$  alors que  $Y$  et  $Z$  n'ont pas la même loi (exercice : le démontrer).

Ceci justifie la définition suivante :

**Définition 3.2.12** - Soit  $m \in \mathbb{R}^n$  et soit  $\Sigma \in \mathcal{M}_n(\mathbb{R})$  une matrice symétrique positive. On désigne par  $\mathcal{N}(m, \Sigma)$  la loi d'un vecteur gaussien d'espérance  $m$  et de matrice de covariance  $\Sigma$ .

- Supposons que loi( $X$ ) =  $\mathcal{N}(m, \Sigma)$ . Lorsque  $m = 0$ , on dit que  $X$  est centré. Lorsque  $m = 0$  et  $\Sigma = Id$ , on dit que  $X$  est centré réduit.

### 3.2.4 Variables aléatoires gaussiennes indépendantes

Les variables aléatoires gaussiennes nous donnent des exemples de vecteurs gaussiens. Inversement, le formalisme introduit plus haut permet d'avoir des résultats très simples sur l'indépendance de gaussiennes.

Nous énonçons une première proposition simple.

**Proposition 3.2.13** *Soient  $X_1, \dots, X_n$  des variables aléatoires gaussiennes indépendantes telles que*

$$\mathbb{E}(X_1) = m_1, \dots, \mathbb{E}(X_n) = m_n \text{ et } \text{Var}(X_1) = \sigma_1^2, \dots, \text{Var}(X_n) = \sigma_n^2.$$

*Alors le vecteur aléatoire  $X = (X_1, \dots, X_n)$  est un vecteur gaussien d'espérance  $m = (m_1, \dots, m_n)$  et de matrice de covariance la matrice diagonale*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_n^2 \end{pmatrix}.$$

**Remarque :** Si  $X_1, \dots, X_n$  sont  $n$  v.a. gaussiennes i.i.d. et si  $X = (X_1, \dots, X_n)$ , alors

$$X \sim \mathcal{N}(\mathbb{E}(X), \sigma^2 Id)$$

où  $Id$  est la matrice identité et  $\sigma^2 = \text{Var}(X_1) = \dots = \text{Var}(X_n)$  et  $\mathbb{E}(X) = (m, \dots, m)$  où  $m = \mathbb{E}(X_1) = \dots = \mathbb{E}(X_n)$ .

*Preuve :* On a vu au Th. 3.1.3 que pour tout  $u \in \mathbb{R}^n$ ,  $\langle u, X \rangle$  est une v.a. gaussienne. Donc  $X$  est un vecteur gaussien, et sa matrice est diagonale car si  $i \neq j$ , son entrée d'indice  $i, j$  est la covariance de  $X_i$  et  $X_j$ , qui est nulle par indépendance de  $X_i$  et  $X_j$ .  $\square$

On peut énoncer et démontrer la réciproque de cette proposition.

**Proposition 3.2.14** *Soient  $X_1, \dots, X_n$   $n$  variables aléatoires telles que le vecteur aléatoire  $X = (X_1, \dots, X_n)$  soit un vecteur gaussien. Si la matrice de covariance de  $X$  est diagonale, alors les variables  $X_1, \dots, X_n$  sont indépendantes.*

**Remarque :** Il faut surtout noter que l'hypothèse de départ n'est pas seulement que toutes les v.a. sont des gaussiennes mais que toutes leurs combinaisons linéaires sont des gaussiennes, ce qui est beaucoup plus fort.

*Preuve :* Pour tout  $t = (t_1, \dots, t_n) \in \mathbb{R}^n$ , on note

$$\phi_X(t) = \mathbb{E}(\exp(it \cdot X)) = \mathbb{E}(\exp(it_1 X_1 + \dots + it_n X_n)).$$

Notons également pour tout  $1 \leq i \leq n$  et tout  $u \in \mathbb{R}$ ,

$$\phi_{X_i}(u) = \mathbb{E}(\exp(iu X_i)).$$

Rappelons que pour que  $X_1, \dots, X_n$  soient indépendantes, il faut et il suffit que l'on ait pour tout  $t = (t_1, \dots, t_n) \in \mathbb{R}^n$ ,

$$\phi_X(t_1, \dots, t_n) = \phi_{X_1}(t_1) \dots \phi_{X_n}(t_n).$$

On note  $\Sigma$  la matrice de covariance de  $X$ . Par hypothèse,  $\Sigma$  est diagonale, ce qui s'écrit

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_n^2 \end{pmatrix}$$

où

$$\sigma_1^2 = \text{Var}(X_1), \dots, \sigma_n^2 = \text{Var}(X_n).$$

Posons par ailleurs  $m = (m_1, \dots, m_n) = \mathbb{E}(X)$ . Comme  $X$  est supposé être un vecteur gaussien nous pouvons dire grâce à la proposition 3.2.9 que l'on a

$$\mathbb{E}(\exp(it.X)) = \exp(it.m - \frac{1}{2}q_\Sigma(t)).$$

Or,

$$q_\Sigma(t) = \sigma_1^2 t_1^2 + \dots + \sigma_n^2 t_n^2,$$

de sorte que

$$\begin{aligned} \phi_X(t_1, \dots, t_n) = \mathbb{E}(\exp(it.X)) &= \exp(it_1 m_1 + \dots + it_n m_n - \frac{\sigma_1^2 t_1^2}{2} - \dots - \frac{\sigma_n^2 t_n^2}{2}) \\ &= \exp(it_1 - \frac{\sigma_1^2 t_1^2}{2}) \dots \exp(it_n - \frac{\sigma_n^2 t_n^2}{2}) \\ &= \phi_{X_1}(t_1) \dots \phi_{X_n}(t_n) \end{aligned}$$

où on a utilisé la proposition 3.1.2, (2). Par conséquent,  $X_1, \dots, X_n$  sont indépendantes.  $\square$

Pour résumer les deux propositions précédentes, énonçons un théorème qui les englobe.

**Théorème 3.2.15** Soient  $X_1, \dots, X_n$   $n$  variables aléatoires telles que le vecteur  $X = (X_1, \dots, X_n)$  soit gaussien d'espérance  $\mathbb{E}(X)$  et de matrice de covariance  $\Sigma(X)$ .

- (1) Les variables  $X_1, \dots, X_n$  sont indépendantes si et seulement si la matrice  $\Sigma(X)$  est diagonale.  
(2) Les variables  $X_1, \dots, X_n$  sont i.i.d. si et seulement si existent des réels  $m$  et  $\sigma$  tels que

$$\mathbb{E}(X) = (m, \dots, m) \text{ et } \Sigma(X) = \sigma^2 Id.$$

*Preuve :* La partie (1) a déjà été démontrée. Pour (2), on suppose d'abord que les variables  $X_1, \dots, X_n$  sont i.i.d. D'après (1),  $\Sigma(X)$  est diagonale. Comme les variables  $X_1, \dots, X_n$  ont par ailleurs même loi, on a

$$\mathbb{E}(X_1) = \dots = \mathbb{E}(X_n) = m$$

et

$$\text{Var}(X_1) = \dots = \text{Var}(X_n) = \sigma^2$$

de sorte que

$$\mathbb{E}(X) = (m, \dots, m) \text{ et } \Sigma(X) = \sigma^2 Id.$$

Si à présent, on suppose que

$$\mathbb{E}(X) = (m, \dots, m) \text{ et } \Sigma(X) = \sigma^2 Id,$$

en particulier,  $\Sigma(X)$  est diagonale donc on sait d'après (1) que les variables  $X_1, \dots, X_n$  sont indépendantes. D'autre part, toutes les variables  $X_i$  sont gaussiennes de même espérance  $m$  et de même variance  $\sigma^2$ . D'après la proposition 3.2.9, les variables  $X_1, \dots, X_n$  ont donc toute la même loi  $\mathcal{N}(m, \sigma^2)$ .  $\square$

**Exercice 3.2.3** Soit

$$A = [a_{i,j}]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

la matrice de covariance d'un vecteur aléatoire. Montrer, en utilisant l'inégalité de Cauchy-Schwartz, que

$$\forall i, j, |a_{i,j}| \leq \sqrt{a_{i,i}} \sqrt{a_{j,j}}. \quad (3.1)$$

**Exercice 3.2.4** Soit  $Z$  une variable aléatoire réelle de loi normale centrée réduite. Soit  $\sigma_1, \dots, \sigma_n > 0$ . On pose  $X_1 = \sigma_1 Z$ ,  $X_2 = \sigma_2 Z, \dots, X_n = \sigma_n Z$ . Montrer que  $X = (X_1, \dots, X_n)$  est gaussien, d'espérance nulle, et de matrice de covariance la matrice

$$A = [a_{i,j}]_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}}$$

avec pour tout  $i, j$ ,  $a_{i,j} = \sigma_i \sigma_j$ .



**Remarque :** On remarque que pour l'exemple de l'exercice précédent, on a égalité dans l'inégalité (3.1). Ainsi, dans le cas où les  $X_i$  sont proportionnels, qui correspond à une dépendance maximale, les entrées de la matrice de covariance sont maximales compte tenu de la contrainte de l'inégalité (3.1). A l'inverse, dans le cas où les  $X_i$  sont indépendants, la matrice de covariance est diagonale, i.e.  $a_{i,j} = 0$  dès que  $i \neq j$ . Ainsi, quand  $X = (X_1, \dots, X_n)$  est un vecteur gaussien, les rapports

$$\frac{|\text{Cov}(X_i, X_j)|}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}}, \quad (i \neq j)$$

permettent de *quantifier l'indépendance* : plus ils sont élevés, moins  $X_i$  et  $X_j$  sont indépendants.

### 3.2.5 Image d'un vecteur gaussien par une application linéaire

A présent, nous allons énoncer un théorème qui nous dit ce que devient un vecteur gaussien lorsqu'on lui applique une application linéaire.

**Proposition 3.2.16** *Soit  $X = (X_1, \dots, X_n)$  un vecteur gaussien d'espérance  $m$  et de matrice de covariance  $\Sigma$ . Soit également  $A = (a_{i,j})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \in \mathcal{M}_{m,n}(\mathbb{R})$ . Soit  $Y$  le vecteur aléatoire de  $\mathbb{R}^m$  défini par*

$$Y = AX.$$

*Alors,  $Y$  est un vecteur gaussien d'espérance  $Am$  et de matrice de covariance  $A\Sigma A^T$ , où  $A^T$  désigne la transposée de  $A$ . On résume ceci sous la forme*

$$\text{loi}(AX) = \mathcal{N}(A\mathbb{E}(X), A\Sigma(X)A^T).$$

## 3.3 Le théorème central limite vectoriel

Nous avons déjà vu au chapitre 2 que les variables aléatoires gaussiennes apparaissent de manière naturelle dans les théorèmes limites concernant les variables aléatoires i.i.d.. Nous pouvons donner à présent une version multi-dimensionnelle du théorème central-limite. Nous dirons qu'un vecteur aléatoire est dans  $L^2$  si et seulement si toutes ses coordonnées sont dans  $L^2$ .

**Théorème 3.3.1 (Théorème central limite vectoriel)** *Soit  $(X_n)_{n \geq 1}$  une suite de **vecteurs aléatoires i.i.d.** de  $\mathbb{R}^d$  appartenant à  $L^2$ , d'espérance  $m \in \mathbb{R}^d$  et de matrice de covariance  $\Sigma \in \mathcal{M}_d(\mathbb{R})$ . Notons pour tout  $n \geq 1$*

$$S_n = X_1 + \dots + X_n.$$

*On a la convergence en loi*

$$\text{loi}\left(\frac{S_n - nm}{\sqrt{n}}\right) \xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \Sigma).$$

**Remarques :** On pourra écrire ainsi une sorte de développement limité des moyennes empiriques de vecteurs i.i.d. sous la forme

$$\frac{S_n}{n} \simeq m + \frac{1}{\sqrt{n}}G$$

où  $G$  est un vecteur gaussien de loi  $\mathcal{N}(0, \Sigma)$ .

## 3.4 Carrés de gaussiennes

Nous allons introduire quelques lois apparaissant fréquemment en présence de carrés de gaussiennes et qui nous seront très utiles pour les chapitres de statistique suivants.

### 3.4.1 Les lois gamma

**Définition 3.4.1** Soient  $a, b > 0$ . La loi  $\gamma_{a,b}$  est la loi sur  $\mathbb{R}$  de densité

$$c(a, b) \exp(-ax)x^{b-1}1_{x>0}dx$$

où  $c(a, b)$  est tel que  $\int_0^\infty c(a, b) \exp(-ax)x^{b-1}dx = 1$ , i.e.

$$c(a, b) = \left( \int_0^\infty \exp(-ax)x^{b-1}dx \right)^{-1}.$$

Les lois gamma forment une classe assez importante de lois. En particulier, cette classe contient les lois exponentielles. En effet, la loi  $\gamma_{a,1}$  n'est autre que la **loi exponentielle de paramètre  $a$** . Nous pouvons calculer la transformée de Laplace d'une loi  $\gamma_{a,b}$ .

**Proposition 3.4.2** Soient  $a, b > 0$  et  $X_{a,b}$  une variable aléatoire de loi  $\gamma_{a,b}$ . Pour tout  $t \geq 0$ ,

$$\mathbb{E}(\exp(-tX_{a,b})) = \left( \frac{a}{a+t} \right)^b.$$

*Preuve :* Nous allons faire la démonstration, qui est simple. Soit  $t \geq 0$ . Nous avons par définition de  $\gamma_{a,b}$ ,

$$\begin{aligned} \mathbb{E}(\exp(-tX_{a,b})) &= c(a, b) \int_0^\infty \exp(-ax)x^{b-1} \exp(-tx)dx \\ &= c(a, b) \int_0^\infty \exp(-(a+t)x)x^{b-1}dx. \end{aligned}$$

Nous faisons le changement de variables  $x = \frac{a}{a+t}u$ . Il vient

$$\begin{aligned} \mathbb{E}(\exp(-tX_{a,b})) &= \left( \frac{a}{a+t} \right)^b c(a, b) \int_0^\infty \exp(-au)u^{b-1}du \\ &= \left( \frac{a}{a+t} \right)^b. \end{aligned}$$

□

Le calcul de cette transformée de Laplace nous permet d'obtenir une propriété de somme des lois gamma qui rend leur usage particulièrement pratique. On admet que la transformée de Laplace jouit des mêmes propriétés que la transformée de Fourier.

**Proposition 3.4.3** Soit  $a, b, b' > 0$  et soient  $X_{a,b}$  et  $X_{a,b'}$  deux variables aléatoires **indépendantes** de lois respectives  $\gamma_{a,b}$  et  $\gamma_{a,b'}$ . La variable aléatoire

$$Y = X_{a,b} + X_{a,b'}$$

suit alors une loi  $\gamma_{a,b+b'}$ .

*Preuve :* Nous allons calculer la transformée de Laplace de  $Y$ . Soit donc  $t \geq 0$ . Nous avons par indépendance de  $X_{a,b}$  et  $X_{a,b'}$

$$\begin{aligned} \mathbb{E}(\exp(-tY)) &= \mathbb{E}(\exp(-t(X_{a,b} + X_{a,b'}))) \\ &= \mathbb{E}(\exp(-tX_{a,b}) \exp(-tX_{a,b'})) \\ &= \mathbb{E}(\exp(-tX_{a,b}))\mathbb{E}(\exp(-tX_{a,b'})). \end{aligned}$$

D'après la proposition 3.4.2, nous obtenons donc

$$\begin{aligned} \mathbb{E}(\exp(-tY)) &= \left( \frac{a}{a+t} \right)^b \left( \frac{a}{a+t} \right)^{b'} \\ &= \left( \frac{a}{a+t} \right)^{b+b'}. \end{aligned}$$

La transformée de Laplace de  $Y$  est donc celle d'une loi  $\gamma_{a,b+b'}$ . Comme les transformées de Laplace caractérisent les lois,  $Y$  suit une loi  $\gamma_{a,b+b'}$ . □

### 3.4.2 Carrés de gaussiennes - Lois du chi-deux

Les lois gamma vont nous être utiles en présence de carrés ou de somme de carrés de gaussiennes. Nous commençons par une proposition simple.

**Proposition 3.4.4** *Soit  $X$  une variable aléatoire gaussienne centrée réduite. La loi de  $X^2$  est  $\gamma_{\frac{1}{2}, \frac{1}{2}}$ , loi appelée aussi loi du chi-deux à un degré de liberté et notée  $\chi^2(1)$ .*

*Preuve :* Soit  $f$  une fonction continue et bornée. Posons  $Y = X^2$ . On a

$$\begin{aligned}\mathbb{E}(f(Y)) = \mathbb{E}(f(X^2)) &= (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(x^2) \exp\left(-\frac{x^2}{2}\right) dx. \\ &= 2(2\pi)^{-\frac{1}{2}} \int_0^{+\infty} f(x^2) \exp\left(-\frac{x^2}{2}\right) dx.\end{aligned}$$

On fait le changement de variables  $x = u^{\frac{1}{2}}$  de sorte que

$$\mathbb{E}(f(Y)) = (2\pi)^{-\frac{1}{2}} \int_0^{+\infty} f(u) \exp\left(-\frac{u}{2}\right) u^{-\frac{1}{2}} du.$$

La variable aléatoire  $Y$  a donc pour densité

$$(2\pi)^{\frac{1}{2}} \exp\left(-\frac{u}{2}\right) u^{-\frac{1}{2}} 1_{u>0} du$$

et suit bien une loi  $\gamma_{\frac{1}{2}, \frac{1}{2}}$ .  $\square$

Grâce à la Proposition 3.4.3 nous allons connaître la loi de la somme de  $n$  carrés de gaussiennes indépendantes.

**Proposition 3.4.5** *Soient  $X_1, \dots, X_n$   $n$  v.a. gaussiennes centrées réduites **indépendantes**. La v.a.*

$$Y = X_1^2 + \dots + X_n^2$$

*suit la loi  $\gamma_{\frac{1}{2}, \frac{n}{2}}$ , loi appelée loi du chi-deux à  $n$  degrés de liberté et notée  $\chi^2(n)$ .*

*Preuve :* Puisque  $X_i^2$  suit une loi  $\gamma_{\frac{1}{2}, \frac{1}{2}}$  pour tout  $1 \leq i \leq n$  par la proposition 3.4.4, et comme les variables  $X_1^2, \dots, X_n^2$  sont indépendantes, on peut appliquer la proposition 3.4.3 à  $n - 1$  reprises pour obtenir que  $Y$  suit une loi  $\gamma_{\frac{1}{2}, \frac{n}{2}}$ .  $\square$

On admet le lemme suivant :

**Lemme 3.4.6** *On appelle matrice de projection orthogonale toute matrice  $A \in \mathcal{M}_n(\mathbb{R})$  telle que ( $A^T$  désignant la transposée de  $A$ )*

$$A^T = A \text{ et } A^2 = A.$$

*Soit  $A = [a_{i,j}]$  une matrice de projection orthogonale. On a alors :*

1. *le rang de  $A$  est  $\text{rg}(A) = \sum_{i=1}^n a_{i,i}$*
2. *il existe une matrice orthogonale  $P \in \mathcal{M}_n(\mathbb{R})$ , c'est-à-dire tel que  $P$  est inversible et  $P^T = P^{-1}$  pour laquelle*

$$A = P^T \begin{pmatrix} Id_r & 0 \\ 0 & 0 \end{pmatrix} P.$$

*En outre,  $r$  est le rang de  $A$ .*

Nous terminons par un théorème important en statistique qui nous donne un critère pour reconnaître une lois du chi-deux. On rappelle qu'un vecteur gaussien centré réduit est un vecteur gaussien d'espérance le vecteur nul et de matrice de covariance la matrice identité, i.e. un vecteur aléatoire dont les coordonnées sont des v.a.i.i.d. de loi  $\mathcal{N}(0, 1)$ .

**Proposition 3.4.7** Soit  $X = (X_1, \dots, X_n)^T$  un vecteur gaussien centré réduit (le symbole de transposition  $T$  signifie qu'on le considère comme un vecteur colonne). Soit  $A \in \mathcal{M}_n(\mathbb{R})$  une matrice de projection orthogonale. Alors, la variable aléatoire

$$Y = q_A(X)$$

suit une loi  $\chi^2(r)$  où  $r = \text{rg}(A)$ .

**Remarque :** Le fait qu'on considère  $X$  comme un vecteur colonne permet d'avoir l'équation :

$$q_A(X) = X^T A X.$$

*Preuve :* Soit donc  $r = \text{rg}(A)$  et une matrice  $P$  associée à  $A$  comme dans le lemme précédent. Posons  $Z = PX = (Z_1, \dots, Z_r)$ . Par la proposition 3.2.16,  $Z$  est un vecteur gaussien d'espérance  $P\mathbb{E}(X) = 0$  et de matrice de covariance

$$PCov(X)P^T = P^T P = Id_n$$

où l'on a utilisé le fait que  $X$  est centré réduit et l'identité  $P^T = P^{-1}$ . Ainsi, le vecteur gaussien  $Z$  est également centré réduit et donc  $Z_1, \dots, Z_r$  sont  $n$  variables aléatoires gaussiennes centrées réduites et i.i.d.. On note par ailleurs que

$$q_A(X) = X^T A X = X^T P^T \begin{pmatrix} Id_r & 0 \\ 0 & 0 \end{pmatrix} P X = Z^T \begin{pmatrix} Id_r & 0 \\ 0 & 0 \end{pmatrix} Z$$

et par conséquent,

$$Y = Z_1^2 + \dots + Z_r^2.$$

Comme  $Z_1, \dots, Z_r$  sont gaussiennes centrées réduites et i.i.d., on peut affirmer d'après la proposition 3.4.5 que  $Y$  suit la loi  $\chi^2(r)$ .  $\square$

L'exercice suivant étend ce résultat au cas où  $X$  n'est pas centré réduit.

**Exercice 3.4.1** Soit  $X = (X_1, \dots, X_n)^T$  un vecteur gaussien d'espérance nulle et de matrice de covariance  $C$  (le symbole de transposition  $T$  signifie qu'on le considère comme un vecteur colonne). On admet l'existence d'une matrice symétrique  $M$  telle que  $C = M^2$ .

- a) Montrer que la loi de  $X$  est celle de  $MY$ , où  $Y$  est un vecteur gaussien de  $\mathbb{R}^n$  centré réduit (considéré comme un vecteur colonne).  
 b) Montrer que la loi de  $X_1^2 + \dots + X_n^2$  est une loi du chi-deux à  $r$  degrés de liberté, où  $r$  est le rang de  $C$ .

Indication : Montrer que  $X_1^2 + \dots + X_n^2 = X^T X$ . La v.a.  $X_1^2 + \dots + X_n^2$  a donc la même loi que  $Z^T C Z$ , puis établir  $Z^T C Z = q_C(Z)$ .

### 3.4.3 Lois de Student

Introduisons encore une nouvelle loi, qui apparaît aussi souvent en statistiques.

**Définition 3.4.8** Soit  $X$  une v.a. de loi normale centrée réduite et soit  $Y$  une v.a. indépendante de  $X$  suivant une loi  $\chi^2(n)$ . Alors, la v.a.

$$\frac{X}{\sqrt{\frac{1}{n}Y}}$$

suit une loi de densité sur  $\mathbb{R}$

$$c_n \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx$$

où

$$c_n = \left( \int_{-\infty}^{+\infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx \right)^{-1}.$$

Cette loi est appelée **loi de Student de paramètre  $n$** .

# Chapitre 4

## Statistiques

### Introduction : problème fondamental de la Statistique

Nous cherchons à étudier le caractère aléatoire de  $n$  expériences fournissant les résultats chiffrés  $x_1, \dots, x_n$ . Dans tout ce chapitre, nous nous placerons dans le cas où ces  $n$  expériences sont réalisées à l'identique et indépendamment les unes des autres. Nous faisons donc l'hypothèse que les résultats  $x_1, \dots, x_n$  (on appelle aussi  $(x_1, \dots, x_n)$  une observation) proviennent de  $n$  variables aléatoires  $X_1, \dots, X_n$  i.i.d.. La famille  $(X_1, \dots, X_n)$  est appelée un **n-échantillon** et la loi commune des  $X_i$  est notée  $\mathcal{L}(X)$ . L'aléa lié à notre expérience est tel que

$$X_1 = x_1, \dots, X_n = x_n.$$

Dans ce qui suit, nous tacherons de désigner les variables aléatoires par des majuscules et les résultats des expériences par des minuscules pour bien distinguer ce qui est aléatoire de ce qui ne l'est pas.

Le problème fondamental de la Statistique est de déterminer la loi commune des  $X_i$  supposée inconnue à partir des résultats  $(x_1, \dots, x_n)$ . Il s'agit donc du problème inverse de la théorie des probabilités dont l'objet est l'étude des propriétés de variables aléatoires de lois données. Ce problème est dit "mal posé" dans le sens où il est illusoire d'espérer obtenir une réponse exacte. Autrement dit, les valeurs  $x_1, \dots, x_n$  peuvent très bien être les mêmes avec des lois différentes pour  $X_1, \dots, X_n$  (regarder par exemple le cas de variables de Bernoulli). Ainsi, le statisticien ne donnera pas de réponse absolue ; il cherchera à déterminer une réponse approchée "la plus probable" selon des critères définis à l'avance.

## 4.1 Modèles à paramètres

### 4.1.1 Définition

Pour obtenir des résultats précis et intéressants, nous faisons un certain nombre d'hypothèses liées à l'expérience effectuée sur la loi commune  $\mathcal{L}(X)$  des variables  $X_1, \dots, X_n$  de l'échantillon. Nous restreindrons notre étude à la **statistique paramétrique**. Dans cette approche les différentes lois considérées possibles a priori pour les variables  $X_i$  sont les  $P_\theta$ ,  $\theta \in \Theta$ , où  $\{P_\theta / \theta \in \Theta\}$  est un ensemble de lois de probabilités sur  $\mathbb{R}$  :

$$\mathcal{L}(X) \in \{P_\theta ; \theta \in \Theta\}.$$

**Définition 4.1.1 (Modèle à paramètre)** *Un modèle à paramètre (ou modèle paramétrique) est un ensemble  $\{P_\theta / \theta \in \Theta\}$  de lois de probabilités sur  $\mathbb{R}$ .*

Le statisticien cherche alors à déterminer le paramètre  $\theta \in \Theta$  inconnu pour lequel  $\mathcal{L}(X) = P_\theta$ . L'inconnue n'est plus une probabilité, c'est un élément de  $\Theta$ , ce qui est bien sûr beaucoup plus maniable quand  $\Theta$  est un ensemble assez simple comme une partie de  $\mathbb{R}^n$ .

## 4.1.2 Exemples de modèles paramétriques

### Le modèle de Bernoulli

Lorsque l'expérience admet deux résultats possible uniquement, que nous noterons simplement 0 et 1, le modèle paramétrique à utiliser est le **modèle de Bernoulli**. L'hypothèse à faire est que

$$\mathcal{L}(X) \in \{\mathcal{B}(p) ; 0 \leq p \leq 1\},$$

où  $\mathcal{B}(p)$  désigne une loi de Bernoulli de paramètre  $p$ . L'ensemble des paramètres est donc  $\Theta = [0, 1] \subset \mathbb{R}$ . Pour déterminer  $\mathcal{L}(X)$ , il s'agit donc de déterminer  $p$ .

### Le modèle gaussien

Un des modèles statistiques les plus importants est le modèle gaussien. On suppose en général inconnues l'espérance et la variance des variables  $X_i$ . Autrement dit

$$\mathcal{L}(X) \in \{\mathcal{N}(m, \sigma^2) ; m \in \mathbb{R}, \sigma \geq 0\}.$$

L'ensemble des paramètres est  $\Theta = \mathbb{R} \times \mathbb{R}_+ \subset \mathbb{R}^2$ . La famille  $(X_1, \dots, X_n)$  est appelée **échantillon gaussien**

### Autres modèles

Il existe bien sûr un très grand nombre de modèles. On peut distinguer en particulier les modèles discrets où la loi est supposée être une loi discrète (par exemple une loi de Bernoulli  $\mathcal{B}(p)$ ,  $0 \leq p \leq 1$ , de Poisson  $\mathcal{P}(\theta)$ ,  $\theta \geq 0 \dots$ ) et les modèles à densité où la loi de  $X$  est supposée avoir une densité de la forme  $f_\theta(x)$  où  $\theta \in \Theta$  est un paramètre inconnu (par exemple une loi  $\gamma_{a,b}$ ,  $a, b > 0$ , une loi normale  $\mathcal{N}(m, \sigma^2)$ ,  $m \in \mathbb{R}, \sigma \geq 0 \dots$ ).

## 4.2 Estimateurs

### 4.2.1 Estimateur du maximum de vraisemblance

#### Cas discret

Supposons que  $(X_1, \dots, X_n)$  soit un  $n$ -échantillon tel que la loi commune  $\mathcal{L}(X)$  de  $X_1, \dots, X_n$  soit une loi discrète. Autrement dit  $X$  ne prend qu'un nombre fini de valeurs dans un ensemble  $E$ . Supposons également que nous disposions d'un modèle paramétrique  $\{\mathbb{P}_\theta / \theta \in \Theta\}$  tel que

$$\mathcal{L}(X) \in \{\mathbb{P}_\theta ; \theta \in \Theta\}.$$

La probabilité d'obtenir une observation  $(x_1, \dots, x_n)$  dépend du paramètre  $\theta$  du modèle. Elle est égale à

$$\begin{aligned} \mathbb{P}((X_1, \dots, X_n) = (x_1, \dots, x_n)) &= \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) \text{ par indépendance} \\ &= \prod_{i=1}^n \mathbb{P}_\theta(x_i) \text{ si } \mathcal{L}(X) = \mathbb{P}_\theta. \end{aligned}$$

On définit donc la **vraisemblance** du modèle comme la fonction des variables  $\theta, x_1, \dots, x_n$  définie par

$$l_n(x_1, \dots, x_n, \theta) = \prod_{i=1}^n \mathbb{P}_\theta(x_i),$$

pour tous  $\theta \in \Theta$ ,  $x_1, \dots, x_n \in E$ . La **log-vraisemblance** du modèle est quant à elle définie par

$$L_n(x_1, \dots, x_n, \theta) = \log l_n(x_1, \dots, x_n, \theta) = \sum_{i=1}^n \log P_\theta(x_i).$$

On pose  $\log(0) = -\infty$ , de sorte que la log-vraisemblance peut valoir  $-\infty$  si un des  $P_\theta(x_i)$  vaut 0. Si nous disposons d'une observation  $(x_1, \dots, x_n)$ , la **méthode du maximum de vraisemblance** consiste à rechercher le paramètre  $\theta$  pour lequel la vraisemblance est la plus élevée, c'est-à-dire pour lequel l'observation est la plus vraisemblable. En général, sous de bonnes conditions, il existe  $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$  tel que

$$l_n(x_1, \dots, x_n, \hat{\theta}_n) = \max_{\theta \in \Theta} l_n(x_1, \dots, x_n, \theta).$$

Le paramètre  $\hat{\theta}_n$  est alors un estimateur appelée **estimateur du maximum de vraisemblance**. Remarquons qu'on a aussi

$$L_n(x_1, \dots, x_n, \hat{\theta}_n) = \max_{\theta \in \Theta} L_n(x_1, \dots, x_n, \theta).$$

### Cas continu

Supposons à présent que le modèle que nous considérons soit continu. La loi de  $X$  est supposée avoir pour densité  $f_{\theta_0}(x)dx$  où  $\theta_0 \in \Theta$ . Si on regarde un  $n$ -échantillon  $(X_1, \dots, X_n)$ , la vraisemblance est définie comme la fonction des variables  $x_1, \dots, x_n$  et  $\theta$  définie par

$$l_n(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f_\theta(x_i), \quad \theta \in \Theta,$$

tandis que la log-vraisemblance est définie par

$$L_n(x_1, \dots, x_n, \theta) = \sum_{i=1}^n \log f_\theta(x_i)$$

L'estimateur du maximum de vraisemblance est alors défini par

$$l_n(x_1, \dots, x_n, \hat{\theta}_n) = \max_{\theta \in \Theta} l_n(x_1, \dots, x_n, \theta).$$

Sous de bonnes conditions, cet estimateur est convergent.

### 4.2.2 Estimateurs sans biais, estimateurs convergents

La méthode du maximum de vraisemblance est la méthode la plus naturelle pour estimer  $\theta \in \Theta$  à partir de  $n$  observations, mais son utilisation suppose deux choses :

- $\forall x \in E^n$ , la fonction  $\theta \mapsto l(x_1, \dots, x_n, \theta)$  admet un maximum en un unique point de  $\Theta$
- on sait calculer ce maximum

Ce n'est pas toujours le cas. Aussi, on va s'intéresser à des estimateurs qui permettent pas d'identifier  $\theta$  directement, mais qui donnent une approximation d'une certaine fonction  $Y$  de  $\theta$ . Considérons un  $n$ -échantillon  $(X_1, \dots, X_n)$ . La loi commune des  $X_i$  est toujours notée  $\mathcal{L}(X)$ .

Supposons que nous ayons choisi un modèle paramétrique pour la loi  $\mathcal{L}(X)$ . Il existe un ensemble  $\Theta$  et une famille de lois  $P_\theta$ ,  $\theta \in \Theta$ , telle que

$$\mathcal{L}(X) \in \{P_\theta ; \theta \in \Theta\}.$$

Nous cherchons à estimer le paramètre  $\theta$  inconnu tel que  $\mathcal{L}(X) = P_\theta$  ou plus généralement une fonction de la forme  $Y(\theta)$  dépendant de  $\theta$ , c'est-à-dire en fait de la loi  $\mathcal{L}(X)$ . Supposons avoir

choisi un estimateur  $Y$ , c'est à dire une fonction  $Y_n : \mathbb{R}^n \rightarrow \mathbb{R}$  telle que la v.a.  $Y_n(X_1, \dots, X_n)$  est censée nous donner une valeur approchée de  $Y(\theta)$ . Si notre expérience nous a fourni les résultats

$$X_1 = x_1, \dots, X_n = x_n,$$

la valeur  $\hat{y} = Y_n(x_1, \dots, x_n)$  sera la valeur choisie pour approcher (ou estimer) l'inconnue  $Y(\theta)$ .

Il s'agit ensuite d'étudier la qualité de cette estimation. La valeur  $\hat{y}$  est aléatoire, elle dépend des résultats de notre expérience. Une propriété souhaitable est qu'en moyenne, elle soit égale à  $Y(\theta)$ . C'est ce qui nous amène à donner la définition suivante :

**Définition 4.2.1 (Estimateur sans biais)** Soit  $\{P_\theta/\theta \in \Theta\}$  un modèle à paramètre, i.e. un ensemble de lois de probabilités sur  $\mathbb{R}$ . Soit  $Y : \Theta \rightarrow \mathbb{R}$ . Soit, pour tout  $n \geq 1$ ,  $Y_n : \mathbb{R}^n \rightarrow \mathbb{R}$ .

On dit que  $(Y_n)$  est une suite d'estimateurs **sans biais** pour  $Y$  (ou de  $Y$ ) si pour tout  $\theta \in \Theta$ , pour toute suite  $(X_i)_{i \geq 1}$  de v.a.i.i.d. de loi  $P_\theta$ ,

$$\forall n \geq 1, \mathbb{E}(Y_n(X_1, \dots, X_n)) = Y(\theta).$$

Notons que cette relation doit être vraie quel que soit le paramètre  $\theta$  dont nous ne connaissons pas la valeur a priori.

Considérons à présent une suite infinie de variables aléatoires i.i.d.  $X_n$ ,  $n \geq 1$ , de loi commune

$$\mathcal{L}(X) \in \{P_\theta ; \theta \in \Theta\}.$$

Supposons que nous cherchions à estimer une quantité  $Y(\theta)$  et que nous disposions pour cela d'une suite d'estimateurs  $(Y_n)_{n \geq 1}$ , de telle sorte que  $Y_n$  soit un estimateur associé à l'échantillon  $(X_1, \dots, X_n)$  pour tout  $n \geq 1$ . Il est très souhaitable que notre suite d'estimateurs se rapproche de  $Y(\theta)$  quand  $n$  augmente, d'autant plus que le théorème fondamental de la statistique affirme que cela est possible. C'est ce qui nous amène à donner la définition suivante :

**Définition 4.2.2 (Suite convergente d'estimateurs)** Soit  $\{P_\theta/\theta \in \Theta\}$  un modèle à paramètre, i.e. un ensemble de lois de probabilités sur  $\mathbb{R}$ . Soit  $Y : \Theta \rightarrow \mathbb{R}$ . Soit, pour tout  $n \geq 1$ ,  $Y_n : \mathbb{R}^n \rightarrow \mathbb{R}$ . On dit que  $(Y_n)$  est une suite **convergente** d'estimateurs pour  $Y$  (ou de  $Y$ ) si pour tout  $\theta \in \Theta$ , pour toute suite  $(X_i)_{i \geq 1}$  de v.a.i.i.d. de loi  $P_\theta$ ,

$$Y_n(X_1, \dots, X_n) \xrightarrow[n \rightarrow \infty]{} Y(\theta)$$

Il sera très judicieux de choisir des estimateurs sans biais et convergents.

Un problème important est de mesurer la "qualité" d'un estimateur et de pouvoir alors comparer deux estimateurs. Il n'y a pas de manière simple et absolue de le faire. Nous pouvons cependant introduire un concept appelé "risque quadratique". Supposons disposer d'un échantillon  $(X_1, \dots, X_n)$  tel que  $\mathcal{L}(X) \in \{P_\theta ; \theta \in \Theta\}$ .

**Définition 4.2.3 (Risque quadratique d'une suite d'estimateurs)** Soit  $\{P_\theta/\theta \in \Theta\}$  un modèle à paramètre, i.e. un ensemble de lois de probabilités sur  $\mathbb{R}$ . Soit  $Y : \Theta \rightarrow \mathbb{R}$ . Soit, pour tout  $n \geq 1$ ,  $Y_n : \mathbb{R}^n \rightarrow \mathbb{R}$ . Pour tout  $n \geq 1$ , on définit le **risque quadratique** de l'estimateur  $Y_n$  de  $Y$  comme la fonction  $R_n$  définie sur  $\Theta$  par

$$R_n(\theta) = \mathbb{E}((Y_n(X_1, \dots, X_n) - Y(\theta))^2)$$

pour  $(X_i)_{i \geq 1}$  v.a.i.i.d. de loi  $P_\theta$ .

**Remarques :**

- Lorsque les estimateurs sont sans biais, pour tout  $\theta \in \Theta$ , on a

$$R_n(\theta) = \text{Var}(Y_n(X_1, \dots, X_n))$$

pour  $(X_i)_{i \geq 1}$  v.a.i.i.d. de loi  $P_\theta$ .

- Intuitivement, plus le risque quadratique est faible, plus l'estimateur est proche de la quantité qu'il estime. Il sera donc préférable de choisir des estimateurs au risque quadratique le plus faible possible.

- Le risque quadratique donne aussi un ordre de grandeur de l'écart entre  $Y_n(X_1, \dots, X_n)$  et  $Y(\theta)$  (voir la remarque (4.2.6)).



### 4.2.3 Estimateurs de l'espérance et de la variance

Soit  $(X_1, \dots, X_n)$  un n-échantillon. Il est très classique de devoir estimer l'espérance et la variance de  $X$ . Dans un modèle de Bernoulli par exemple, l'espérance caractérise la loi. Dans un modèle gaussien, l'espérance et la variance caractérisent la loi.

**Définition 4.2.4 (Estimateur pour l'espérance)** Soit  $\{P_\theta/\theta \in \Theta\}$  un modèle à paramètre, i.e. un ensemble de lois de probabilités sur  $\mathbb{R}$ . On suppose que pour tout  $\theta \in \Theta$  toute v.a. de loi  $P_\theta$  est intégrable, d'espérance notée  $\mathbb{E}_\theta(X)$ . Soit  $Y : \Theta \rightarrow \mathbb{R}$ ,  $\theta \mapsto \mathbb{E}_\theta(X)$ . Soit, pour tout  $n \geq 1$ ,  $M_n : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x \mapsto \frac{1}{n} \sum_{i=1}^n x_i$ .

Alors  $(M_n)$  est une suite **convergente et sans biais** d'estimateurs pour  $Y$ .

*Preuve :* Soit  $\theta \in \Theta$ , soit  $X_i$  une suite de v.a.i.i.d. de loi  $P_\theta$ .

Soit  $n \geq 1$ .

$$\mathbb{E}(M_n(X_1, \dots, X_n)) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \mathbb{E}_\theta(X).$$

Donc les estimateurs sont sans biais, le caractère convergent est donné par la LGN.  $\square$

**Proposition 4.2.5** Soit  $\{P_\theta/\theta \in \Theta\}$  un modèle à paramètre, i.e. un ensemble de lois de probabilités sur  $\mathbb{R}$ . On suppose que pour tout  $\theta \in \Theta$  toute v.a. de loi  $P_\theta$  est de carré intégrable, d'espérance notée  $\mathbb{E}_\theta(X)$  et de variance notée  $V_\theta(X)$ . Alors pour tout  $n \geq 1$ , le risque quadratique de l'estimateur  $M_n$  de l'espérance est donné par la formule :

$$\forall \theta \in \Theta, R_n(\theta) = \frac{V_\theta(X)}{n}.$$

**Remarque 4.2.6** On en déduit que si, à la suite de  $n$  observations  $x_1, \dots, x_n$ , on assimile  $M_n(x_1, \dots, x_n)$  à l'espérance des  $X_i$ , l'erreur commise est en moyenne de l'ordre de  $\frac{\sigma}{\sqrt{n}}$ , où  $\sigma^2$  est la variance des  $X_i$ .

*Preuve :* Soit  $\theta \in \Theta$ , soit  $X_i$  une suite de v.a.i.i.d. de loi  $P_\theta$ .

Soit  $n \geq 1$ .

$$\mathbb{E}((M_n(X_1, \dots, X_n) - \mathbb{E}_\theta(X))^2) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{V_\theta(X)}{n}.$$

$\square$

**Définition 4.2.7 (Estimateur pour la variance)** Soit  $\{P_\theta/\theta \in \Theta\}$  un modèle à paramètre, i.e. un ensemble de lois de probabilités sur  $\mathbb{R}$ . On suppose que pour tout  $\theta \in \Theta$  toute v.a. de loi  $P_\theta$  est de carré intégrable, d'espérance notée  $\mathbb{E}_\theta(X)$  et de variance notée  $V_\theta(X)$ . Soit  $V : \Theta \rightarrow \mathbb{R}$ ,  $\theta \mapsto V_\theta(X)$ . Soit, pour tout  $n \geq 1$ ,

$$\begin{aligned} \bar{V}_n : \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\mapsto \frac{1}{n} \sum_{i=1}^n (x_i - M_n(x))^2 \end{aligned}$$

Soit  $V_n = \frac{n}{n-1} \bar{V}_n$ .

Alors  $(\bar{V}_n)$  et  $(V_n)$  sont des suites **convergentes** d'estimateurs pour  $V$ , et  $(V_n)$  est **sans biais**.

*Preuve :* Soit  $\theta \in \Theta$ , soit  $X_i$  une suite de v.a.i.i.d. de loi  $P_\theta$ .

Soit  $n \geq 1$ . Notons que

$$\begin{aligned}\bar{V}_n &= \frac{1}{n} \sum_{i=1}^n (X_i^2 - 2X_i M_n + M_n^2) \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - 2 \frac{M_n}{n} \sum_{i=1}^n X_i + M_n^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - M_n^2 \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \left( \frac{1}{n} \sum_{i=1}^n X_i \right)^2.\end{aligned}$$

La loi des grands nombres nous dit que

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(X^2),$$

et que

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(X).$$

Par conséquent,

$$\bar{V}_n \xrightarrow[n \rightarrow \infty]{} \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \text{Var}(X).$$

L'estimateur  $\bar{V}_n$  est donc un estimateur convergent de la variance donc  $V_n$  aussi (car  $\lim_{n \rightarrow \infty} \frac{n}{n-1} = 1$ ).

Calculons l'espérance de  $\bar{V}_n$  :

$$\begin{aligned}\mathbb{E}(\bar{V}_n) &= \frac{1}{n} \mathbb{E} \left( \sum_{i=1}^n X_i^2 \right) - \frac{1}{n^2} \mathbb{E} \left( \left( \sum_{i=1}^n X_i \right)^2 \right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i^2) - \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \mathbb{E}(X_i X_j) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i^2) - \frac{1}{n^2} \sum_{1 \leq i, j \leq n, i \neq j} \mathbb{E}(X_i X_j) - \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}(X_i^2).\end{aligned}$$

En utilisant le caractère i.i.d. des variables  $X_i$ , on obtient

$$\begin{aligned}\mathbb{E}(\bar{V}_n) &= \frac{1}{n} n \mathbb{E}(X^2) - \frac{1}{n^2} n(n-1) \mathbb{E}(X)^2 - \frac{1}{n^2} n \mathbb{E}(X^2) \\ &= \frac{n-1}{n} (\mathbb{E}(X^2) - \mathbb{E}(X)^2) \\ &= \frac{n-1}{n} \text{Var}(X).\end{aligned}$$

Donc  $V_n$  est sans biais.  $\square$

L'estimateur  $\bar{V}_n$  n'est donc pas un estimateur sans biais de la variance de  $X$ . En revanche,

$$V_n = \frac{n}{n-1} \bar{V}_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2 = \frac{1}{n-1} \sum_{i=1}^n X_i^2 - \frac{1}{n(n-1)} \left( \sum_{i=1}^n X_i \right)^2$$

est bien un estimateur sans biais et convergent de la variance. C'est cet estimateur qui est généralement utilisé. Néanmoins, lorsque  $n$  est très élevé, il est très proche de  $\bar{V}_n$ .

## 4.3 Intervalles de confiance non asymptotiques

### 4.3.1 Définitions

Nous disposons d'un  $n$  échantillon  $(X_1, \dots, X_n)$  et d'un modèle paramétrique de telle sorte que  $\mathcal{L}(X) \in \{P_\theta ; \theta \in \Theta\}$ . Nous cherchons des renseignements sur une fonction de la forme

$Y(\theta)$  à partir d'une observation  $(x_1, \dots, x_n)$ . Plus précisément, nous cherchons une région (nous nous restreindrions ici à un intervalle) déterminée à partir de  $(x_1, \dots, x_n)$  à laquelle appartient "vraisemblablement"  $Y(\theta)$ .

**Définition 4.3.1** Soit  $Y$  une fonction de  $\Theta$  vers  $\mathbb{R}$ . soit, pour  $n \geq 1$ ,  $a_n$  et  $b_n$  des fonctions de  $\mathbb{R}^n$  vers  $\mathbb{R}$ . Soit  $0 < \alpha < 1$ .

- On dit que  $[a_n, b_n]$  est un intervalle de confiance de niveau  $1 - \alpha$  si et seulement si pour tout  $\theta \in \Theta$ ,

$$P(a_n(X_1, \dots, X_n) \leq Y(\theta) \leq b_n(X_1, \dots, X_n)) = 1 - \alpha$$

pour  $X_1, \dots, X_n$  v.a.i.i.d. de loi  $P_\theta$ .

- On dit que  $[a_n, b_n]$  est un intervalle de confiance par excès de niveau  $1 - \alpha$  si et seulement si pour tout  $\theta \in \Theta$ ,

$$P(a_n(X_1, \dots, X_n) \leq Y(\theta) \leq b_n(X_1, \dots, X_n)) \geq 1 - \alpha$$

pour  $X_1, \dots, X_n$  v.a.i.i.d. de loi  $P_\theta$ .

**Remarque :** Si  $[a_n, b_n]$  est un intervalle de confiance (respectivement par excès) au niveau  $1 - \alpha$ , et  $P_\theta$  la loi des  $X_i$ , alors

$$P(Y(\theta) \notin [a_n(X_1, \dots, X_n), b_n(X_1, \dots, X_n)]) = \alpha \text{ (respectivement } \leq \alpha).$$

En pratique, si on dispose d'un estimateur  $Y_n$  de  $Y(\theta)$  et on cherche un intervalle de confiance de niveau  $1 - \alpha$  de la forme  $[Y_n - \varepsilon, Y_n + \varepsilon]$ . Supposons avoir construit un tel intervalle. Si par exemple  $\alpha = 0.05$ , la probabilité que le paramètre recherché soit dans l'intervalle construit  $[Y_n - \varepsilon, Y_n + \varepsilon]$  est de 95%, soit une probabilité très proche de 1. Etant donnée une observation  $(x_1, \dots, x_n)$ , nous dirons qu'au niveau de confiance 95%,  $Y(\theta)$  vaut  $Y_n(x_1, \dots, x_n) \pm \varepsilon$ .

Pour déterminer des intervalles de confiance, nous aurons très souvent besoin de la notion de **quantile** d'une loi.

**Définition 4.3.2** Soit  $X$  une variable aléatoire de loi  $\mathcal{L}$  et soit  $0 < \beta < 1$ . On appelle quantile de niveau (ou d'ordre)  $\beta$  de  $\mathcal{L}$  un réel  $a_\beta$  tel que

$$P(X \leq a_\beta) = \beta.$$

**Remarque :** Pour trouver un quantile d'ordre  $\beta$ , si  $F$  est la fonction de répartition de  $X$ , il faut trouver  $a_\beta$  tel que

$$F(a_\beta) = \beta.$$

Il s'agit donc quand cela est possible d'inverser  $F$  et de prendre  $a_\beta = F^{-1}(\beta)$ . Nous pourrions toujours le faire lorsque  $X$  admet une densité continue strictement positive sur  $\mathbb{R}$ .

### 4.3.2 Exemple : le modèle de Bernouilli

Soit, pour  $\theta \in \Theta = ]0, 1[$ ,  $P_\theta$  la loi de Bernouilli de paramètre  $\theta$ .

**Proposition 4.3.3** Soit  $\theta \in ]0, 1[$ . Soit  $X_1, \dots, X_n$  une famille de v.a.i.i.d. de loi  $P_\theta$ . Alors pour tout  $\varepsilon > 0$ ,

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \theta\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}.$$

*Preuve :* Soit  $\varepsilon > 0$ . On a

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \theta\right| \geq \varepsilon\right) = P\left(\left|\sum_{i=1}^n X_i - n\theta\right| \geq n\varepsilon\right) \leq \frac{1}{n^2\varepsilon^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \text{ par Chebichef}$$

$$\text{donc par indep. des } X_i, P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \theta\right| \geq \varepsilon\right) \leq \frac{1}{n^2\varepsilon^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\theta(1-\theta)}{n\varepsilon^2}.$$

Or le maximum de la fonction  $\theta \mapsto \theta(1-\theta)$  sur  $[0, 1]$  est  $1/4$ , d'où le résultat.  $\square$

**Corollaire 4.3.4** Pour tout  $\alpha \in ]0, 1[$ , pour tout  $n \geq 1$ , en prenant  $\varepsilon > 0$  tel que  $\alpha = \frac{1}{4n\varepsilon^2}$ , les fonctions de  $\mathbb{R}^n$  vers  $\mathbb{R}$   $a_n(x) = \frac{1}{n} \sum_{i=1}^n x_i - \varepsilon$  et  $b_n(x) = \frac{1}{n} \sum_{i=1}^n x_i + \varepsilon$  définissent un intervalle de confiance par excès  $[a_n, b_n]$  de niveau  $1 - \alpha$  pour estimer  $\theta$  dans le modèle paramétrique de Bernouilli.

**Exercice 4.3.1** On considère un sondage dans lequel deux réponses sont possibles, “oui” et “non”. Combien faut-il interroger de personnes pour avoir une estimation à 3% près du pourcentage de gens en faveur de “oui”, et ce au moins 19 fois sur 20.

*Solution :* Soit  $\theta \in ]0, 1[$  la proportion de la population totale favorable au “oui”. Quand on interroge  $n$  personnes prises au hasard, indépendamment, comme c’est le cas dans les sondages, en notant 1 lorsque la réponse est “oui” et 0 lorsque la réponse est “non”, on observe  $n$  v.a.i.i.d. de loi  $\mathcal{B}(\theta)$ . On sait, par ce qui précède, que si  $x_1, \dots, x_n$  sont les réponses successives, on a, pour tout  $\varepsilon > 0$ ,

$$P\left(\frac{1}{n} \sum_{i=1}^n x_i - \varepsilon \leq \theta \leq \frac{1}{n} \sum_{i=1}^n x_i + \varepsilon\right) \geq 1 - \frac{1}{4n\varepsilon^2}.$$

On veut  $n$  tel que pour  $\varepsilon = 0,03$ ,  $\frac{1}{4n\varepsilon^2} \geq \frac{1}{20}$ . Il faut donc  $n \geq \frac{5 \cdot 10^4}{9}$ .

### 4.3.3 Exemple : le modèle gaussien

Considérons un échantillon gaussien  $(X_1, \dots, X_n)$ . Autrement dit  $X_1, \dots, X_n$  sont des variables aléatoires i.i.d. de loi  $\mathcal{N}(m, \sigma^2)$ , où  $m \in \mathbb{R}$  et  $\sigma \geq 0$ .

#### Résultats préliminaires

Nous avons des résultats très précis sur les estimateurs  $M_n$  et  $V_n$  introduits plus hauts.

**Proposition 4.3.5** Supposons que  $(X_1, \dots, X_n)$  soit un  $n$ -échantillon gaussien avec  $\mathcal{L}(X) = \mathcal{N}(m, \sigma^2)$ ,  $m \in \mathbb{R}$ ,  $\sigma > 0$ . Soient

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i$$

et

$$V_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2.$$

(1)  $M_n$  suit une loi  $\mathcal{N}(m, \frac{\sigma^2}{n})$ .

(2)  $\frac{n-1}{\sigma^2} V_n$  suit une loi  $\chi^2(n-1)$  (loi du chi-deux à  $n-1$  degrés de liberté = loi  $\gamma_{\frac{1}{2}, \frac{n-1}{2}}$ ).

(3)  $M_n$  et  $V_n$  sont indépendantes.

(4)  $\sqrt{n} \frac{M_n - m}{\sqrt{V_n}}$  suit une loi  $t_{n-1}$  (loi de Student à  $n-1$  degrés de liberté).

**Remarque 4.3.6** On rappelle que pour tout  $k \in \mathbb{N}^*$ , la loi de Student à  $k$  degrés de liberté est la loi de densité

$$c_k \frac{1}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}} \quad \text{avec} \quad c_k = \int_{\mathbb{R}} \frac{dt}{\left(1 + \frac{t^2}{k}\right)^{\frac{k+1}{2}}}.$$

*Preuve :*

1. Notons déjà que le vecteur  $X = (X_1, \dots, X_n)$  est un vecteur gaussien de loi  $\mathcal{N}((m, \dots, m), \sigma^2 Id)$  d'après un théorème du chapitre précédent. Par conséquent, la v.a.  $M_n$  comme combinaison linéaire de  $X_1, \dots, X_n$  suit une loi normale. En outre,  $\mathbb{E}(M_n) = m$  ( $M_n$  est sans biais) et puisque  $X_1, \dots, X_n$  sont indépendantes,

$$\text{Var}(M_n) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Ainsi,  $M_n$  suit une loi  $\mathcal{N}(m, \frac{\sigma^2}{n})$  et (1) est démontré.

2. Pour tout  $1 \leq i \leq n$ , notons  $Y_i = \frac{X_i - m}{\sigma}$ , de sorte que  $Y_1, \dots, Y_n$  sont  $n$  gaussiennes centrées réduites indépendantes. Notons que

$$\begin{aligned} \frac{n-1}{\sigma^2} V_n &= \sum_{i=1}^n \left( \frac{X_i - M_n}{\sigma} \right)^2 \\ &= \sum_{i=1}^n \left( Y_i - \frac{1}{n} \sum_{j=1}^n Y_j \right)^2 \\ &= \sum_{i=1}^n Y_i^2 - \frac{1}{n} \left( \sum_{i=1}^n Y_i \right)^2. \end{aligned} \quad (4.1)$$

Un petit peu de calcul matriciel et l'introduction d'une forme quadratique nous permettra d'y voir plus clair dans cette formule. Posons  $Z = (1, \dots, 1)^T$  (le symbole  $T$  est mis pour la transposition, de sorte que  $Z$  est une matrice à  $n$  lignes et 1 colonne). Soit  $A = Id - \frac{1}{n} ZZ^T$ , dont on notera les coef.  $a_{i,j}$ .  $ZZ^T$  est la matrice carré de taille  $n$  dont tous les coef. valent 1 :

$$ZZ^T = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \vdots & \vdots \\ 1 & \dots & 1 \end{pmatrix}$$

donc  $A$  est la matrice carré de taille  $n$  de termes diagonaux égaux à  $1 - \frac{1}{n}$  et de termes non diagonaux égaux à  $-\frac{1}{n}$ . La forme quadratique  $q_A$  associée à  $A$  est donc donnée par la formule :  $\forall x \in \mathbb{R}^n$ ,

$$\begin{aligned} q_A(x) &= \sum_{i,j} a_{i,j} x_i x_j \\ &= \sum_{i=1}^n x_i^2 - \frac{1}{n} \sum_{i,j=1}^n x_i x_j \\ &= \sum_{i=1}^n x_i^2 - \frac{1}{n} \left( \sum_{i=1}^n x_i \right)^2. \end{aligned}$$

En combinant les équations (4.1) et (4.2), on a  $\frac{n-1}{\sigma^2} V_n = q_A(Y_1, \dots, Y_n)$ . D'autre part, on a donc  $A^T = A$  et

$$\begin{aligned} A^2 &= \left( Id - \frac{1}{n} ZZ^T \right) \left( Id - \frac{1}{n} ZZ^T \right) \\ &= Id - \frac{2}{n} ZZ^T + \frac{1}{n^2} ZZ^T ZZ^T. \end{aligned}$$

Or  $Z^T Z$  est la matrice à 1 ligne et 1 colonne de coef  $1 + \dots + 1 = n$ . Donc

$$A^2 = Id - \frac{1}{n} Z^T Z = A.$$

Donc d'après la dernière proposition du chapitre Vecteurs Gaussiens,  $\frac{n-1}{\sigma^2} V_n = q_A(Y_1, \dots, Y_n)$  suit une loi  $\chi^2(r)$  où  $r = \sum_{i=1}^n a_{i,i} = n - 1$ .

Ainsi,  $\frac{n-1}{\sigma^2} V_n$  suit une loi du chi-deux à  $n - 1$  degrés de liberté, ce qui prouve (2).

3. Intéressons-nous à présent à la propriété (3). Posons  $G_i = X_i - M_n$  pour tout  $i$  compris entre 1 et  $n$ . En particulier,

$$V_n = \frac{1}{n-1} \sum_{i=1}^n G_i^2.$$

Pour montrer que  $M_n$  est indépendant de  $V_n$ , il suffit par conséquent de montrer que  $M_n$  est indépendant du vecteur  $G = (G_1, \dots, G_n)$ . Or, le vecteur aléatoire de  $\mathbb{R}^{n+1}$   $(M_n, G_1, \dots, G_n)$  est gaussien. D'après le chapitre Vecteurs Gaussiens, il suffit donc de montrer que  $\text{Cov}(M_n, G_i) = 0$  pour tout  $1 \leq i \leq n$ . Or

$$\begin{aligned} \text{Cov}(M_n, G_i) &= \text{Cov}(M_n, X_i - M_n) \\ &= \text{Cov}(M_n, X_i) - \text{Var}(M_n) \\ &= \frac{1}{n} \sum_{j=1}^n \text{Cov}(X_j, X_i) - \frac{\sigma^2}{n} \\ &= \frac{1}{n} \text{Var}(X_i) - \frac{\sigma^2}{n} \\ &= 0. \end{aligned}$$

Ainsi, (3) est démontrée.

4. Pour montrer (4), on note que

$$\sqrt{n} \frac{M_n - m}{\sqrt{V_n}} = \frac{\sqrt{n} \frac{M_n - m}{\sigma}}{\sqrt{\frac{1}{n-1} \left( \frac{n-1}{\sigma^2} V_n \right)}}.$$

Or d'après (1),

$$L = \sqrt{n} \frac{M_n - m}{\sigma} \stackrel{\text{loi}}{\sim} \mathcal{N}(0, 1)$$

et d'après (2),

$$K = \frac{n-1}{\sigma^2} V_n \stackrel{\text{loi}}{\sim} \chi^2(n-1).$$

En outre, d'après (3),  $L$  et  $K$  sont indépendantes. Par définition

$$\frac{L}{\sqrt{\frac{1}{n-1} K}} = \sqrt{n} \frac{M_n - m}{\sqrt{V_n}}$$

suit donc une loi de Student à  $n-1$  degrés de liberté.

□

### Intervalle de confiance pour $m$ lorsque $\sigma$ est connu

Nous supposons dans un premier temps connue la variance  $\sigma^2$  de  $X$  alors que l'espérance  $m$  est inconnue. On travaille donc avec le modèle paramétrique  $\{\mathcal{N}(m, \sigma^2) / m \in \mathbb{R}\}$ . Soit  $Y_n = \frac{\sqrt{n}(M_n - m)}{\sigma}$ . D'après la proposition 4.3.5, (1),

$$\mathcal{L}(Y_n) = \mathcal{N}(0, 1).$$

Fixons un niveau de confiance  $0 < \alpha < 1$ . Soit  $a = a_{1-\frac{\alpha}{2}}$  un quantile de niveau  $1 - \frac{\alpha}{2}$  de  $\mathcal{N}(0, 1)$ . Nous avons donc

$$\text{P}(\mathcal{N}(0, 1) \leq a) = 1 - \frac{\alpha}{2}.$$

De plus

$$\text{P}(\mathcal{N}(0, 1) \geq a) = \frac{\alpha}{2}$$

et par symétrie de la loi normale,

$$\begin{aligned} \text{P}(|\mathcal{N}(0, 1)| \leq a) &= 1 - \text{P}(\mathcal{N}(0, 1) \geq a) - \text{P}(\mathcal{N}(0, 1) \leq -a) \\ &= 1 - 2\text{P}(\mathcal{N}(0, 1) \geq a) \\ &= 1 - \alpha. \end{aligned}$$

Donc, nous avons également

$$\mathbb{P}(|Y_n| \leq a) = \mathbb{P}\left(-a \leq \frac{\sqrt{n}(M_n - m)}{\sigma} \leq a\right) = 1 - \alpha,$$

ce que nous pouvons écrire

$$\mathbb{P}\left(M_n - a \frac{\sigma}{\sqrt{n}} \leq m \leq M_n + a \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

On a donc montré la proposition suivante.

**Proposition 4.3.7** *En prenant a le quantile de niveau  $1 - \frac{\alpha}{2}$  de  $\mathcal{N}(0, 1)$ , les fonctions de  $\mathbb{R}^n$  vers  $\mathbb{R}$   $a_n(x) = \frac{1}{n} \sum_{i=1}^n x_i - a \frac{\sigma}{\sqrt{n}}$  et  $b_n(x) = \frac{1}{n} \sum_{i=1}^n x_i + a \frac{\sigma}{\sqrt{n}}$  définissent un intervalle de confiance  $[a_n, b_n]$  de niveau  $1 - \alpha$  pour estimer  $m$  dans le modèle paramétrique  $\{\mathcal{N}(m, \sigma^2)/m \in \mathbb{R}\}$ .*

### Intervalle de confiance pour $m$ lorsque $\sigma$ est inconnu

Abordons à présent le cas plus réaliste où l'écart-type  $\sigma$  est inconnu. On travaille donc avec le modèle paramétrique  $\{\mathcal{N}(m, \sigma^2)/m \in \mathbb{R}, \sigma > 0\}$ . Soient  $M_n$  et  $V_n$  respectivement l'estimateur de l'espérance et de la variance de  $X$  introduits précédemment. D'après la proposition 4.3.5, (4),

$$\text{loi} \left( \frac{\sqrt{n}(M_n - m)}{\sqrt{V_n}} \right) = t_{n-1} \text{ (loi de Student à } n - 1 \text{ degrés de liberté).}$$

Fixons un niveau de confiance  $0 < \alpha < 1$ . Soit  $a = a_{1-\frac{\alpha}{2}, n-1}$  un quantile de niveau  $1 - \frac{\alpha}{2}$  pour la loi de Student à  $n - 1$  degrés de liberté. Comme précédemment, on montre que

$$\mathbb{P}(|t_{n-1}| \leq a) = 1 - \alpha.$$

Ainsi,

$$\mathbb{P}\left(-a \leq \frac{\sqrt{n}(M_n - m)}{\sqrt{V_n}} \leq a\right) = 1 - \alpha$$

ce qui donne

$$\mathbb{P}\left(M_n - a \frac{\sqrt{V_n}}{\sqrt{n}} \leq m \leq M_n + a \frac{\sqrt{V_n}}{\sqrt{n}}\right) = 1 - \alpha.$$

On a donc montré la proposition suivante.

**Proposition 4.3.8** *En prenant a le quantile de niveau  $1 - \frac{\alpha}{2}$  de  $t_{n-1}$ ,  $m_n$  et  $v_n$  les fonctions de  $\mathbb{R}^n$  vers  $\mathbb{R}$  :  $m_n(x) = \frac{1}{n} \sum_{i=1}^n x_i$  et  $v_n(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - m_n(x))^2$ ,  $[m_n - a \frac{\sqrt{v_n}}{\sqrt{n}}, m_n + a \frac{\sqrt{v_n}}{\sqrt{n}}]$  est un intervalle de confiance de niveau  $1 - \alpha$  pour estimer  $m$  dans le modèle paramétrique  $\{\mathcal{N}(m, \sigma^2)/m \in \mathbb{R}, \sigma > 0\}$ .*

### Intervalle de confiance pour $\sigma$ lorsque $m$ est inconnu

On travaille encore avec le modèle paramétrique  $\{\mathcal{N}(m, \sigma^2)/m \in \mathbb{R}, \sigma > 0\}$ . Cherchons à présent un intervalle de confiance pour  $\sigma^2$ . D'après la proposition 4.3.5,  $\frac{n-1}{\sigma^2} V_n$  suit une loi  $\chi^2(n-1)$  (loi du chi-deux à  $n - 1$  degrés de liberté). Fixons un niveau de confiance  $0 < \alpha < 1$ . Soient alors  $a = a_{\alpha, n}$  et  $b = b_{\alpha, n}$  deux réels positifs avec  $a < b$  (on autorise le cas  $b = +\infty$ ) tels que

$$\mathbb{P}(a \leq \chi^2(n-1) \leq b) = 1 - \alpha.$$

On a alors

$$\mathbb{P}\left(a \leq \frac{n-1}{\sigma^2} V_n \leq b\right) = 1 - \alpha.$$

D'où

$$\mathbb{P}\left(\frac{(n-1)V_n}{b} \leq \sigma^2 \leq \frac{(n-1)V_n}{a}\right) = 1 - \alpha.$$

On a donc montré la proposition suivante.

**Proposition 4.3.9** *En prenant  $a < b \leq \infty$  tels que*

$$P(a \leq \chi^2(n-1) \leq b) = 1 - \alpha,$$

*$v_n$  la fonction de  $\mathbb{R}^n$  vers  $\mathbb{R}$  définie à la proposition (4.3.8),  $[\frac{(n-1)v_n}{b}, \frac{(n-1)v_n}{a}]$  est un intervalle de confiance de niveau  $1 - \alpha$  pour estimer  $\sigma^2$  dans le modèle paramétrique  $\{\mathcal{N}(m, \sigma^2)/m \in \mathbb{R}, \sigma > 0\}$ .*

**Remarque :** Si  $b = +\infty$ , ce qui signifie que  $P(\chi^2(n-1) \geq a) = 1 - \alpha$ , l'intervalle

$$\left[0, \frac{(n-1)v_n}{a}\right]$$

est un un intervalle de confiance de niveau  $1 - \alpha$  pour  $\sigma^2$ .

**Région de confiance pour le couple  $(\sigma^2, m)$**

On travaille encore avec le modèle paramétrique  $\{\mathcal{N}(m, \sigma^2)/m \in \mathbb{R}, \sigma^2 > 0\}$ . On dispose d'un  $n$ -échantillon  $X_1, \dots, X_n$  de v.a.i.i.d. de loi gaussienne  $\mathcal{N}(m, \sigma^2)$ . On voudrait maintenant trouver une région de  $\mathbb{R} \times \mathbb{R}^+$  où se trouve "vraisemblablement"  $(\sigma^2, m)$ . Plus précisément, étant donné  $\alpha \in ]0, 1[$  fixé, on voudrait trouver une partie de  $A$  de  $\mathbb{R}^+ \times \mathbb{R}$  (la plus petite possible, sinon ça n'a pas d'intérêt) telle que la probabilité que  $(\sigma^2, m)$  se trouve dans  $A$  soit égale à  $1 - \alpha$ .

**Définition 4.3.10** *Soit  $d \geq 1$  et  $\Theta$  une partie de  $\mathbb{R}^d$ . Soit  $\alpha \in ]0, 1[$ . Soit  $\{P_\theta/\theta \in \Theta\}$  un modèle paramétrique. Une suite de fonctions  $(A_n)_{n \geq 1}$  tels que pour tout  $n$*

$$A_n : \mathbb{R}^n \rightarrow \mathcal{P}(\Theta) \text{ (ensemble des parties de } \Theta)$$

*est une suite de régions de confiance de niveau  $1 - \alpha$  pour  $\theta$  si pour tout  $n$ , pour tout  $\theta \in \Theta$ , pour tout  $X_1, \dots, X_n$  v.a.i.i.d. de loi  $P_\theta$ ,*

$$P(\theta \in A_n(X_1, \dots, X_n)) = 1 - \alpha.$$

Choisissons  $\beta, \gamma \in ]0, 1[$  tels que  $1 - \alpha = (1 - \beta)(1 - \gamma)$ .  
Soit  $c$  un quantile de niveau  $1 - \frac{\beta}{2}$  de  $\mathcal{N}(0, 1)$ .  
Soit  $a < b \leq +\infty$  tels que

$$P(a \leq \chi^2(n-1) \leq b) = 1 - \gamma.$$

On sait tout d'abord que

$$P\left(-c \leq \frac{\sqrt{n}}{\sqrt{\sigma^2}}(M_n - m) \leq c\right) = 1 - \beta$$

i.e.

$$P\left(M_n - c \frac{\sqrt{\sigma^2}}{\sqrt{n}} \leq m \leq M_n + c \frac{\sqrt{\sigma^2}}{\sqrt{n}}\right) = 1 - \beta. \quad (4.2)$$

Soit, pour  $\lambda \in \mathbb{R}$  et  $\tau > 0$ ,  $\mathcal{C}_{\lambda, \tau}^+$  et  $\mathcal{C}_{\lambda, \tau}^-$  les courbes d'équations d'équations respectives  $y = \lambda + \tau\sqrt{x}$  et  $y = \lambda - \tau\sqrt{x}$  pour  $x \in [0, +\infty[$ . On remarque que  $\mathcal{C}_{\lambda, \tau}^+$  est située au dessus de  $\mathcal{C}_{\lambda, \tau}^-$ . L'équation (4.2) nous dit que la probabilité que  $(\sigma^2, m)$  soit dans le domaine situé entre les deux courbes  $\mathcal{C}_{M_n, \frac{c}{\sqrt{n}}}^+$  et  $\mathcal{C}_{M_n, \frac{c}{\sqrt{n}}}^-$  vaut  $1 - \alpha$ .

Par ailleurs, on a

$$P\left(a \leq \frac{n-1}{\sigma^2} V_n \leq b\right) = 1 - \gamma.$$

D'où

$$P\left(\frac{(n-1)V_n}{b} \leq \sigma^2 \leq \frac{(n-1)V_n}{a}\right) = 1 - \alpha, \quad (4.3)$$



i.e. la probabilité que  $(\sigma^2, m)$  aie son abscisse entre  $\frac{(n-1)V_n}{b}$  et  $\frac{(n-1)V_n}{a}$  vaut  $1 - \gamma$ .

En combinant les équations (4.2) et (4.3), et par indépendance de  $V_n$  et de  $M_n$  (proposition 4.3.5),

$$P\left(-c \leq \frac{\sqrt{n}}{\sigma^2}(M_n - m) \leq c \text{ et } a \leq \frac{n-1}{\sigma^2}V_n \leq b\right) = (1 - \beta)(1 - \gamma) = 1 - \alpha,$$

i.e. la probabilité que  $(\sigma^2, m)$  soit dans le domaine situé entre les deux courbes  $\mathcal{C}_{M_n, \frac{c}{\sqrt{n}}}^+$  et  $\mathcal{C}_{M_n, \frac{c}{\sqrt{n}}}^-$  et aie son abscisse entre  $\frac{(n-1)V_n}{b}$  et  $\frac{(n-1)V_n}{a}$  vaut  $1 - \alpha$ .

On a donc montré la proposition suivante.

**Proposition 4.3.11** *Soit, pour  $n \geq 1$ ,  $A_n : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^+ \times \mathbb{R})$  qui à  $x \in \mathbb{R}^n$  associe le domaine de  $\mathbb{R}^+ \times \mathbb{R}$  situé entre les courbes  $\mathcal{C}_{m_n(x), \frac{c}{\sqrt{n}}}^+$  et  $\mathcal{C}_{m_n(x), \frac{c}{\sqrt{n}}}^-$  et délimité par les droites verticales d'abscisses  $\frac{(n-1)v_n(x)}{b}$  et  $\frac{(n-1)v_n(x)}{a}$ . Alors  $(A_n)$  est une suite de régions de confiance de niveau  $1 - \alpha$  pour  $(\sigma^2, m)$  dans le modèle  $\{\mathcal{N}(m, \sigma^2)/m \in \mathbb{R}, \sigma^2 > 0\}$ .*

## 4.4 Intervalles de confiance asymptotiques

### 4.4.1 Définition

Les cas des modèles de Bernoulli et gaussiens sont très particuliers. En général, on ne possède pas de renseignements aussi précis sur la loi des estimateurs. En revanche, lorsque le nombre d'observations est très grand, on sait souvent que les lois des estimateurs sont proches de lois connues et souvent des lois gaussiennes par le théorème central limite. C'est ce principe qui est à la base du concept d'intervalles de confiance asymptotiques. L'autre intérêt des intervalles de confiance asymptotique est que, même pour le modèle de Bernoulli, il peut donner des résultats plus précis et plus pratiques à utiliser que ceux que l'on a déjà vu lorsque  $n$  est grand.

On suppose disposer d'une suite de variables aléatoires i.i.d.  $X_n, n \geq 0$  avec  $\mathcal{L}(X) \in \{\mathcal{P}_\theta ; \theta \in \Theta\}$ . On cherche à obtenir un intervalle de confiance pour une fonction de  $\theta$  notée  $Y(\theta)$ .

**Définition 4.4.1** *Soit  $0 < \alpha < 1$ .*

*Soit  $(a_n), (b_n)$  deux suites de fonctions de  $\mathbb{R}^n$  vers  $\mathbb{R}$ .*

*On dit que les intervalles  $[a_n, b_n]$  sont des intervalles de confiance asymptotiques de niveau  $1 - \alpha$  pour  $Y$  si et seulement si pour tout  $\theta \in \Theta$ , pour toute suite  $X_n$  de v.a.i.i.d. de loi  $\mathcal{P}_\theta$ ,*

$$P(a_n(X_1, \dots, X_n) \leq Y(\theta) \leq b_n(X_1, \dots, X_n)) \xrightarrow[n \rightarrow \infty]{} 1 - \alpha.$$

**Remarque :** On pourrait aussi définir les intervalles de confiance asymptotique par excès.

### 4.4.2 Intervalles de confiance asymptotiques pour l'espérance - Cas général

Considérons une suite de v.a. i.i.d.  $X_n, n \geq 1$  telle que  $X \in L^2$  et  $X$  est non constante. Posons  $m = \mathbb{E}(X)$  et  $\sigma^2 = \text{Var}(X)$ . Nous cherchons des intervalles de confiance asymptotiques pour  $m$  à partir de l'estimateur

$$M_n = \frac{1}{n} \sum_{i=1}^n X_i$$

et de l'estimateur sans biais de la variance de  $X$  donné par  $V_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2$ . Le théorème central limite permet d'affirmer que

$$\frac{\sqrt{n}(M_n - m)}{\sigma} \xrightarrow[n \rightarrow \infty]{\text{loi}} \mathcal{N}(0, 1).$$

Nous savons par ailleurs que

$$V_n \xrightarrow[n \rightarrow \infty]{} \sigma^2.$$

On peut assez facilement en déduire que l'on a aussi :

$$\frac{\sqrt{n}(M_n - m)}{\sqrt{V_n}} \xrightarrow[n \rightarrow \infty]{loi} \mathcal{N}(0, 1).$$

Fixons  $0 < \alpha < 1$ . Soit alors  $a = a_{1-\frac{\alpha}{2}}$  le quantile d'ordre  $1 - \frac{\alpha}{2}$  de  $\mathcal{N}(0, 1)$ , de sorte que

$$P(-a \leq \mathcal{N}(0, 1) \leq a) = 1 - \alpha.$$

Nous avons donc

$$\begin{aligned} P\left(M_n - \frac{a\sqrt{V_n}}{\sqrt{n}} \leq m \leq M_n + \frac{a\sqrt{V_n}}{\sqrt{n}}\right) &= P\left(-a \leq \frac{\sqrt{n}(M_n - m)}{\sqrt{V_n}} \leq a\right) \\ &\xrightarrow[n \rightarrow \infty]{} P(-a \leq \mathcal{N}(0, 1) \leq a) = 1 - \alpha. \end{aligned}$$

On a donc montré la proposition suivante.

**Proposition 4.4.2** *En prenant  $a$  le quantile d'ordre  $1 - \frac{\alpha}{2}$  de  $\mathcal{N}(0, 1)$ ,  $m_n$  et  $v_n$  les fonctions de  $\mathbb{R}^n$  vers  $\mathbb{R}$  définies à la proposition 4.3.8,  $[m_n - \frac{a\sqrt{v_n}}{\sqrt{n}}, m_n + \frac{a\sqrt{v_n}}{\sqrt{n}}]$  est un intervalle de confiance asymptotique de niveau  $1 - \alpha$  pour estimer  $m_\theta$  où pour tout  $\theta \in \Theta$ ,  $m_\theta$  est l'espérance d'une v.a. de loi  $P_\theta$ .*

Lorsque  $n$  est grand (en pratique, quand  $n \geq 30$ ), nous dirons (en faisant une approximation) qu'à  $100(1 - \alpha)\%$ ,  $m$  est dans l'intervalle  $[M_n - \frac{a\sqrt{V_n}}{\sqrt{n}}, M_n + \frac{a\sqrt{V_n}}{\sqrt{n}}]$ .

**Exercice 4.4.1** On veut estimer la taille moyenne des hommes adultes dans une population. Un échantillon de 100 individus a donné une moyenne de

$$\begin{aligned} \frac{1}{100} \sum_{\text{individus testés}} \text{taille} &= 175\text{cm}, \\ \frac{1}{99} \sum_{\text{individus testés}} (\text{taille} - 175)^2 &= 49\text{cm}^2. \end{aligned}$$

Donner un intervalle de confiance à 95% pour la taille moyenne de la population.

*Solution :* Soit  $c$  tel que  $P(-c \leq \mathcal{N}(0, 1) \leq c) = 0,95$ , i.e.  $c = 1,96$ . L'intervalle  $[175 - \frac{1,96\sqrt{49}}{\sqrt{100}}, 175 + \frac{1,96\sqrt{49}}{\sqrt{100}}]$  convient.

#### 4.4.3 Intervalles de confiance asymptotiques pour l'espérance - Modèle de Bernoulli

Un cas particulier fréquent est celui des modèles de Bernoulli. On suppose que  $\mathcal{L}(X) \in \{\mathcal{B}(\theta) ; 0 < \theta < 1\}$ . On garde les mêmes définitions pour  $M_n$  et  $V_n$ .

On a la convergence en loi, pour  $p$  paramètre de la loi des  $X_i$ ,

$$\frac{\sqrt{n}(M_n - p)}{\sqrt{V_n}} \xrightarrow[n \rightarrow \infty]{loi} \mathcal{N}(0, 1).$$

En utilisant le fait que  $X_i^2 = X_i$ , on a  $V_n = \frac{n}{n-1}M_n(1 - M_n)$ .

On en déduit que l'on a aussi la convergence en loi

$$\frac{\sqrt{n}(M_n - p)}{\sqrt{M_n(1 - M_n)}} \xrightarrow[n \rightarrow \infty]{loi} \mathcal{N}(0, 1).$$

Si on fixe  $0 < \alpha < 1$  et si  $a = a_{1-\frac{\alpha}{2}}$  est le quantile d'ordre  $1 - \frac{\alpha}{2}$  de  $\mathcal{N}(0, 1)$ , alors

$$P\left(p \in \left[M_n - \frac{a\sqrt{M_n(1-M_n)}}{\sqrt{n}}, M_n + \frac{a\sqrt{M_n(1-M_n)}}{\sqrt{n}}\right]\right) \xrightarrow{n \rightarrow \infty} 1 - \alpha.$$

Notons que

$$x(1-x) \leq \frac{1}{2} \left(1 - \frac{1}{2}\right) = \frac{1}{4},$$

pour tout  $0 \leq x \leq 1$ , si bien que

$$\left[M_n - \frac{a\sqrt{M_n(1-M_n)}}{\sqrt{n}}, M_n + \frac{a\sqrt{M_n(1-M_n)}}{\sqrt{n}}\right] \subset \left[M_n - \frac{a}{2\sqrt{n}}, M_n + \frac{a}{2\sqrt{n}}\right].$$

On a donc montré la proposition suivante.

**Proposition 4.4.3** *En prenant  $a$  le quantile d'ordre  $1 - \frac{\alpha}{2}$  de  $\mathcal{N}(0, 1)$   $m_n$  la fonction de  $\mathbb{R}^n$  vers  $\mathbb{R}$  définie à la proposition (4.3.8),  $[m_n - \frac{a}{2\sqrt{n}}, m_n + \frac{a}{2\sqrt{n}}]$  est un intervalle de confiance asymptotique de niveau  $1 - \alpha$  pour estimer  $\theta$  dans le modèle paramétrique  $\{\mathcal{B}(\theta) ; 0 < \theta < 1\}$ .*

Par exemple, si  $\alpha = 5\%$ ,  $a = 1,96$  et à  $95\%$ , on pourra dire que  $p$  est dans l'intervalle  $[\hat{p}_n - \frac{0,98}{\sqrt{n}}, \hat{p}_n + \frac{0,98}{\sqrt{n}}]$ . Si  $n = 1000$ ,  $\frac{0,98}{\sqrt{n}} \simeq 3\%$ . L'erreur due à l'estimation est au plus de l'ordre de  $3\%$ .

**Exercice 4.4.2** Sur le site <http://www.csa-tmo.fr/fra/dataset/data2003/opi20030321a.htm>, on peut lire que lors d'un sondage réalisé en France par le CSA pour le journal Le Parisien en mars 2003 que sur 1000 personnes de 18 ans et plus choisies au hasard, 750 ont répondu "non" à la question "Pensez-vous que l'intervention américaine en Irak est de nature à ramener la paix et la stabilité dans cette région du monde?". Donner un intervalle de confiance à  $95\%$  de la proportion des habitants de la France majeurs qui auraient répondu "oui". On pourra utiliser le fait que pour  $c = 1,96$ ,  $P(-c \leq \mathcal{N}(0, 1) \leq c) = 0,95$ .

**Exercice 4.4.3** Refaire l'exercice 4.3.1 à la lumière de ce qu'on vient de faire, et comparer les réponses. Commenter.

*Solution :* Soit  $\theta \in ]0, 1[$  la proportion de la population totale favorable au "oui". Quand on interroge  $n$  personnes prises au hasard, indépendamment, comme c'est le cas dans les sondages, en notant 1 lorsque la réponse est "oui" et 0 lorsque la réponse est "non", on observe  $n$  v.a.i.i.d. de loi  $\mathcal{B}(\theta)$ . Comme pour  $c = 1,96$ ,  $P(-c \leq \mathcal{N}(0, 1) \leq c) = 0,95$ , on sait, par ce qui précède, que si  $x_1, \dots, x_n$  sont les réponses successives, on a, pour  $n$  grand,

$$P\left(\frac{1}{n} \sum_{i=1}^n x_i - \frac{c}{2\sqrt{n}} \leq \theta \leq \frac{1}{n} \sum_{i=1}^n x_i + \frac{c}{2\sqrt{n}}\right) \simeq 0,95 (= 19/20).$$

On veut donc  $n$  assez grand tel que  $\frac{c}{2\sqrt{n}} \leq 0,03$ , soit  $n \geq \frac{1,96^2}{4 \cdot 0,03^2} = 1067,1$ . Il faut donc  $n \geq 1068$ , ce qui est inférieur à la borne précédente. Cette estimation est plus précise pour de grandes valeurs de  $n$ .

**Exercice 4.4.4** Sur le site <http://pollingreport.com/religion.htm>, on peut lire que lors d'un sondage réalisé aux USA en 2003 pour la chaîne de télévision nationale FOX News que sur 900 citoyens américains votant choisis au hasard dans les Etats Unis, 639 ont répondu oui à la question "Croyez-vous personnellement au diable" (aussi impressionnante que soit la proportion, ce sondage est très sérieux). Donner un intervalle de confiance à  $95\%$  pour la proportion des citoyen américains croyant au diable. Avec quelle probabilité peut-on dire que cette proportion est située entre  $67\%$  et  $75\%$  ?

## 4.5 Tests

Nous nous intéressons à présent aux problèmes de **tests d'hypothèses**.

### 4.5.1 Formulation

Nous disposons d'un  $n$ -échantillon  $(X_1, \dots, X_n)$ . Les variables aléatoires  $X_1, \dots, X_n$  sont donc i.i.d. de loi commune  $\mathcal{L}(X)$ . Nous supposons avoir un modèle paramétrique de telle sorte que

$$\mathcal{L}(X) \in \{P_\theta ; \theta \in \Theta\}.$$

Nous cherchons à tester une hypothèse faite a priori sur la loi de  $X$ , c'est-à-dire en fait sur  $\theta$ , à partir d'une observation  $(x_1, \dots, x_n)$ . Plus précisément, on fixe un ensemble  $\Theta_0 \subset \Theta$  et on veut "tester" l'hypothèse

$$H_0 : \theta \in \Theta_0$$

contre son alternative

$$H_1 : \theta \in \Theta_1 := \Theta \setminus \Theta_0.$$

Nous voudrions une "règle" qui, en fonction de  $(x_1, \dots, x_n)$  nous dise d'accepter  $H_0$  ou de rejeter  $H_0$ , autrement dit d'accepter  $H_1$ .

En général, le problème sera résolu de la manière suivante. On fixe un **seuil de tolérance**  $0 < \alpha < 1$  ( $\alpha$  est en général petit, par exemple  $\alpha = 10, 5, 1\%$ ). Soit une fonction  $Y_n : \mathbb{R}^n \rightarrow \mathbb{R}$  (si  $\theta \in \mathbb{R}$ ,  $Y_n$  est souvent un estimateur de  $\theta$ ) et  $D_n$  une fonction de  $\mathbb{R}^n$  vers l'ensemble des parties de  $\mathbb{R}$ . Dans ce cadre, un estimateur sera aussi appelée une **statistique**. On se propose d'accepter  $H_0$  si  $Y_n(X_1, \dots, X_n)$  appartient à un sous-ensemble  $D_n(X_1, \dots, X_n)$  bien choisi de  $\mathbb{R}$  (la région de rejet) et de la rejeter si  $Y_n(X_1, \dots, X_n) \notin D_n(X_1, \dots, X_n)$ . Pour un tel test, le **niveau du test** est la quantité

$$\alpha = \sup_{\theta \in \Theta_0} P \left( Y_n(X_1^{(\theta)}, \dots, X_n^{(\theta)}) \notin D_n(X_1^{(\theta)}, \dots, X_n^{(\theta)}) \right),$$

où pour tout  $\theta \in \Theta$ ,  $X_1^{(\theta)}, \dots, X_n^{(\theta)}$  sont des v.a.i.i.d. de loi  $P_\theta$ .

Cette quantité est le plus petit majorant de la probabilité de rejeter  $H_0$  lorsque  $H_0$  a effectivement lieu. Avec ce test, si l'hypothèse est vérifiée, on se trompera, i.e. on rejettera l'hypothèse alors qu'elle a effectivement lieu avec une probabilité inférieure à  $\alpha(D_n)$ . Ainsi, le statisticien fait en sorte de choisir  $D_n$  tel que le niveau du test soit égal au seuil de tolérance, c'est-à-dire que

$$\alpha(D_n) = \alpha,$$

ceci afin que le niveau du test reste faible.

Un tel estimateur  $Y_n$  et une telle région  $D_n$  étant choisis, on appelle **puissance du test** la fonction définie sur  $\Theta_1 = \Theta \setminus \Theta_0$  par

$$\pi(\theta) = P \left( Y_n(X_1^{(\theta)}, \dots, X_n^{(\theta)}) \notin D_n(X_1^{(\theta)}, \dots, X_n^{(\theta)}) \right)$$

où pour tout  $\theta \in \Theta$ ,  $X_1^{(\theta)}, \dots, X_n^{(\theta)}$  sont des v.a.i.i.d. de loi  $P_\theta$ .

Autrement dit,  $\pi_{D_n}(\theta)$  est la probabilité de rejeter l'hypothèse  $H_0$  lorsque  $H_0$  est fautive. Il est donc très souhaitable que la puissance soit grande. A l'extrême limite, si on accepte systématiquement l'hypothèse  $H_0$ , on a un test de niveau 0 (c'est bien) mais de puissance 0 (ce qui hote tout intérêt au test).

Il est souhaitable mais difficile d'avoir un test à la fois de niveau bas et de puissance haute.

Il faut bien comprendre que **les hypothèses  $H_0$  et  $H_1$  ne jouent pas le même rôle**. Lorsqu'on teste  $H_0$  contre  $H_1$ , on veut avoir une faible probabilité de rejeter l'hypothèse  $H_0$  alors qu'elle est vraie (cette probabilité est au plus égale au **niveau du test**), mais on ne peut pas en même temps exiger d'avoir une faible probabilité de retenir  $H_0$  alors qu'elle est fautive (cette probabilité est donnée par la **puissance du test**). Voici un exemple simple où apparaît l'asymétrie entre les

rôles de  $H_0$  et de  $H_1$ . Les parents d'un nouveau né trouvent un matin leur enfant avec un tout petit peu de fièvre. Ils ne savent pas si c'est dû au fait qu'il est en train de "faire une dent" (auquel cas il n'y a pas à s'alarmer) ou si leur enfant a attrapé une petite maladie (auquel cas il faut le soigner). Ils vont alors tester l'hypothèse *il est malade* contre *il fait sa dent* et non l'inverse. En effet, ils ne veulent pas ne pas s'apercevoir qu'il est malade, mais ça n'est pas grave s'ils l'emmènent chez le médecin pour rien. Tester l'hypothèse *il fait sa dent* contre *il est malade* aurait revenu à tenir à ne pas aller chez le médecin pour rien, au prix éventuel d'une maladie non décelée chez l'enfant.

**Exercice 4.5.1** Dans la campagne présidentielle espagnole de 2004, le président Aznar avait fait de la lutte contre les terroristes indépendantistes basques de l'ETA un des piliers de son argumentaire. Le 11 mars 2004, peu de temps avant les élections, des attentats font de nombreuses victimes à Madrid. Pendant une courte période, l'origine de ces attentats est restée incertaine. Lequel du président Aznar ou de son rival Zapatero avait intérêt à tester l'hypothèse "l'ETA est à l'origine de ces attentats" contre "l'ETA n'est pas à l'origine de ces attentats", et lequel avait intérêt à faire le test inverse ?

**Exercice 4.5.2** Lors de la campagne présidentielle américaine 2004, lequel des deux candidats avait intérêt à tester l'hypothèse "la guerre en Irak a rendu le monde plus sûr" contre "la guerre en Irak a rendu le monde moins sûr", et lequel avait intérêt à faire le test inverse ?

## 4.5.2 Lien avec les intervalles de confiance

**Proposition 4.5.1** Soit  $\{P_\theta/\theta \in \Theta\}$  un modèle paramétrique. Soit  $Y$  une fonction de  $\Theta$  vers  $\mathbb{R}$ . Soit  $y_0 \in \mathbb{R}$ . On veut tester l'hypothèse  $H_0 : Y(\theta) = y_0$ , contre  $H_1 : Y(\theta) \neq y_0$ . Soit, pour  $n \geq 1$ ,  $a_n$  et  $b_n$  des fonctions de  $\mathbb{R}^n$  vers  $\mathbb{R}$  telles que  $[a_n, b_n]$  est un intervalle de confiance de niveau  $1 - \alpha$  par excès pour  $Y$ .

Alors le test défini par : on retient  $H_0$  si  $y_0 \in [a_n(X_1, \dots, X_n), b_n(X_1, \dots, X_n)]$  et on la rejette sinon est un test de niveau au moins  $\alpha$  de  $H_0$  contre  $H_1$ .

*Preuve :* Soit  $\theta$  tel que  $Y(\theta) = y_0$ . Soit  $X_1, \dots, X_n$  des v.a.i.i.d. de loi  $P_\theta$ . On a alors

$$P(y_0 \in [a_n(X_1, \dots, X_n), b_n(X_1, \dots, X_n)]) = P(Y(\theta) \in [a_n(X_1, \dots, X_n), b_n(X_1, \dots, X_n)]) \geq 1 - \alpha.$$

□

**Remarque :** La puissance de ce test est la fonction définie sur  $\Theta \setminus Y^{-1}(\{y_0\})$  par

$$\pi(\theta) = P\left(y_0 \notin [a_n(X_1^{(\theta)}, \dots, X_n^{(\theta)}), b_n(X_1, \dots, X_n)]\right)$$

où pour tout  $\theta \in \Theta$ ,  $X_1^{(\theta)}, \dots, X_n^{(\theta)}$  sont des v.a.i.i.d. de loi  $P_\theta$ .

**Exercice 4.5.3** Donner un test pour l'hypothèse  $\theta = \theta_0$  (contre son alternative) dans le modèle de Bernouilli  $\{\mathcal{B}(\theta)/0 < \theta < 1\}$ , dans le cas où  $n$  est considéré comme grand, et dans le cas  $n$  quelconque.

*Solution :* C'est une application immédiate de la proposition précédente, et de la proposition 4.4.3 dans le cas où  $n$  est considéré comme grand, et du corollaire 4.3.4 dans le cas  $n$  quelconque.

**Exercice 4.5.4** Donner un test pour l'hypothèse  $\theta \leq \theta_0$  (contre son alternative) dans le modèle de Bernouilli  $\{\mathcal{B}(\theta)/0 < \theta < 1\}$ , dans le cas où  $n$  est considéré comme grand.

*Solution :* On sait, par le paragraphe 4.4.3, que pour  $n$  grand, la v.a.  $\frac{\sqrt{n}(M_n - \theta)}{\sqrt{M_n(1 - M_n)}}$  suit approximativement une loi  $\mathcal{N}(0, 1)$ . Or si  $\theta \leq \theta_0$ , cette v.a. est  $\geq$  à  $\frac{\sqrt{n}(M_n - \theta_0)}{\sqrt{M_n(1 - M_n)}}$ , donc pour tout  $a \in \mathbb{R}$ ,

$$P\left(\frac{\sqrt{n}(M_n - \theta)}{\sqrt{M_n(1 - M_n)}} \geq a\right) \geq P\left(\frac{\sqrt{n}(M_n - \theta_0)}{\sqrt{M_n(1 - M_n)}} \geq a\right),$$

i.e., modulo notre approximation,

$$P(\mathcal{N}(0, 1) \geq a) \geq P\left(\frac{\sqrt{n}(M_n - \theta_0)}{\sqrt{M_n(1 - M_n)}} \geq a\right).$$

Si  $a > 0$  est un quantile de niveau au moins  $1 - \alpha$  de  $\mathcal{N}(0, 1)$ , on a donc

$$\alpha \geq P\left(\frac{\sqrt{n}(M_n - \theta_0)}{\sqrt{M_n(1 - M_n)}} \geq a\right),$$

ce qui s'écrit aussi

$$\alpha \geq P\left(M_n \geq \theta_0 + \frac{a\sqrt{M_n(1 - M_n)}}{\sqrt{n}}\right).$$

Or  $M_n \in [0, 1]$  et  $a > 0$ , donc  $a\sqrt{M_n(1 - M_n)} \leq a/2$ , donc  $M_n$  a plus de chances d'être  $\geq$  à  $\theta_0 + \frac{a\sqrt{M_n(1 - M_n)}}{\sqrt{n}}$  que à  $\theta_0 + \frac{a}{2\sqrt{n}}$ . Ainsi,

$$\alpha \geq P\left(M_n \geq \theta_0 + \frac{a}{2\sqrt{n}}\right),$$

et en acceptant  $\theta \leq \theta_0$  si

$$M_n \geq \theta_0 + \frac{a}{2\sqrt{n}},$$

on a un test de niveau au moins  $\alpha$  de  $\theta \leq \theta_0$  contre  $\theta > \theta_0$ .

**Exercice 4.5.5** Sur le site <http://pollingreport.com/guns.htm>, on peut lire que lors d'un sondage réalisé aux USA en 2004 pour The Gallup Poll que sur 1012 américains adultes choisis au hasard dans les Etats Unis, 385 ont répondu oui à la question "Avez-vous une arme à votre domicile?" (il convient néanmoins ce relativiser ce chiffre impressionnant : ces armes peuvent provenir d'héritages, et n'avoir pas été acquises par leur propriétaires eux mêmes, de la même façon que de nombreux français possèdent un fusil de chasse ayant appartenu à un aïeul). Testez à niveau 5%, à partir de ces données, l'hypothèse "Plus d'un tiers des américains adultes ont une arme chez eux" contre l'hypothèse inverse. Quel est le plus grand  $p$  tel que le test de niveau 5% de l'hypothèse "Plus de  $p\%$  des américains adultes ont une arme chez eux" contre l'hypothèse inverse a pour réponse "Plus de  $p\%$  des américains adultes ont une arme chez eux"?

*Solution* : C'est une application facile de l'exercice précédent.

### 4.5.3 L'exemple des modèles gaussiens

L'exemple le plus classique concerne le cas important des modèles gaussiens. Considérons dans un premier temps un échantillon gaussien  $(X_1, \dots, X_n)$  avec  $\mathcal{L}(X) \in \{\mathcal{N}(m, \sigma^2) ; m \in \mathbb{R}, \sigma > 0\}$ . Nous allons voir comment on peut tester différentes hypothèses sur les paramètres  $m$  et  $\sigma$  du modèle, paramètres supposés inconnus. L'hypothèse que nous cherchons à tester sera toujours notée  $H_0$ . Les estimateurs classiques de l'espérance et de la variance de  $X$  seront notés  $M_n$  et  $V_n$ . La probabilité  $P_{(m, \sigma)}$  désignera la probabilité sous laquelle  $X$  suit une loi normale  $\mathcal{N}(m, \sigma^2)$ . Nous fixons un seuil  $0 < \alpha < 1$ . les fonctions  $\mathbf{m}_n$  et  $\mathbf{v}_n$ , de  $\mathbb{R}^n$  vers  $\mathbb{R}$  sont celles définies à la proposition 4.3.8.

1.  $H_0 : m = m_0$

Il s'agit donc d'une hypothèse sur l'espérance de la loi recherchée. D'après la proposition 4.3.5, (4), la statistique

$$U_n = \frac{\sqrt{n}(M_n - m_0)}{\sqrt{V_n}}$$

suit, par la proposition 4.3.5, une loi de Student à  $n - 1$  degrés de liberté sous  $P_{(m_0, \sigma)}$ , pour tout  $\sigma > 0$  (i.e. si  $H_0$  est vérifiée). Soit  $t_{1-\frac{\alpha}{2}}$  un quantile d'ordre  $1 - \frac{\alpha}{2}$  de la loi de Student à  $n - 1$  degrés de liberté, de sorte que pour tout  $\sigma > 0$ ,

$$P_{(m_0, \sigma)}(|U_n| > t_{1-\frac{\alpha}{2}}) = \alpha.$$

En définissant  $u_n$  la fonction de  $\mathbb{R}^n$  vers  $\mathbb{R}$

$$u_n = \frac{\sqrt{n}(m_n - m_0)}{\sqrt{v_n}},$$

étant donnée une observation  $(x_1, \dots, x_n)$ , le **test de Student** consiste à rejeter  $H_0$  (et donc à accepter  $H_1$ ) si  $|u_n(x_1, \dots, x_n)| > t_{1-\frac{\alpha}{2}}$  et à accepter  $H_0$  sinon. C'est bien un test de niveau  $\alpha$ .

**Remarque :** Ce test est le même que celui donné par la proposition (4.5.1) et l'intervalle de confiance pour  $m$  dans le modèle  $\{\mathcal{N}(m, \sigma^2) ; m \in \mathbb{R}, \sigma > 0\}$  lorsque  $\sigma^2$  est inconnu.

2.  $H_0 : m \geq m_0$ .

Nous allons utiliser la même statistique pour ce test, i.e.

$$U_n = \frac{\sqrt{n}(M_n - m_0)}{\sqrt{V_n}}.$$

Si  $m \geq m_0$ ,

$$U_n \geq Z_n := \frac{\sqrt{n}(M_n - m)}{\sqrt{V_n}}.$$

Ainsi, pour tout  $t \in \mathbb{R}$ , tout  $m \geq m_0$  et tout  $\sigma > 0$ ,

$$P_{(m, \sigma)}(U_n < t) \leq P_{(m, \sigma)}(Z_n < t).$$

Par ailleurs, par la proposition 4.3.5,  $Z_n$  suit dans tous les cas une loi de Student à  $n - 1$  degrés de liberté. Soit donc  $t_\alpha$  un quantile de niveau  $1 - \alpha$  de la loi de Student à  $n - 1$  degrés de liberté. Pour tout  $m \geq m_0$  et tout  $\sigma > 0$ ,

$$P_{(m, \sigma)}(U_n < t_\alpha) \leq P_{(m, \sigma)}(Z_n < t_\alpha) = \alpha.$$

Par ailleurs, si  $\sigma > 0$ ,

$$P_{(m_0, \sigma)}(U_n < t_\alpha) = P_{(m_0, \sigma)}(Z_n < t_\alpha) = \alpha.$$

En résumé,

$$\sup_{m \geq m_0} P_{(m, \sigma)}(U_n < t_\alpha) = \alpha.$$

On conserve la même définition de la fonction  $u_n$ . Etant donnée une observation  $(x_1, \dots, x_n)$ , nous rejeterons donc  $H_0$  lorsque  $u_n(x_1, \dots, x_n) < t_\alpha$  et nous acceptons  $H_0$  sinon. C'est bien un test de niveau  $\alpha$ .

3.  $H_0 : m \leq m_0$ .

Ce cas est très similaire au précédent. Nous n'entrerons pas dans les détails. Nous utilisons la même statistique  $U$  que ci-dessus. Soit  $t_{1-\alpha}$  un quantile d'ordre  $1 - \alpha$  de la loi de Student à  $n - 1$  degrés de liberté. Etant donnée une observation  $(x_1, \dots, x_n)$ , si  $u_n(x_1, \dots, x_n) > t_{1-\alpha}$ , on rejette  $H_0$ . Sinon, on accepte  $H_0$ . C'est bien un test de niveau  $\alpha$ .

4.  $H_0 : \sigma = \sigma_0$ .

Introduisons la statistique

$$W_n = \frac{(n-1)V_n}{\sigma_0^2}.$$

D'après la proposition 4.3.5, (2), pour tout  $m \in \mathbb{R}$ , la statistique  $W_n$  suit sous  $P_{(m, \sigma_0)}$  une loi du chi-deux à  $n - 1$  degrés de liberté. Soit donc  $0 < a_\alpha < b_\alpha \leq +\infty$  tels que

$$P(a_\alpha \leq \chi^2(n-1) \leq b_\alpha) = 1 - \alpha.$$

Alors si  $\sigma = \sigma_0$ ,

$$P(a_\alpha \leq W_n \leq b_\alpha) = 1 - \alpha,$$

En définissant  $w_n$  la fonction de  $\mathbb{R}^n$  vers  $\mathbb{R}$

$$w_n = \frac{(n-1)v_n}{\sigma_0^2},$$

étant donnée une observation  $(x_1, \dots, x_n)$ , le test consiste à rejeter  $H_0$  (et donc à accepter  $H_1$ ) si  $a_\alpha \leq w_n(x_1, \dots, x_n) \leq b_\alpha$  et à accepter  $H_0$  sinon. C'est bien un test de niveau  $\alpha$ .

**Remarque :** Ce test est le même que celui donné par la proposition 4.5.1 et l'intervalle de confiance pour  $\sigma^2$  dans le modèle  $\{\mathcal{N}(m, \sigma^2) ; m \in \mathbb{R}, \sigma > 0\}$ .

5.  $H_0 : \sigma \geq \sigma_0$ .

Nous allons utiliser la même statistique pour ce test, i.e.

$$W_n = \frac{(n-1)V_n}{\sigma_0^2}.$$

Si  $\sigma \geq \sigma_0$ ,

$$W_n \geq R_n := \frac{(n-1)V_n}{\sigma^2}.$$

Ainsi, pour tout  $t \in \mathbb{R}$ , tout  $m \in \mathbb{R}$  et tout  $\sigma \geq \sigma_0$ ,

$$P_{(m, \sigma)}(R_n \geq t) \leq P_{(m, \sigma)}(W_n \geq t).$$

Par ailleurs, par la proposition 4.3.5,  $R_n$  suit dans tous les cas une loi du chi-deux à  $n - 1$  degrés de liberté. Soit donc  $t_\alpha$  un quantile d'ordre  $\alpha$  de la loi du chi-deux à  $n - 1$  degrés de liberté. Pour tout  $m$  et tout  $\sigma \geq \sigma_0$ ,

$$1 - \alpha = P_{(m, \sigma)}(R_n \geq t_\alpha) \leq P_{(m, \sigma)}(W_n \geq t_\alpha).$$

Par ailleurs, si  $m \in \mathbb{R}$ ,

$$P_{(m, \sigma_0)}(R_n \geq t_\alpha) = P_{(m, \sigma_0)}(W_n \geq t_\alpha) = 1 - \alpha.$$

En résumé,

$$\inf_{\sigma \geq \sigma_0} P_{(m, \sigma)}(W_n \geq t_\alpha) = 1 - \alpha.$$

On conserve la même définition de la fonction  $w_n$ . Etant donnée une observation  $(x_1, \dots, x_n)$ , nous rejeterons donc  $H_0$  lorsque  $w_n(x_1, \dots, x_n) \geq t_\alpha$  et nous acceptons  $H_0$  sinon. C'est bien un test de niveau  $\alpha$ .

6.  $H_0 : \sigma \leq \sigma_0$ .

Nous allons utiliser les mêmes statistiques pour ce test, i.e.

$$W_n = \frac{(n-1)V_n}{\sigma_0^2} \text{ et } R_n := \frac{(n-1)V_n}{\sigma^2}.$$

Si  $\sigma \leq \sigma_0$ ,

$$W_n \leq R_n.$$



Ainsi, pour tout  $t \in \mathbb{R}$ , tout  $m \in \mathbb{R}$  et tout  $\sigma \leq \sigma_0$ ,

$$P_{(m,\sigma)}(R_n \leq t) \leq P_{(m,\sigma)}(W_n \leq t).$$

Par ailleurs, par la proposition 4.3.5,  $R_n$  suit dans tous les cas une loi du chi-deux à  $n - 1$  degrés de liberté. Soit donc  $t_{1-\alpha}$  un quantile d'ordre  $1 - \alpha$  de la loi du chi-deux à  $n - 1$  degrés de liberté. Pour tout  $m$  et tout  $\sigma \geq \sigma_0$ ,

$$1 - \alpha = P_{(m,\sigma)}(R_n \leq t_{1-\alpha}) \leq P_{(m,\sigma)}(W_n \leq t_{1-\alpha}).$$

Par ailleurs, si  $m \in \mathbb{R}$ ,

$$P_{(m,\sigma_0)}(R_n \leq t_{1-\alpha}) = P_{(m,\sigma_0)}(W_n \leq t_{1-\alpha}) = 1 - \alpha.$$

En résumé,

$$\inf_{\sigma \geq \sigma_0} P_{(m,\sigma)}(W_n \leq t_{1-\alpha}) = 1 - \alpha.$$

On conserve la même définition de la fonction  $w_n$ . Etant donnée une observation  $(x_1, \dots, x_n)$ , nous rejeterons donc  $H_0$  lorsque  $w_n(x_1, \dots, x_n) \leq t_{1-\alpha}$  et nous acceptons  $H_0$  sinon. C'est bien un test de niveau  $\alpha$ .

## 4.6 Modèle de regression simple

### 4.6.1 Introduction

Une usine manufacture une certaine pièce du moteur de voitures, et la comande (donc la production) pour cette pièce varie d'année en année. On aimerait trouver une relation entre le nombre de pièces produites et le nombre d'heures de travail. On numérote dix années successives de 1 à 10, à chaque année  $i$  correspond une production de  $x_i$  pièces, qui a nécessité un travail de  $Y_i$  heures. Voici ces données :

$i =$	1	2	3	4	5	6	7	8	9	10
$x_i =$	30	20	60	80	40	50	60	30	40	60
$Y_i =$	73	50	128	170	87	108	130	74	88	139

Un graphe faisant apparaître les dix points de coordonnées  $(x_i, Y_i)$  donne l'impression qu'ils sont disséminés le long d'une droite (non verticale), mais pas tout à fait sur la droite (par exemple, les points correspondant à  $i = 1$  et 8, de même abscisse mais d'ordonnée différente, ne peuvent appartenir à la même droite).

Dans cette situation, où les  $Y_i$  ont toutes les chances d'être aléatoires, on pourra, dans une modélisation, supposer que la relation liant les  $x_i$  et les  $Y_i$  est du type :

$$\forall i, Y_i = ax_i + b + U_i,$$

où  $a$  et  $b$  sont des constantes, et les  $U_i$  sont des v.a.i.i.d. de loi  $\mathcal{N}(0, \sigma^2)$ .

Il faut ensuite déterminer  $a, b$  et  $\sigma^2$ , avec des estimateurs ou des intervalles de confiance, et puis on fera des tests sur leurs valeurs.

### 4.6.2 Théorie

Avertissement : ce passage est très calculatoire. Les calculs pourraient être éclaircis par une vision géométrique des choses, mais cette approche géométrique requière la connaissance de la théorie des espaces euclidiens.

Une famille de variables deterministes (i.e. non aléatoires)  $x_1, \dots, x_n$  est liée à une famille  $Y_1, \dots, Y_n$  de v.a. par la relation

$$\forall k \in \{1, \dots, n\}, Y_k = ax_k + b + U_k$$

où  $a, b \in \mathbb{R}$  et les  $U_k$  sont des v.a.i.i.d. de loi  $\mathcal{N}(0, \sigma^2)$ .

Les  $x_i$  sont supposées connues, non toutes égales, et fixées une fois pour toutes. Le problème est d'estimer  $a, b$  et  $\sigma^2$  à partir des observations  $y_1, \dots, y_n$  de  $Y_1, \dots, Y_n$ .

**Définition 4.6.1** On appelle estimateurs des moindres carrés les fonctions  $\hat{b}, \hat{a}$  de  $\mathbb{R}^n$  vers  $\mathbb{R}$  définies pour  $y \in \mathbb{R}^n$ , par

$$\begin{aligned}\hat{a}(y) &= \frac{\sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})}{\sum_{k=1}^n (x_k - \bar{x})^2} \\ \hat{b}(y) &= \bar{y} - \hat{a}(y)\bar{x} \\ \text{où } \bar{x} &= \frac{1}{n} \sum_{k=1}^n x_k \\ \text{et } \bar{y} &= \frac{1}{n} \sum_{k=1}^n y_k\end{aligned}$$

On définit aussi la fonction de  $\mathbb{R}^n$  vers  $\mathbb{R}$

$$\hat{\sigma}^2(y) = \frac{1}{n-2} \sum_{k=1}^n \left( y_k - \hat{b}(y) - \hat{a}(y)x_k \right)^2.$$

**Théorème 4.6.2** Pour tout  $y \in \mathbb{R}^n$ , le couple  $(\hat{b}(y), \hat{a}(y))$  réalise l'unique minimum, sur  $\mathbb{R}^2$ , de la fonction

$$Q(\alpha, \beta) = \sum_{k=1}^n (y_k - \alpha x_k - \beta)^2.$$

Ce résultat a été démontré dans un des TD du chapitre 3 en optimisation dans le problème nommé *Regression linéaire - Méthode des moindres carrés*.

**Remarque :** (pour ceux qui connaissent les espaces euclidiens) Ce théorème signifie que pour tout  $y \in \mathbb{R}^n$ ,  $\hat{a}(y)$  et  $\hat{b}(y)$  sont les deux réels tels que le projeté orthogonal de  $y$  sur  $\text{Vect}(x, 1)$  s'écrit  $\hat{b}(y)1 + \hat{a}(y)x$ , où  $1$  est le vecteur de  $\mathbb{R}^n$  dont toutes les coordonnées valent 1. D'autre part,  $\hat{\sigma}^2(y)$  est le carré de la norme euclidienne de  $y - (\hat{b}(y)1 + \hat{a}(y)x)$ .

**Théorème 4.6.3 (Qualité des estimateurs  $\hat{b}$ ,  $\hat{a}$  et  $\hat{\sigma}^2$  de  $a, b$  et  $\sigma^2$ )** Ils sont sans biais :  $\mathbb{E}(\hat{b}(Y)) = a$ ,  $\mathbb{E}(\hat{a}(Y)) = b$  et  $\mathbb{E}(\hat{\sigma}^2(Y)) = \sigma^2$ , et minimisent le risque quadratique.

**Remarque :** L'estimateur  $(\hat{b}, \hat{a})$  de  $(a, b)$  est aussi l'estimateur du maximum de vraisemblance de  $(a, b)$ .

*Preuve (partielle) :* On ne calcule que les espérances  $\mathbb{E}(\hat{a}(Y))$  et  $\mathbb{E}(\hat{b}(Y))$ .

$$\mathbb{E}(\hat{a}(Y)) = \frac{\sum_{k=1}^n (x_k - \bar{x})(\mathbb{E}(Y_k) - (1/n) \sum_{l=1}^n \mathbb{E}(Y_l))}{\sum_{k=1}^n (x_k - \bar{x})^2}$$

Or pour tout  $k$ ,  $\mathbb{E}(Y_k) = ax_k + b$ , donc

$$\begin{aligned}\mathbb{E}(\hat{a}(Y)) &= \frac{\sum_{k=1}^n (x_k - \bar{x})(ax_k - (1/n) \sum_{l=1}^n ax_l)}{\sum_{k=1}^n (x_k - \bar{x})^2} \\ &= \frac{\sum_{k=1}^n (x_k - \bar{x})(ax_k - a\bar{x})}{\sum_{k=1}^n (x_k - \bar{x})^2} \\ &= a\end{aligned}$$

$$\mathbb{E}(\widehat{b}(Y)) = (1/n) \sum_{l=1}^n \mathbb{E}(Y_l) - \bar{x} \mathbb{E}(\widehat{a}(Y))$$

Or pour tout  $k$ ,  $\mathbb{E}(Y_k) = ax_k + b$ , donc

$$\begin{aligned} \mathbb{E}(\widehat{b}(Y)) &= (1/n) \sum_{l=1}^n (ax_l + b) - \bar{x}a \\ &= a\bar{x} + b - \bar{x}a = b \end{aligned}$$

□

**Théorème 4.6.4** - la loi du couple  $(\widehat{b}(Y), \widehat{a}(Y))$  est celle d'un vecteur gaussien de  $\mathbb{R}^2$  de moyenne  $(a, b)$  et de matrice de covariance

$$\frac{\sigma^2}{\sum_{k=1}^n (x_k - \bar{x})^2} \begin{pmatrix} 1 & -\bar{x} \\ -\bar{x} & \frac{1}{n} \sum_{k=1}^n x_k^2 \end{pmatrix}$$

- la loi de  $\frac{(n-2)\widehat{\sigma}^2(Y)}{\sigma^2}$  est  $\chi^2(n-2)$

- le couple  $(\widehat{b}(Y), \widehat{a}(Y))$  et  $\widehat{\sigma}^2(Y)$  sont indépendants

*Preuve* : Ces résultats sont assez faciles à démontrer avec le théorème de Cochran quand on connaît la structure euclidienne de  $\mathbb{R}^n$ . Sans utiliser ces outils, nous n'allons démontrer que (1).

On pose  $\bar{Y} = \frac{1}{n} \sum_{k=1}^n Y_k$  et  $\bar{U} = \frac{1}{n} \sum_{k=1}^n U_k$ . On a alors  $\bar{Y} = a\bar{x} + b + \bar{U}$ .

- les  $U_i$  étant i.i.d. de loi gaussienne, toute combinaison linéaire des  $U_i$  et de constantes a une loi gaussienne, donc tout combinaison linéaire de  $\widehat{b}(Y)$  et  $\widehat{a}(Y)$  a une loi gaussienne. Ainsi, la loi du couple  $(\widehat{b}(Y), \widehat{a}(Y))$  est celle d'un vecteur gaussien de  $\mathbb{R}^2$  de moyenne  $(a, b)$ .

Calculons sa matrice de covariance.

Commençons par calculer  $\widehat{a}(Y) - a$  :

$$\begin{aligned} a - \widehat{a}(Y) &= \frac{\sum_{k=1}^n (x_k - \bar{x})^2 a - \sum_{k=1}^n (x_k - \bar{x}) (Y_k - \bar{Y})}{\sum_{l=1}^n (x_l - \bar{x})^2} \\ &= \frac{\sum_{k=1}^n (x_k - \bar{x}) (a(x_k - \bar{x}) - (Y_k - \bar{Y}))}{\sum_{l=1}^n (x_l - \bar{x})^2} \end{aligned}$$

Or  $Y_k - \bar{Y} = -a(x_k - \bar{x}) + U_k - \bar{U}$ , donc

$$\begin{aligned} a - \widehat{a}(Y) &= \frac{\sum_{k=1}^n (x_k - \bar{x}) (\bar{U} - U_k)}{\sum_{l=1}^n (x_l - \bar{x})^2} \\ &= \bar{U} \frac{\sum_{k=1}^n (x_k - \bar{x})}{\sum_{l=1}^n (x_l - \bar{x})^2} - \frac{\sum_{k=1}^n (x_k - \bar{x}) U_k}{\sum_{l=1}^n (x_l - \bar{x})^2}. \end{aligned}$$

Or

$$\sum_{k=1}^n (x_k - \bar{x}) = n\bar{x} - n\bar{x} = 0$$

donc

$$a - \widehat{a}(Y) = - \frac{\sum_{k=1}^n (x_k - \bar{x}) U_k}{\sum_{l=1}^n (x_l - \bar{x})^2},$$

Donc

$$\begin{aligned} \mathbb{E}((\widehat{a}(Y) - a)^2) &= \text{Var}(\widehat{a}(Y) - a) \\ &= \sigma^2 \frac{\sum_{l=1}^n (x_l - \bar{x})^2}{(\sum_{l=1}^n (x_l - \bar{x})^2)^2} \\ &= \frac{\sigma^2}{\sum_{l=1}^n (x_l - \bar{x})^2}. \end{aligned}$$

Calculons  $\widehat{b}(Y) - b$  :

$$\begin{aligned}\widehat{b}(Y) - b &= \bar{Y} - \bar{x}\widehat{a}(Y) - b \\ &= a\bar{x} + b + \bar{U} - \bar{x}\widehat{a}(Y) - b \\ &= \bar{x}(a - \widehat{a}(Y)) + \bar{U}\end{aligned}$$

donc par ce qui précède

$$\begin{aligned}\widehat{b}(Y) - b &= -\bar{x} \frac{\sum_{k=1}^n (x_k - \bar{x})U_k}{\sum_{l=1}^n (x_l - \bar{x})^2} + \bar{U} \\ &= \frac{1}{n} \sum_{k=1}^n \left( 1 - \frac{n\bar{x}(x_k - \bar{x})}{\sum_{l=1}^n (x_l - \bar{x})^2} \right) U_k.\end{aligned}$$

Donc

$$\begin{aligned}\mathbb{E} \left( (\widehat{b}(Y) - b)^2 \right) &= \text{Var} \left( \widehat{b}(Y) - b \right) \\ &= \frac{1}{n^2} \sum_{k=1}^n \left( 1 - \frac{n\bar{x}(x_k - \bar{x})}{\sum_{l=1}^n (x_l - \bar{x})^2} \right)^2 \text{Var}(U_k) \\ &= \frac{\sigma^2}{n^2} \sum_{k=1}^n \left( 1 - \frac{n\bar{x}(x_k - \bar{x})}{\sum_{l=1}^n (x_l - \bar{x})^2} \right)^2 \\ &= \frac{\sigma^2}{n^2} \left( n - 2n\bar{x} \underbrace{\frac{\sum_{k=1}^n (x_k - \bar{x})}{\sum_{l=1}^n (x_l - \bar{x})^2}}_{=0} + n^2 \bar{x}^2 \frac{\sum_{l=1}^n (x_l - \bar{x})^2}{(\sum_{k=1}^n (x_k - \bar{x})^2)^2} \right) \\ &= \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{l=1}^n (x_l - \bar{x})^2} \right) \\ &= \sigma^2 \frac{(1/n) \sum_{k=1}^n (x_k - \bar{x})^2 + (1/n \sum_{k=1}^n (x_k - \bar{x}))^2}{\sum_{l=1}^n (x_l - \bar{x})^2} \\ &= \sigma^2 \frac{(1/n) \sum_{k=1}^n x_k^2}{\sum_{l=1}^n (x_l - \bar{x})^2}.\end{aligned}$$

Calculons  $(\widehat{a}(Y) - a)(\widehat{b}(Y) - b)$  :

$$\widehat{b}(Y) - b = \bar{x}(a - \widehat{a}(Y)) + \bar{U}$$

donc

$$\begin{aligned}(\widehat{a}(Y) - a)(\widehat{b}(Y) - b) &= -\bar{x}(a - \widehat{a}(Y))^2 - \bar{U}(a - \widehat{a}(Y)) \\ &= -\bar{x}(a - \widehat{a}(Y))^2 + \bar{U} \frac{\sum_{k=1}^n (x_k - \bar{x})U_k}{\sum_{l=1}^n (x_l - \bar{x})^2} \\ \mathbb{E} \left( (\widehat{a}(Y) - a)(\widehat{b}(Y) - b) \right) &= -\bar{x} \frac{\sigma^2}{\sum_{l=1}^n (x_l - \bar{x})^2} + \sigma^2 \underbrace{\frac{\sum_{k=1}^n (x_k - \bar{x})}{n \sum_{l=1}^n (x_l - \bar{x})^2}}_{=0} \\ &= -\bar{x} \frac{\sigma^2}{\sum_{l=1}^n (x_l - \bar{x})^2}\end{aligned}$$

□

### 4.6.3 Test d'hypothèses

#### Tests sur $a$

- Question** Donnez une v.a. observable lorsque l'on connaît  $a$ , et dont la loi est connue. On sait que

$$\left( \frac{1}{\sum_{k=1}^n (x_k - \bar{x})^2} \right)^{-\frac{1}{2}} (\hat{a}(Y) - a)$$

suit une loi normale centrée réduite, et que cette v.a. est indépendante de  $\frac{(n-2)\hat{\sigma}^2(Y)}{\sigma^2}$ , qui suit une loi  $\chi^2(n-2)$ . Donc la v.a.

$$\frac{\left( \frac{\sigma^2}{\sum_{k=1}^n (x_k - \bar{x})^2} \right)^{-\frac{1}{2}} (\hat{a}(Y) - a)}{\frac{\hat{\sigma}(Y)}{\sigma}},$$

soit

$$\frac{(\sum_{k=1}^n (x_k - \bar{x})^2)^{\frac{1}{2}} (\hat{a}(Y) - a)}{\hat{\sigma}(Y)}$$

suit une loi du Student à  $n-2$  degrés de liberté.

- Déduisez-en un test de niveau  $\alpha$  de l'hypothèse  $a = a_0$**

Notons  $t_{1-\frac{\alpha}{2}}(n-2)$  le quantile de niveau  $1 - \frac{\alpha}{2}$  de la loi de Student à  $n-2$  degrés de liberté. Alors en acceptant l'hypothèse lorsque

$$\frac{(\sum_{k=1}^n (x_k - \bar{x})^2)^{\frac{1}{2}} |\hat{b}(Y) - a_0|}{\hat{\sigma}(Y)} \leq t_{1-\frac{\alpha}{2}}(n-2),$$

on a un test de niveau  $\alpha$ .

- Déduisez-en un test de niveau  $\alpha$  de l'hypothèse  $a \leq a_0$**

Notons  $t_{1-\alpha}(n-2)$  le quantile de niveau  $1 - \alpha$  de la loi de Student à  $n-2$  degrés de liberté. Alors en acceptant l'hypothèse lorsque

$$\frac{(\sum_{k=1}^n (x_k - \bar{x})^2)^{\frac{1}{2}} (\hat{b}(Y) - a_0)}{\hat{\sigma}(Y)} \leq t_{1-\alpha}(n-2),$$

on a un test de niveau  $\alpha$ .

#### Tests sur $b$

- Question** Donnez une v.a. observable lorsque l'on connaît  $b$ , et dont la loi est connue. On sait que

$$\left( \frac{\sum_{k=1}^n x_k^2 \sigma^2}{n \sum_{k=1}^n (x_k - \bar{x})^2} \right)^{-\frac{1}{2}} (\hat{a}(Y) - b)$$

suit une loi normale centrée réduite, et que cette v.a. est indépendante de  $\frac{(n-2)\hat{\sigma}^2(Y)}{\sigma^2}$ , qui suit une loi  $\chi^2(n-2)$ . Donc la v.a.

$$\frac{\left( \frac{\sum_{k=1}^n x_k^2 \sigma^2}{n \sum_{k=1}^n (x_k - \bar{x})^2} \right)^{-\frac{1}{2}} (\hat{a}(Y) - b)}{\frac{\hat{\sigma}(Y)}{\sigma}},$$

soit

$$\frac{(n \sum_{k=1}^n (x_k - \bar{x})^2)^{\frac{1}{2}} (\hat{a}(Y) - b)}{(\sum_{k=1}^n x_k^2)^{\frac{1}{2}} \hat{\sigma}(Y)}$$

suit une loi du Student à  $n-2$  degrés de liberté.

2. **Déduisez-en un test de niveau  $\alpha$  de l'hypothèse  $b = b_0$**

Notons  $t_{1-\frac{\alpha}{2}}(n-2)$  le quantile de niveau  $1 - \frac{\alpha}{2}$  de la loi de Student à  $n-2$  degrés de liberté. Alors en acceptant l'hypothèse lorsque

$$\frac{(n \sum_{k=1}^n (x_k - \bar{x})^2)^{\frac{1}{2}} |\hat{a}(Y) - b_0|}{(\sum_{k=1}^n x_k^2)^{\frac{1}{2}} \hat{\sigma}(Y)} \leq t_{1-\frac{\alpha}{2}}(n-2),$$

on a un test de niveau  $\alpha$ .

3. **Déduisez-en un test de niveau  $\alpha$  de l'hypothèse  $b \leq b_0$**

Notons  $t_{1-\alpha}(n-2)$  le quantile de niveau  $1 - \alpha$  de la loi de Student à  $n-2$  degrés de liberté. Alors en acceptant l'hypothèse lorsque

$$\frac{(n \sum_{k=1}^n (x_k - \bar{x})^2)^{\frac{1}{2}} (\hat{a}(Y) - b_0)}{(\sum_{k=1}^n x_k^2)^{\frac{1}{2}} \hat{\sigma}(Y)} \leq t_{1-\alpha}(n-2),$$

on a un test de niveau  $\alpha$ .

**Tests sur  $\sigma^2$**

1. **Question** Donnez une v.a. observable lorsque l'on connaît  $\sigma$ , et dont la loi est connue.

$\frac{(n-2)\hat{\sigma}^2(Y)}{\sigma^2}$  suit une loi du chi-deux à  $n-2$  degrés de liberté.

2. **Déduisez-en un test de niveau  $\alpha$  de l'hypothèse  $\sigma^2 \leq \sigma_0^2$**

Notons  $t_{1-\alpha}(n-2)$  le quantile de niveau  $1 - \alpha$  de la loi du chi-deux à  $n-2$  degrés de liberté. Alors en acceptant l'hypothèse lorsque

$$\frac{(n-2)\hat{\sigma}^2(Y)}{\sigma^2} \leq t_{1-\alpha}(n-2),$$

on a un test de niveau  $\alpha$ .