

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

**UNIVERSITÉ MOHAMED KHIDER, BISKRA**

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

**DÉPARTEMENT DE MATHÉMATIQUES**



Mémoire présenté en vue de l'obtention du Diplôme :

**MASTER en Mathématiques**

Option : **Statistique**

Par

**KRAZTI Nassira**

Titre :

# Comparaison entre les méthodes d'estimation

Membres du Comité d'Examen :

Dr. DJABER Ibtissem	UMKB	Président
Pr. BRAHIMI Brahim	UMKB	Encadreur
Dr. BERKANE Hassiba	UMKB	Examineur

Juin 2018

## DÉDICACE

*Je dédise ce travail :*

*À mes parents...*

*À mes soeurs et frères...*

*À mes jolis neveux...*

*À toute ma famille...*

*À tous mes amis...*

*À tous ceux que j'aime et ceux qui m'aiment.*

مباركة

## REMERCIEMENTS

Je tiens tout d'abord à remercier **Dieu** de m'avoir donné le courage, le moral et la santé pour mener à bien ce travail.

Je remercie chaleureusement et présente mon plus profond respect à mon encadreur **Pr.BRAHIMI Brahim** pour ses conseils précieux qui m'ont été énormément utiles ainsi que pour le soutien constant et bienveillant qu'elle m'a apporté durant toutes les étapes de réalisation de ce mémoire.

Je tiens à exprimer mes plus vifs remerciements à Dr.DJABER Ibtissem de bien vouloir présider le jury de soutenance ainsi un grand merci à Dr.BERKANE Hassiba d'avoir accepté la pénible tâche de lire ce manuscrit et surtout pour tout son soutien et encouragement.

J'adresse aussi un remerciement tout particulier avec tous mes sentiments de respectueuse gratitude mes enseignants et particulièrement l'équipe du département de mathématiques qui ont été à la hauteur et je leur en suis reconnaissante.

J'exprime aussi mes remerciements plus généralement à tous mes enseignants de toutes les années d'études de m'avoir donné ce qui a peut-être été le meilleur conseil de toute ma carrière d'étude, ainsi que pour son disponibilité surtout dans les moments délicats...!

Je remercie également toutes les personnes qui m'ont aidé, et qui ont contribué de proche ou de loin à la réalisation de ce travail, sans oublier mes collègues de l'université de parcours qui sont la source d'une bonne ambiance, humeur et complicité qui ont régné pendant ces années.

Merci enfin à mes parents, mes sœurs, mes frères pour son soutiennent qui n'a jamais discontinuité.

Encore une fois merci à tous : **Nassira Krazti.**

# Table des matières

Remerciements	ii
Table des matières	iii
Liste des figures	v
Liste des tableaux	vi
Résumé	vii
Abstract	ix
Introduction	1
<b>1 Estimation</b>	<b>3</b>
1.1 Estimation . . . . .	3
1.2 Comparaison des estimateurs . . . . .	5
1.2.1 Fonction de risque, estimateur admissible et approche minimax . . .	5
1.2.2 L'approche de Rao-Blackwell-Lehmann-Scheffé . . . . .	7
1.3 Méthodes d'estimation paramétrique . . . . .	8
1.3.1 Estimation ponctuel . . . . .	8
1.3.2 Estimation par intervalle de confiance . . . . .	14
1.4 Borne de Cramér-Rao . . . . .	15

1.5	Autres estimateurs . . . . .	16
1.5.1	Estimateurs bayésiens . . . . .	16
1.5.2	Le maximum a posteriori (MAP) . . . . .	17
<b>2</b>	<b>Méthodes d'estimation</b>	<b>18</b>
2.1	Exemples sur l'estimation ponctuel . . . . .	18
2.1.1	Méthode du maximum de vraisemblance . . . . .	18
2.1.2	Méthode des moments . . . . .	21
2.1.3	Méthode de régression . . . . .	22
2.1.4	Méthode des moindres carrés . . . . .	24
2.1.5	Estimateur bayésien . . . . .	25
2.2	Exemple sur l'estimation par intervalle de confiance . . . . .	26
	<b>Conclusion</b>	<b>35</b>
	<b>Bibliographie</b>	<b>37</b>
	<b>Annexe A : Définitions</b>	<b>39</b>
	<b>Annexe B : Abréviations et Notations</b>	<b>42</b>

# Table des figures

1.1	Le 1 <sup>er</sup> cas . . . . .	11
1.2	Le 2 <sup>ème</sup> cas . . . . .	11
1.3	Le 3 <sup>ème</sup> cas . . . . .	12
1.4	Le 4 <sup>ème</sup> cas . . . . .	12
1.5	La droite des moindres carrés . . . . .	13

# Liste des tableaux

2.1	La durée de vie en heures du produit. . . . .	32
-----	---	----

# Résumé



R.A Fisher (17/02/1890 – 29/07/1962)

Les méthodes statistiques sont aujourd'hui utilisées dans presque tous les secteurs de l'activité humaine et font partie des connaissances de base de l'ingénieur, du gestionnaire, de l'économiste, du biologiste, de l'informaticien... Parmi les innombrables applications citons dans le domaine industriel : la fiabilité des matériels, le contrôle de qualité, l'analyse des résultats de mesures et leur planification, la prévision, et dans le domaine de l'économie et des sciences de l'homme : les études quantitatives de marché, etc.

Nous allons tenter de préciser dans les paragraphes suivants les notions fondamentales de la statistique et les rapports qu'elle entretient avec la théorie des probabilités ainsi que ce qu'on entend par démarche statistique.

La notion de distribution d'échantillonnage est à la base des méthodes d'inférence statistique dont les deux principales applications sont les problèmes d'estimation et les tests d'hypothèse. Les premiers ont pour but d'estimer, à partir d'un échantillon, la valeur numérique d'un ou de plusieurs paramètres de la population, et de déterminer la précision de cette ou de ces estimations. Les seconds ont pour but de vérifier la véracité d'une hypothèse émise au départ au sujet d'une ou de plusieurs populations.

En supposant que l'échantillon est aléatoire et simple  $(X_1, \dots, X_n)$ , la variable aléatoire  $T$  possède une distribution de probabilité, dite distribution d'échantillonnage, est donc la distribution des différentes valeurs que peut prendre la variable aléatoire  $T$ , pour les différents échantillons possibles. Son écart-type  $\sigma_T$  est appelé erreur standard.

Les principales distributions d'échantillonnage sont la distribution d'échantillonnage de la moyenne, la distribution d'échantillonnage de la variance et la distribution d'échantillonnage de la proportion.

On se place dans un modèle statistique paramétrique on cherche à estimer au mieux le paramètre  $\theta$  à partir de l'observation  $x$  à l'aide d'une statistique  $T(x)$ . Les résultats seront d'abord introduits dans le cas simple où  $\theta$  est de dimension 1, mais nous regarderons rarement le cas où  $\theta$  est de dimension  $d$  quelconque en abordant la notion d'information de Fisher [3].

Les résultats que nous allons présenter ne doivent pas être pris excessivement au premier degré dans les applications pratiques, car le modèle est généralement approximatif, d'autant plus que l'échantillon est grand; d'un point de vue pratique, la vraisemblance doit être vue comme un contraste.

## Mots clés :

Estimation paramétrique, estimateur, fonction de risque, borne de Cramér-Rao, maximum de vraisemblance, la méthode de régression, intervalle de confiance, estimation de la variance.

# Abstract

The statistical methods are today used in almost all the sectors of the human activity and belong to basic knowledge of the engineer, the manager, the economist, the biologist, the data processing specialist... Among the innumerable applications let us quote in the industrial field : the reliability of the materials, quality control, analysis of the results of measurement and their planning, the forecast, and in the field of the economy and the social sciences : studies quantitative of market, etc.

We will try to specify in the following paragraphs the basic concepts of the statistics and the reports which it maintains with the theory of probability as what one understands by statistical step.

The concept of distribution of sampling is at the base of the methods of statistical inference whose two principal applications are the problems of the estimate and the tests of assumptions. The purpose of the first are to estimate, starting from a sample, the numerical value of one or several parameters of the population, and to determine the precision of this or these estimates. The purpose of the seconds is to check the veracity of an assumption put forth at the beginning about one or several populations.

By supposing that the sample is random and simple  $(X_1, \dots, X_n)$ , the random variable  $T$  has a probability distribution, known as distribution of sampling, is thus the distribution of the various values which the random variable can take  $T$ , for the various possible samples. Its standard deviation  $\sigma_T$  is called standard error.

The principal distributions of sampling are the distribution of sampling of the average,

the distribution of sampling of the variance and the distribution of sampling of the proportion.

One place oneself in a parametric statistical model one seeks to estimate as well as possible the parameter  $\theta$  starting from observation  $x$  using statistics  $T(x)$ . The results will be initially introduced into the simple case where  $\theta$  is of dimension 1, but we will seldom look at the case where  $\theta$  is of unspecified size  $d$  by approaching the concept of information of Fisher [3].

The results which we will present should not be taken excessively with the first degree in the practical applications, because the model is generally approximate, more especially as the sample is large ; from a practical point of view, the probability must be seen like a contrast.

### Key words :

Parametric estimate, estimator, function of risk, limit of Cramér-Rao, maximum of vraisemblance, the method of regression, confidence interval, estimate of the variance.

# Introduction

Le problème d'estimation se réduit à estimer un paramètre  $\theta$  qu'est une valeur numérique qui n'est pas calculée par le modèle et qui n'est pas une variable d'entrée mesurée ou observée à partir d'une observation de  $X$ , ayant à sa disposition une famille des lois (pour  $X$ ) indexées par  $\theta$ ; cette réduction diminue souvent considérablement le vecteur d'observation, car  $X$  aura typiquement la dimension de  $\theta$ , mais son utilisation peut compliquer la famille de lois.

Ce domaine aussi est riche en méthodes et modèles d'estimation qui ont été étudié et amélioré afin d'obtenir des estimateurs d'effort de plus en plus réalistes. Les études ont montré que, pour produire des bons estimateurs, selon le modèle d'estimation sélectionné pour un paramètre et la pertinence des estimations diffèrent. Face à la multitude des modèles d'estimation existants, tels que l'information asymptotique existe, la question qui se pose par Fisher est alors souvent de savoir quelques estimateurs qui sont sans biais optimaux, comme sa raison de nous représenter la quantité d'informations de Fisher. D'autre part, pour un modèle statistique dominé  $P_{\Theta}$ , on trouve la notion de la vraisemblance; cette idée date de Gauss (vers 1821) dans ses études de lois qui portent son nom. Cependant, la fondation d'une entière méthodologie développant, cette idée revient au statisticien anglais R.A. Fisher (dans l'année 1920).

En ce travail, nous intéresserons d'abord à étudier les méthodes classiques d'estimateurs, on verra différentes notions de comparaison d'estimateurs et la recherche d'un estimateur optimal associé à une expérience statistique. En étudiant en particulier les méthodes pa-

ramétriques ponctuelles (maximum de vraisemblance, des moments,..) et par intervalle de confiance, de plus nous nous intéresserons à différents exemples théoriques, avec en particulier des observations qui ne seraient plus forcément iid.

# Chapitre 1

## Estimation

### 1.1 Estimation

Soit  $X$  une variable aléatoire de loi  $P_\theta$ , où  $\theta$  désigne un paramètre réel (à valeurs dans  $\Theta \subseteq \mathbb{R}$ , pour éviter que les sup ne soient infinis, il faut généralement supposer  $\Theta$  compact) inconnu à estimer, dans un problème d'estimation, on s'intéresse à identifier  $g(\theta)$ , une transformation du paramètre  $\theta$ , avec :

$$g : \Theta \rightarrow \mathbb{R}^d, \text{ mesurable ;}$$

très souvent  $d = 1$  et on notera  $S = g(\Theta)$ .

Un estimateur de  $g(\theta)$  est une v.a. de la forme  $T(X_1, \dots, X_n)$  où :  $T$  est une fonction mesurable de  $X_n$  dans  $\mathbb{R}^d$  (ou  $\mathbb{Z}^d$ ), une valeur particulière  $T(x_1, \dots, x_n)$  d'un estimateur sur un échantillon numérique  $x_1, \dots, x_n$  s'appelle une estimation de  $g(\theta)$  (il est important de réaliser qu'une estimation de  $\theta$  est une grandeur numérique alors qu'un estimateur de  $\theta$  est une variable aléatoire).

L'estimateur  $T$  doit vérifier certaines propriétés pour être de bonne qualité, un estimateur  $T$  est **sans biais** si :

$$E_\theta [T] = g(\theta), \forall \theta \in \Theta,$$

plus généralement, on appelle le biais d'un estimateur  $T$  la quantité  $b_{\theta,n}(T)$ , dans le cas où :

$$\lim_{n \rightarrow \infty} b_{\theta,n}(T) = 0, \forall \theta \in \Theta,$$

on trouve que  $T$  est **asymptotiquement sans biais**.

On dit qu'un estimateur  $T$  est **exhaustif** (suffisant) si pour tout ensemble  $\mathcal{A}$ , il existe une version de l'espérance conditionnelle vérifie :

$$\mathbb{E}_{\theta}[\mathbb{1}_{X \in \mathcal{A}} | T] \equiv \theta.$$

L'exemple le plus simple de statistique exhaustive est la moyenne empirique des  $T(X_i)$  où les  $X_i$  sont i.i.d issues d'une famille exponentielle.

D'autre part, une statistique  $T$  est **complète** si pour toute fonction  $\psi$  telle que :

$$\mathbb{E}_{\theta}(|\psi(T)|) < +\infty, \forall \theta \in \Theta,$$

on a :

$$\mathbb{E}_{\theta}(\psi(T)) = 0, \forall \theta \in \Theta \implies \psi(T) = 0 \text{ "P}_{\theta}p.s", \forall \theta.$$

**Remarque 1.1.1** Dans le cas où  $T$  est complète, avec  $\psi_1(T)$  et  $\psi_2(T)$  sont deux estimateurs sans biais de  $g(\theta)$ , alors :

$$\psi_1(T) = \psi_2(T).$$

De plus, on dit qu'un estimateur  $T$  sans biais de  $g(\theta)$  est **uniformément de variance minimale** parmi les estimateurs sans biais si, pour tout estimateur sans biais  $U$  de  $g(\theta)$ ,

on a :

$$\forall \theta \in \Theta, R(T, g(\theta)) \leq R(U, g(\theta)),$$

où  $R$  désigne le risque quadratique.

La variable aléatoire  $T_n$  est un estimateur **convergent** si :

$$E(T_n) \xrightarrow{p} g(\theta), \text{ lorsque } n \text{ tend vers } +\infty.$$

La précision d'un estimateur est liée à sa variance, mais il ne suffit pas d'imposer une variance minimale pour obtenir le meilleur estimateur, car tous les estimateurs **certain** (i.e : de valeur constante) sont de variance nulle ! Il faut donc ajouter une condition sur l'espérance mathématique.

La variable aléatoire  $T_n$  est un estimateur **efficace** s'il est sans biais et s'il est de variance minimale parmi les estimateurs sans biais, et on dira que  $T$  est **asymptotiquement efficace** si  $\forall \theta \in \Theta$  :

$$\mathbb{E}_\theta [(T - \theta) (T - \theta)^t] \rightarrow I(\theta)^{-1}.$$

## 1.2 Comparaison des estimateurs

### 1.2.1 Fonction de risque, estimateur admissible et approche minimax

La comparaison des différents estimateurs s'effectue à l'aide des fonctions de risque, premièrement, on définit la fonction de perte :

$$L : \Theta \times S \rightarrow [0, +\infty[ ,$$

vérifiant :

$$L(\theta, g(\theta)) = 0 \text{ et mesurable,}$$

typiquement :

$$L(\theta, g(\theta)) = |\theta - g(\theta)|^2 ,$$

le risque (perte moyenne) d'un estimateur  $T$  de  $g(\theta)$  est :

$$\theta \mapsto R_\theta(T) = E_\theta [L(\theta, T)] = \int \dots \int L(\theta, T(x_1, \dots, x_n)) dP_\theta(x_1) \dots dP_\theta(x_n).$$

C'est cette fonction qui caractérise la qualité de l'estimateur, la comparaison d'estimateurs est donc une comparaison de fonctions, c'est pourquoi il est difficile de mettre en avant un estimateur optimal, même théoriquement.

On dit qu'un estimateur  $T$  est **préférable** à un autre estimateur  $U$ , noté  $T \geq U$ , si :

$$R_\theta(T) \leq R_\theta(U);$$

il est meilleur, noté  $T > U$ , si de plus,  $R_\theta(T) < R_\theta(U)$  pour au moins un  $\theta \in \Theta$ . La relation  $\geq$  est réflexive et transitive : c'est un préordre (ordre partiel).

Un estimateur  $T$  est **admissible** s'il n'existe pas d'autre estimateur  $U$  strictement meilleur au sens où :

$$R_\theta(U) \leq R_\theta(T) \forall \theta, \text{ et } : R_{\theta_0}(U) < R_{\theta_0}(T) \text{ pour certain } \theta_0.$$

Dans le cas contraire, un estimateur  $T$  est **inadmissible** s'il existe un estimateur  $U$  qui lui est meilleur. Cette notion d'admissibilité permet de s'intéresser uniquement à la classe des estimateurs admissibles : Il est admissible dans le cas contraire, on verra que les estimateurs bayésiens sont généralement admissibles, ce qui fournit une grande classe d'exemples.

Par ailleurs, un estimateur  $T$  est dit **minimax** sur  $\Theta$  si et seulement si pour tout autre estimateur  $U$  on a :

$$\sup_{\theta \in \Theta} R_\theta(T) \leq \sup_{\theta \in \Theta} R_\theta(U),$$

**Remarque 1.2.1** *il est difficile de trouver des estimateurs minimax et si possible admissibles. De plus, il y a rarement unicité, ce qui fait que ces deux critères ne permettent pas*

de caractériser un estimateur privilégié.

Un estimateur sera dit **uniformément de risque minimal** (uniformly minimum risk, *UMR*) si sa fonction de risque est partout inférieure aux autres fonctions de risque :

$$\forall \theta \in \Theta, \forall U : R_\theta(T) \leq R_\theta(U),$$

où  $U$  varie parmi tous les estimateurs.

**Remarque 1.2.2** *C'est une propriété extrêmement forte ; un estimateur UMR est bien entendu optimal. Mais en général un tel estimateur n'existe pas car le membre de droite a pour chacun  $\theta$  un minimum nul : il suffit de prendre  $U = \theta$ .*

**Remarque 1.2.3** *Si l'objectif est de minimiser le risque quadratique, on a parfois intérêt à considérer des estimateurs qui sont biaisés.*

## 1.2.2 L'approche de Rao-Blackwell-Lehmann-Scheffé

Cette théorie a pour but de mettre en évidence des estimateurs *UMR* dans un cadre non asymptotique. On a besoin d'introduire quelques concepts qui décrivent comment une statistique concentre l'information nécessaire à l'estimation de  $\theta$ .

Si  $T$  est un estimateur, alors :

$$T_* = \mathbb{E}_\theta[T|X]$$

est un meilleur estimateur de  $\theta$  au sens de l'erreur quadratique. C'est le théorème de Rao-Blackwell (Rao 1945, Blackwell 1947).

## 1.3 Méthodes d'estimation paramétrique

### 1.3.1 Estimation ponctuel

#### Méthode du maximum de vraisemblance

Considérons un modèle statistique dominé  $P_\Theta = \{P_\theta\}$  dont on note  $\{f(\cdot, \theta)\}$  la famille des fonctions de densité. Étant donné l'observation  $X$ , la probabilité de son apparition suivante  $P_\theta$  peut être mesurée par  $f(X, \theta)$ . Puisque  $\theta$  est inconnue, il semble alors naturel de favoriser les valeurs de  $\theta$  pour lesquelles  $f(X, \theta)$  est élevé : c'est la notion de la vraisemblance de  $\theta$  pour l'observation  $X$ .

On définit la fonction de vraisemblance comme suit :  $\theta \mapsto f(X, \theta)$ , et tout point maximal de cette fonction représente l'estimateur du maximum de vraisemblance :

$$\hat{\theta}_{mv} = \underset{\theta \in \Theta}{\operatorname{argmax}} f(X, \theta).$$

Remarquons qu'en cas de  $n$  observations  $X = (X_1, \dots, X_n)$ , notant  $\{f(\cdot, \theta)\}$  les densités de  $X_1$ , la fonction de vraisemblance devient :

$$\theta \mapsto L(X, \theta) = \prod_{i=1}^n f(X_i, \theta).$$

Pour obtenir l'estimateur  $\hat{\theta}$  du maximum de vraisemblance, on maximise sa log-vraisemblance selon  $\theta$  en résolvant le système d'équations du maximum de vraisemblance :

$$\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \dots, \theta_k; x) = 0; \text{ pour } i = \overline{1, k}.$$

Et si on souhaite estimer  $g(\theta)$  par la méthode du *M.V.*, on peut appliquer la méthode de substitution :

$$\hat{g}(\theta) := g(\hat{\theta}_{mv}).$$

**Remarque 1.3.1** *Les estimateurs du maximum de vraisemblance de  $\theta$  sont équivalents aux estimateurs des moindres carrés de  $\theta$ . On pourra le montrer dans le cas de la régression linéaire.*

*En revanche, certaines propriétés ne sont possibles que sous l'hypothèse de normalité des résidus.*

### Méthode des moments

Pour chacun  $\theta$ , soit le vecteur des  $s$  premiers moments :

$$\Psi(\theta) = [\mathbb{E}_\theta(x_1), \dots, \mathbb{E}_\theta(x_1^s)].$$

Étant donné les observations  $x = (x_1, \dots, x_n)$ , on définit les  $s$  premiers moments empiriques :

$$\hat{\alpha}_k := \frac{1}{n} \sum_{i=1}^n X_i^s, 1 \leq k \leq s.$$

On appelle estimateur des moments de  $\theta$ , tout estimateur  $\hat{\theta}$ , solution des équations des moments :

$$\mathbb{E}_\theta [X_1^k] = \hat{\alpha}_k, k = 1, \dots, s.$$

Ces équations s'écrivent aussi :  $\Psi(\theta) = \hat{\alpha}$  avec le vecteur  $\hat{\alpha} = (\hat{\alpha}_k)_{1 \leq k \leq s}$ .

Par la loi des grands nombres et sous les conditions d'existence des moments, pour  $n$  grand,  $\hat{\alpha} = \Psi(\hat{\theta})$  est proche de  $\Psi(\theta_0)$  où  $\mathbb{P}_{\theta_0}$  dénote la vraie loi des observations. Ainsi, sous de bonnes propriétés d'inversion de  $\Psi$ ,  $\hat{\theta}$  sera proche de  $\theta_0$ .

**Remarque 1.3.2** *La méthode des moments est très facile à implémenter et elle est très efficace en temps de calcul, mais très imprécise si l'échantillon n'est pas normalisé.*

**Méthode de régression**

Nous supposons donc que la fonction de régression est de la forme :

$$\mathbb{E}(Y|X = x) = ax + b,$$

pour  $x \in I$ , de plus nous supposons que la variance de la loi conditionnelle de  $Y$  sachant  $X = x$  ne dépend pas de  $x$  et est égale à  $\sigma^2$ . Enfin nous faisons l'hypothèse que cette loi est gaussienne,  $\forall x$ . Nous reviendrons toutefois par la suite sur cette condition qui n'est pas cruciale. Notons que la linéarité du modèle est relative aux paramètres  $a$  et  $b$ , et que l'on peut substituer à  $X$  des transformées  $\ln X$ ,  $\sqrt{X}$ ,  $X^2$  etc., pour atteindre éventuellement la linéarité du modèle.

Le modèle contient donc trois paramètres inconnus :  $a, b$  et  $\sigma^2$ . Pour estimer ces paramètres nous considérons  $(Y_1, \dots, Y_n)$  situées, respectivement, aux niveaux  $x_1, x_2, \dots, x_n$  de la variable explicative fixé (on suppose naturellement qu'au moins deux de ces valeurs sont distinctes). Ainsi :

$$Y_i \rightsquigarrow \mathcal{N}(ax_i + b, \sigma^2), \forall i.$$

Il est aussi commode d'utiliser la notation :

$$Y_i = ax_i + b + \varepsilon_i, \forall i = \overline{1, n}.$$

On a donc :

$$\varepsilon_i \rightsquigarrow \mathcal{N}(0, \sigma^2), \forall i.$$

L'indépendance des  $Y_i$  entraîne celle des «erreurs»  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  qui sont donc des variables aléatoires i.i.d.

**Remarque 1.3.3** *La méthode de régression est facile à implémenter et elle est très efficace en temps calcul. Les propriétés asymptotiques (convergence et normalité des estimateurs de*

moindres carrés dans une régression linéaire sont bien connues. Pour résoudre ce problème on peut proposer la variante de la méthode dont les résultats sont bien meilleurs pour une plus grande région de l'espace paramétrique. On remarque ces quatre exemples de regression peu convaincantes (Ces jeux de données ont été choisis de manière à définir la même droite de régression, et avec le même coefficient de corrélation  $R^2$ ) :

Le 1<sup>er</sup> cas :

Ce jeu de données est au mieux, très bruité mais on peut douter que les données soient liées par une relation affine, comme :

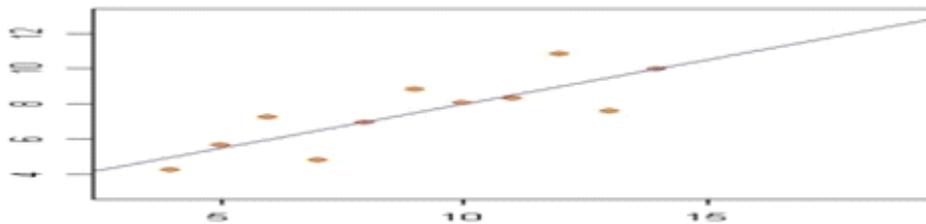


FIG. 1.1 – Le 1<sup>er</sup> cas

Le 2<sup>ème</sup> cas :

Ce jeu correspond assez clairement à une relation quadratique (c'est une courbe  $y = ax^2 + bx + c$ ) qu'il conviendrait d'ajuster :

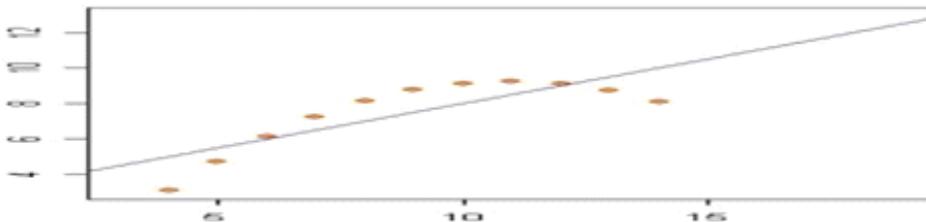


FIG. 1.2 – Le 2<sup>ème</sup> cas

Le 3<sup>ème</sup> cas :

*Ce jeu tous les points sauf un semblent alignés, il y a visiblement un point aberrant dont il faudrait vérifier la provenance (ou la saisie dans le logiciel).*

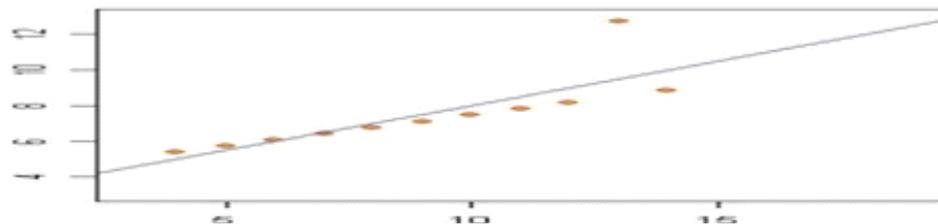


FIG. 1.3 – Le 3<sup>ème</sup> cas

*Le 4<sup>ème</sup> cas :*

*Cette situation est semblable pour le précédent échantillon. Moralité : la régression linéaire donne (presque) toujours une droite, mais il convient de regarder le résultat pour débusquer les situations par trop absurdes.*

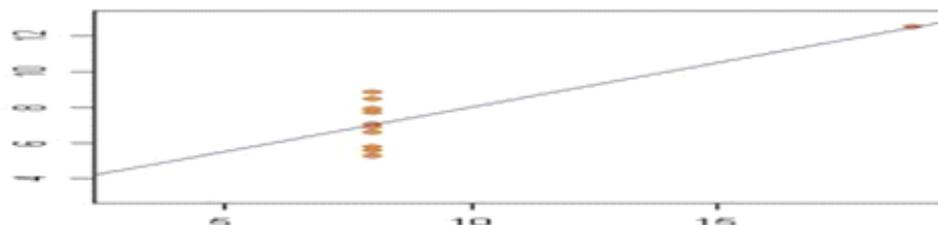


FIG. 1.4 – Le 4<sup>ème</sup> cas

### Méthode des moindres carrés

La méthode des moindres carrés consiste à estimer  $g(\theta)$  en minimisant la somme des carrés des résidus ( $ssr$ ), telle que :

$$\varphi(T) = \min \sum_{i=1}^n (\hat{\epsilon}_i)^2 = \min \sum_{i=1}^n (x_i - \hat{x}_i)^2,$$

Le critère des moindres carrés peut s'écrire aussi de la façon suivante :

$$\|\hat{e}\|^2 = \|g(\theta) - T\|^2.$$

Cette méthode d'estimation ne nécessite pas que l'on pose l'hypothèse de normalité des résidus, comme suit :

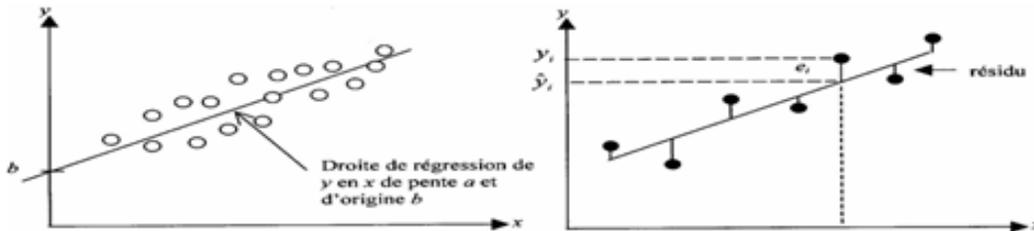


FIG. 1.5 – La droite des moindres carrés

### Méthode d'ajustement

La droite d'équation  $y = a + bx$  que l'on détermine pour représenter «au mieux» un nuage  $\{(x_i, y_i)\}$  de points observés. Dans un contexte du type déterministe avec erreurs de mesure, un type standard de méthode consiste à définir de façon générale une «distance» entre une droite et un nuage de points, et à déterminer la droite d'ajustement d'un nuage donné comme la droite qui minimise la distance à ce nuage. On prend généralement comme distance la somme des carrés des écarts :

$$\sum_{i=1}^n (y_i - (a + bx_i))^2,$$

ce qui donne la droite des moindres carrés.

Dans un contexte de type intrinsèquement aléatoire, où les  $(x_i, y_i)$  sont des observations d'un couple  $(X, Y)$  de v.a, on suppose que l'espérance conditionnelle  $\mathbb{E}(Y|X = x)$  est une

fonction affine de  $x$  :

$$\mathbb{E}(Y|X = x) = a + bx,$$

et on ajuste la droite en minimisant la variance des écarts, ce qui donne la droite de régression. Les deux méthodes ne diffèrent que par la présentation et fournissent la même droite.

**Remarque 1.3.4** *En pratique il est rare que l'on puisse ainsi calculer explicitement l'estimation d'un paramètre par ajustement. On doit alors procéder à une minimisation numérique sur le paramètre inconnu.*

### 1.3.2 Estimation par intervalle de confiance

Remplacement, dans une estimation, de la donnée d'une valeur ponctuelle par celle d'un intervalle entourant cette valeur, dont les extrémités dépendent d'un seuil de confiance fixé a priori. Cet intervalle s'appelle intervalle de confiance du paramètre.

L'estimation par intervalle de confiance consiste à déterminer autour de la valeur estimée un intervalle dont on a de fortes chances de croire qu'il contient la vraie valeur du paramètre recherché.

La détermination théorique d'un intervalle de confiance pour un paramètre  $\theta$  d'une loi de probabilité nécessite d'utiliser un estimateur  $T = T_n(X_1, \dots, X_n)$  de ce paramètre, et consiste à déterminer de part et d'autres de  $T$  les bornes  $T_1$  et  $T_2$  d'un intervalle qui a une forte probabilité de contenir  $\theta$ . Cette probabilité est appelée niveau (ou seuil) de confiance et désignée par  $(1 - \alpha)$ ,  $\alpha$  est alors un risque d'erreur.

Les limites  $T_1$  et  $T_2$  sont telles que :

$$\mathbb{P}(T_1 < T < T_2) = 1 - \alpha.$$

L'intervalle  $[T_1, T_2]$  est appelé intervalle de confiance, la probabilité que le paramètre se

trouve à l'extérieur de cet intervalle est donc :

$$\mathbb{P}(T < T_1) + \mathbb{P}(T > T_2) = \alpha.$$

Le risque total  $\alpha$  peut-être réparti d'une infinité de manières. Généralement, on divise ce risque en deux parties égales, c'est-à-dire on cherche un intervalle à risque symétrique, on peut déterminer, l'intervalle d'acceptation  $I_a(\theta) = [T_1, T_2]$  caractérisé par :

$$\mathbb{P}(\theta < T_1) = \mathbb{P}(\theta > T_2) = \frac{\alpha}{2}.$$

On obtient ainsi un encadrement probabiliste de l'estimation lorsque le paramètre  $\theta$  est fixé.

La détermination d'intervalle de confiance  $I_c$  nécessite d'inverser le procédé, afin d'obtenir un encadrement probabiliste du paramètre lorsque l'estimation  $T = T_n(x_1, x_2, \dots, x_n)$  est fixé.

Dans la présentation de résultats au public, on parle souvent de fourchette(malheureusement trop souvent sans préciser ni la taille de l'échantillon ni le seuil de confiance).

## 1.4 Borne de Cramér-Rao

Le but de ce paragraphe est de définir ce qu'est le meilleur estimateur, il est intéressant de commencer par la borne de Cramér-Rao :

**Théorème 1.4.1** *On suppose l'expérience  $R$ -régulière, alors pour tout estimateur  $T$  non biaisé d'une fonction  $g(\theta) \in C^1$ , tel que :*

$$\mathbb{E}_\theta [ |T|^2 ] < c < \infty, \theta \in \Theta,$$

la variance d'estimation est minorée comme suit :

$$\mathbb{E}_\theta [(T - g(\theta)) (T - g(\theta))^t] \geq \nabla g(\theta) I(\theta)^{-1} \nabla g(\theta)^t$$

où chaque ligne de  $\nabla g$  contient le gradient de la coordonnée correspondante.

**Remarque 1.4.1** Cette borne implique que dans le cas de  $n$  échantillons indépendants, on a pour tout estimateur  $T$  de  $\theta$  :

$$\mathbb{E}_\theta [(T - \theta) (T - \theta)^t] \geq I(\theta)^{-1},$$

où l'information de Fischer est calculée sur la base d'un seul échantillon.

## 1.5 Autres estimateurs

### 1.5.1 Estimateurs bayésiens

La méthode bayésienne consiste à minimiser en  $T$  le risque calculé sur la base de la probabilité totale :

$$R(T) = \int R_\theta(T) \pi d(\theta) = \iint r(\theta, T(x)) p_\theta(x) \pi d(\theta) \omega(dx),$$

puis à définir les estimateurs bayésiens par la propriété :

$T$  est bayésien si pour tout estimateur  $U$  on a :

$$R(T) \leq R(U),$$

Ceci donne :

$$T(x) = \arg \min \int r(\theta, g(\theta)) p_\theta(x) \pi d(\theta) = \arg \min \mathbb{E} [r(\theta, g(\theta)) | X] = x \omega - p.p.$$

### 1.5.2 Le maximum a posteriori (MAP)

L'estimation de  $\theta$  se fait ici par maximisation de  $p(\theta, X)$  c'est-à-dire qu'on choisit le mode de la probabilité a posteriori :

$$p(\theta|X) = p(\theta, X)/p(X) ;$$

le facteur  $1/p(X)$  étant constant, il vient :

$$T = \arg \max \pi(\theta)p_{\theta}(X)$$

Cet estimateur correspond également à l'estimateur bayésien limite ( $\varepsilon \rightarrow \infty$ ) quand :

$$r(\theta, g(\theta)) = \varphi(\|\theta - g(\theta)\|)$$

où  $\varphi$  est continue, nulle sur  $[0, \varepsilon]$ , 1 sur  $[2\varepsilon, +\infty[$ , et monotone  $[\varepsilon, 2\varepsilon]$ .

**Remarque 1.5.1** *Le MAP est utilisé quand la méthode bayésienne est difficile à mettre en œuvre en raison des calculs d'intégrales compliqués qu'elle impose.*

# Chapitre 2

## Méthodes d'estimation

### 2.1 Exemples sur l'estimation ponctuel

#### 2.1.1 Méthode du maximum de vraisemblance

**Exemple 2.1.1** Si  $X$  suit la loi normale  $\mathcal{N}(\mu, \sigma^2)$ , et pour un paramètre multidimensionnel, le principe est le même, mais les calculs d'optimisation sont plus compliqués. Pour les lois normales, deux paramètres sont inconnus. Afin d'éviter les confusions dans les dérivations, nous noterons  $v$  le paramètre de variance, habituellement noté  $\sigma^2$ . Pour  $n$ -uplet de réels  $(x_1, \dots, x_n)$  la vraisemblance vaut :

$$L(x_1, \dots, x_n, \mu, v) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi v}} e^{-\frac{(x_i - \mu)^2}{2v}} = \left( \frac{1}{\sqrt{2\pi v}} \right)^n e^{-\frac{1}{2v} \sum (x_i - \mu)^2}.$$

Son logarithme est :

$$\log(L(x_1, \dots, x_n, \lambda)) = -\frac{n}{2} \log(v) - \frac{n}{2} \log(2\pi) - \frac{1}{2v} \sum (x_i - \mu)^2.$$

Et leurs dérivées partielles par rapport aux paramètres  $\mu$  et  $v$  sont :

$$\frac{\partial \log(L(x_1, \dots, x_n, \lambda))}{\partial \mu} = \frac{1}{v} \sum (x_i - \mu).$$

et :

$$\frac{\partial \log(L(x_1, \dots, x_n, \lambda))}{\partial v} = -\frac{n}{2v} + \frac{1}{2v^2} \sum (x_i - \mu)^2.$$

Elles s'annulent pour :

$$\hat{\mu} = \frac{\sum x_i}{n} \text{ et } \hat{v} = \frac{\sum (x_i - \hat{\mu})^2}{n}.$$

De plus, les dérivées partielles secondes valent :

$$\left\{ \begin{array}{l} \frac{\partial^2 \log(L(x_1, \dots, x_n, \lambda))}{\partial \mu^2} = -\frac{n}{v}, \\ \frac{\partial^2 \log(L(x_1, \dots, x_n, \lambda))}{\partial \mu \partial v} = -\frac{1}{v^2} \sum (x_i - \mu), \\ \frac{\partial^2 \log(L(x_1, \dots, x_n, \lambda))}{\partial v^2} = \frac{n}{2v^2} - \frac{1}{v^3} \sum (x_i - \mu)^2. \end{array} \right.$$

Aussi la matrice hessienne ( matrice des dérivées partielles secondes) au point  $(\hat{\mu}, \hat{v})$  est donc :

$$\begin{pmatrix} -\frac{n}{\hat{v}} & 0 \\ 0 & -\frac{n}{2\hat{v}^2} \end{pmatrix}$$

Ses valeurs propres sont négatives, le point  $(\hat{\mu}, \hat{v})$  est bien un maximum.

Si  $(X_1, \dots, X_n)$  est un échantillon de la loi normale de paramètre  $\mu$  et  $v$ , les estimateurs du maximum de vraisemblance de  $\mu$  et  $v$  sont respectivement la moyenne et la variance empiriques de l'échantillon, comme on pouvait s'y attendre.

**Exemple 2.1.2** Si  $X \rightsquigarrow \mathcal{G}(p)$ , l'ensemble des valeurs possibles pour la loi géométrique est  $\mathbb{N}^*$ . Le paramètre inconnu est  $p \in ]0,1[$ .

Si  $(x_1, \dots, x_n)$  est un échantillon entier, la vraisemblance vaut :

$$L(x_1, \dots, x_n, p) = p^n (1 - p)^{\sum x_i - n}.$$

son logarithme définie par :

$$\log(L(x_1, \dots, x_n, p)) = n \log p + \left( \sum x_i - n \right) \log(1 - p).$$

Et sa dérivée par rapport à  $p$  comme :

$$\frac{\partial \log(L(x_1, \dots, x_n, p))}{\partial p} = n \frac{1}{p} - \left( \sum x_i - n \right) \frac{1}{1-p}.$$

Elle s'annule pour :

$$\hat{p} = \frac{n}{\sum x_i}.$$

De plus, la dérivée seconde est :

$$\frac{\partial^2 \log(L(x_1, \dots, x_n, p))}{\partial p^2} = -n \frac{1}{p^2} - \left( \sum x_i - n \right) \frac{1}{(1-p)^2}.$$

On remarque qu'elle est strictement négative, la valeur  $\hat{p}$  est bien un maximum.

Si  $(X_1, \dots, X_n)$  est un échantillon de la loi géométrique de paramètre  $p$ , l'estimateur du maximum de vraisemblance de  $p$  est :

$$T = \frac{n}{\sum x_i}.$$

D'autre part, et de la même façon pour la loi de poisson  $\mathcal{P}(\lambda)$ , on prend :

$$Q_\theta = \mathbb{P}(\theta),$$

avec :  $\theta \in \Theta = ]0, +\infty[$ , et  $v$  la mesure de comptage sur  $\mathbb{N}$ .

On a :

$$q(\theta, x) = \frac{\theta^x}{x!} e^{-\theta} = e^{-\theta} \exp[x \log(\theta) - \ln(x!)], \text{ pour tout } x \in \mathbb{N}.$$

L'estimateur du maximum de vraisemblance maximisant  $S_n \log(\theta) - n\theta$ , est donné par :

$$T = \bar{X}.$$

## 2.1.2 Méthode des moments

**Exemple 2.1.3** Si  $X \rightsquigarrow \Gamma(\alpha, \lambda)$ , son espérance et sa variance valent :

$$\mathbb{E}[X] = \frac{\alpha}{\lambda} \text{ et } : \text{Var}[X] = \frac{\alpha}{\lambda^2}.$$

On peut donc exprimer  $\alpha$  et  $\lambda$  par :

$$\alpha = \frac{\mathbb{E}[X]^2}{\text{Var}[X]} \text{ et } : \lambda = \frac{\mathbb{E}[X]}{\text{Var}[X]}.$$

On sait que la moyenne empirique  $\bar{X}$  et la variance empirique  $S^2$  sont des estimateurs convergents de  $\mathbb{E}[X]$  et  $\text{Var}[X]$  respectivement, donc on déduit deux estimateurs convergents de  $\alpha$  et  $\lambda$  :

$$A = \frac{\bar{X}^2}{S^2} \text{ et } : \Lambda = \frac{\bar{X}}{S^2}.$$

De plus, si  $X$  suit la loi bêta des paramètres  $\alpha$  et  $b$ , son espérance et sa variance valent :

$$\mathbb{E}[X] = \frac{\alpha}{\alpha + b} \text{ et } : \text{Var}[X] = \frac{\alpha b}{(\alpha + b)^2(\alpha + b + 1)}.$$

On peut exprimer  $\alpha$  et  $b$  par :

$$\alpha = \frac{\mathbb{E}(\mathbb{E} - \mathbb{E}^2 - V)}{V} \text{ et } : b = \frac{\mathbb{E} - 2\mathbb{E}^2 + \mathbb{E}^3 - V + \mathbb{E}V}{V}. \quad (2.1)$$

En ce cas : on en déduit deux estimateurs convergents de  $\alpha$  et  $b$  en remplaçant  $\mathbb{E}$  et  $V$  par leurs estimateurs  $\bar{X}$  et  $S^2$  dans Equation 2.1.

**Exemple 2.1.4** Soit  $X$  suit la loi binomiale négative  $\mathcal{B}(n, p)$ , on veut estimer ses paramètres  $n$  et  $p$ . Premièrement, son espérance et sa variance valent :

$$\mathbb{E}[X] = \frac{n(1-p)}{p} \text{ et } : \text{Var}[X] = \frac{n(1-p)}{p^2}.$$

Aussi on peut exprimer  $n$  et  $p$  par :

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

On en déduit deux estimateurs convergents de  $n$  et  $p$  en remplaçant  $\mathbb{E}[X]$  et  $\text{Var}[X]$  par leurs estimateurs  $\bar{X}$  et  $S^2$  dans ces expressions.

L'inconvénient principal de la méthode des moments est que les estimateurs qu'elle fournit sont en général assez peu précis, et qu'il est difficile d'étudier leur loi autrement que par simulation.

### 2.1.3 Méthode de régression

**Exercice 2.1.1** Pour la régression simple, déterminer la statistique  $\lambda$  du test du rapport de vraisemblance généralisé pour l'hypothèse  $H_0 : a = 0$ , et montrer que c'est une fonction décroissante de la statistique  $F$ .

**Solution 2.1.1** La fonction de vraisemblance des trois paramètres, associée aux réalisations  $y_1, y_2, \dots, y_n$ , est :

$$L(a, b, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} [y_i - (ax_i + b)]^2 \right\}.$$

D'où la log-vraisemblance :

$$\ln L(a, b, \sigma^2) = -\frac{n}{2}(\ln 2\pi + \frac{1}{2}\ln \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (ax_i + b)]^2.$$

Le dénominateur du RVG est obtenu avec les estimations du MV de  $a$ ,  $b$  et  $\sigma^2$  :

$$\begin{aligned} \ln L(\hat{a}, \hat{b}, \hat{\sigma}^2) &= -\frac{n}{2}(\ln 2\pi + \frac{1}{2}\ln \hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n [y_i - (\hat{a}x_i + \hat{b})]^2 \\ &= -\frac{n}{2}(\ln 2\pi + \frac{1}{2}\ln \hat{\sigma}^2) - \frac{n}{2}, \end{aligned}$$

car :

$$\hat{\sigma}^2 = \sum_{i=1}^n [y_i - (\hat{a}x_i + \hat{b})]^2.$$

Pour le numérateur, on se place sous  $H_0 : a = 0$ , c'est-à-dire que les  $Y_i$  sont i.i.d. d'espérance  $b$  et de variance  $\sigma^2$  dont les estimations du MV sont respectivement :  $\bar{y}$  et  $\tilde{s}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ .

Le log-vraisemblance maximale est :

$$\ln L(0, \hat{b}, \tilde{s}^2) = -\frac{n}{2}(\ln 2\pi + \ln \tilde{s}^2) - \frac{n}{2},$$

Finalement :

$$\begin{aligned} \ln RVG &= -\frac{n}{2} \ln \frac{\tilde{s}^2}{\hat{\sigma}^2} = -\frac{n}{2} \ln \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n \hat{\sigma}^2} \\ &= -\frac{n}{2} \ln \left[ \frac{(n-2) \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n \hat{\sigma}^2} - 1}{n-2} + 1 \right]. \end{aligned}$$

Le RVG est une fonction décroissante, la région de rejet de la forme :

$$-2 \ln RVG < k$$

correspond à une région (du test classique) :

$$(n-2) \left[ \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n \hat{\sigma}^2} - 1 \right] < k'.$$

Toutefois la loi asymptotique de  $-2 \ln RVG$  n'est pas en correspondance avec la loi  $F(1, n-2)$  du test classique.

### 2.1.4 Méthode des moindres carrés

**Exemple 2.1.5** Soit  $Y \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$ , et  $Y_1, \dots, Y_n$  les statistiques d'ordre (sont ordonnées de façon décroissante). Si l'hypothèse de normalité est pertinente, alors  $Y_{(i)}$  doit être proche du quantile  $Q_{\mathcal{N}(\mu, \sigma^2)}(i/n)$  de la loi  $\mathcal{N}(\mu, \sigma^2)$ .

Rappelons que si  $X \rightsquigarrow \mathcal{N}(0, 1)$ , alors :

$$Y = \sigma X + \mu \rightsquigarrow \mathcal{N}(\mu, \sigma^2).$$

Donc :

$$Q_{\mathcal{N}(\mu, \sigma^2)}(u) = \sigma Q_{\mathcal{N}(0, 1)}(u) + \mu, \text{ avec } : u \in [0, 1].$$

Et notons  $x_i Q_{\mathcal{N}(0, 1)}(i/n)$  les valeurs de la fonction quantile de la loi  $\mathcal{N}(0, 1)$  aux points  $i/n$ .

Si l'hypothèse de normalité est vérifiée, les points de coordonnées  $(x_i, Y_{(i)})$  devraient être proches de la droite d'équation :

$$y = \sigma x + \mu.$$

Les estimateurs des moindres carrés  $A$  et  $B$  pour la régression linéaire simple des  $Y_{(i)}$  sur les  $x_i$  sont donc des estimateurs de  $\sigma$  et  $\mu$  respectivement.

**Exemple 2.1.6** Soit  $Y$  suit la loi de weibull  $\mathcal{W}(a, \lambda)$ , la fonction quantile de la loi de Weibull  $\mathcal{W}(a, \lambda)$  définie comme :

$$Q_{\mathcal{W}(a, \lambda)}(u) = \left( -\frac{1}{\lambda} \log(1 - u) \right)^{\frac{1}{a}}.$$

La statistique d'ordre  $Y_{(i)}$  doit être proche du quantile  $Q_{\mathcal{W}(a, \lambda)}(i/n)$  :

$$Y_{(i)} \approx \left( -\frac{1}{\lambda} \log\left(1 - \frac{i}{n}\right) \right)^{1/a}.$$

Soit :

$$\log(Y_{(i)}) \approx \frac{1}{a} \log\left(-\log\left(1 - \frac{i}{n}\right)\right) + \frac{1}{a} \log\left(\frac{1}{\lambda}\right).$$

Posons :

$$x_i = \log\left(-\log\left(1 - \frac{i}{n}\right)\right) \text{ et } : Y'_i = \log(Y_{(i)}).$$

Les points  $(x_i, Y'_i)$  devraient être proches de la droite d'équation :

$$y = \frac{1}{a}x + \frac{1}{a} \log\left(\frac{1}{\lambda}\right).$$

Les estimateurs des moindres carrés  $A$  et  $B$  pour la régression linéaire simple des  $Y'_i$  sur les  $x_i$  sont des estimateurs de  $\frac{1}{a}$  et  $\frac{1}{a} \log\left(\frac{1}{\lambda}\right)$  respectivement.

Donc :

$$\hat{a} = \frac{1}{A} \text{ et } : \hat{\lambda} = e^{-\frac{B}{A}}$$

sont des estimateurs de  $a$  et  $\lambda$  respectivement.

### 2.1.5 Estimateur bayésien

**Exemple 2.1.7** Soit  $X \rightsquigarrow \mathcal{G}(\alpha, \theta)$  avec  $\alpha$  supposé connu. Puisque  $\theta$  est un paramètre d'échelle,  $\pi(\theta) = \frac{1}{\theta}$  est une distribution non informative appropriée.

La distribution a posteriori correspondante est  $\mathcal{G}(\alpha, x)$ , et donc :

$$\delta^\pi(x) = \frac{\alpha}{x}$$

est l'estimateur de Bayes généralisé de  $\theta$  sous coût quadratique.

Pour un estimateur de la forme :

$$\delta^\pi(x) = \frac{c}{x},$$

le risque quadratique est :

$$\begin{aligned} R(\theta, \delta_c) &= \mathbb{E}_\theta \left( \frac{c}{x} - \theta \right)^2 \\ &= c^2 \mathbb{E}_\theta(x^{-2}) - 2c\theta \mathbb{E}_\theta(x^{-1}) + \theta^2. \end{aligned}$$

Pour  $\alpha > 2$ , on a :

$$\begin{aligned} \mathbb{E}_\theta(x^{-2}) &= \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} x^{-2} x^{\alpha-1} \theta^\alpha e^{-\theta x} dx \\ &= \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} \theta^\alpha x^{\alpha-3} e^{-\theta x} dx \\ &= \theta^2 \frac{\Gamma(\alpha-2)}{\Gamma(\alpha)} \\ &= \frac{\theta^2}{(\alpha-1)(\alpha-2)}, \end{aligned}$$

et :

$$\mathbb{E}_\theta(x^{-1}) = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} \theta^\alpha x^{\alpha-2} e^{-\theta x} dx = \theta \frac{\Gamma(\alpha-1)}{\Gamma(\alpha)} = \frac{\theta}{\alpha-1}.$$

On en déduit que le meilleur estimateur de la forme  $T = \delta_c$  est associé à :

$$\begin{aligned} c^* &= \frac{\theta \mathbb{E}_\theta(x^{-1})}{\mathbb{E}_\theta(x^{-2})} = \frac{\theta^2 / (\alpha-1)}{\theta^2 / (\alpha-1)(\alpha-2)} \\ &= \alpha - 2. \end{aligned}$$

et donc que  $\delta^\pi$  est dominé par  $\delta_{c^*}$ .

## 2.2 Exemple sur l'estimation par intervalle de confiance

**Exemple 2.2.1** Soit  $(X_1, \dots, X_n)$  un  $n$ -échantillon de v.a.r. de loi  $\mathcal{N}(\mu, \sigma^2)$ .

Estimation de l'espérance  $\mu$  lorsque la variance  $\sigma^2$  connut.

L'estimation par intervalle de confiance consiste à déterminer de part et d'autre de l'estimateur  $T = \bar{X}$  les bornes  $T_1$  et  $T_2$  d'un intervalle qui a un niveau de confiance  $(1 - \alpha)$  de contenir  $\mu$ . Les limites  $T_1$  et  $T_2$  sont telles que :

$$\mathbb{P}(T_1 < \mu < T_2) = 1 - \alpha.$$

ou d'une autre façon :

$$\mathbb{P}(\mu < T_1) = \mathbb{P}(\mu > T_2) = \alpha/2.$$

les limites de confiance peuvent être écrites :

$$T_1 = T - \varepsilon_1 \text{ et } T_2 = T + \varepsilon_2$$

on peut alors écrire :

$$\mathbb{P}(\mu < T - \varepsilon_1) = \mathbb{P}(\mu > T + \varepsilon_2) = \alpha/2,$$

C'est-à-dire que :

$$\mathbb{P}(T - \mu > \varepsilon_1) = \mathbb{P}(\mu - T > \varepsilon_2) = \alpha/2.$$

Comme, pour une population normale, la variable  $\bar{X}$  est elle-même normale de moyenne  $\mu$  et d'écart-type  $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ , on peut écrire :

$$\mathbb{P}\left(\frac{T - \mu}{\sigma/\sqrt{n}} > \frac{\varepsilon_1}{\sigma/\sqrt{n}}\right) = \mathbb{P}\left(\frac{\mu - T}{\sigma/\sqrt{n}} > \frac{\varepsilon_2}{\sigma/\sqrt{n}}\right) = \frac{\alpha}{2},$$

Équivalent à :

$$\begin{aligned} \mathbb{P}\left(\mathcal{Z}_1 > \frac{\varepsilon_1}{\sigma/\sqrt{n}}\right) &= \mathbb{P}\left(\mathcal{Z}_2 > \frac{\varepsilon_2}{\sigma/\sqrt{n}}\right) = \frac{\alpha}{2}, \\ \mathbb{P}\left(\mathcal{Z}_1 < \frac{\varepsilon_1}{\sigma/\sqrt{n}}\right) &= \mathbb{P}\left(\mathcal{Z}_2 < \frac{\varepsilon_2}{\sigma/\sqrt{n}}\right) = 1 - \frac{\alpha}{2}. \end{aligned}$$

Donc :

$$\prod \frac{\varepsilon_1}{\sigma/\sqrt{n}} = \prod \frac{\varepsilon_2}{\sigma/\sqrt{n}} = 1 - \frac{\alpha}{2}.$$

Si on désigne par  $Z_{1-\frac{\alpha}{2}}$  la valeur de la variable normale réduite lue dans la table :

$$\frac{\varepsilon_1}{\sigma/\sqrt{n}} = \frac{\varepsilon_2}{\sigma/\sqrt{n}} = 1 - \frac{\alpha}{2},$$

il en résulte :

$$\varepsilon_1 = \varepsilon_2 = Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Les limites de confiances sont donc :

$$T_1 = T - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \text{ et } T_2 = T + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

On notera l'intervalle de confiance :

$$T \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

C'est un intervalle symétrique par rapport à la moyenne.

### Estimation de l'espérance $\mu$ lorsque la variance $\sigma^2$ est inconnue

Pour une population de distribution de probabilité inconnue, on utilise la quasi-variance comme estimation de la variance de la population. L'intervalle de confiance de la moyenne sera défini selon les cas :

#### **Cas d'un échantillon d'effectif inférieur à 30 ( $n < 30$ ) :**

Dans ce cas, la moyenne d'un échantillon peut toujours être considérée comme une variable  $T$  de Student à  $(n - 1)$  degré de liberté. La valeur  $Z_{1-\frac{\alpha}{2}}$  sera remplacée par la valeur  $T_{1-\frac{\alpha}{2}}$  à  $(n - 1)$  degré de liberté.

L'intervalle de confiance est alors :

$$T \pm \mathcal{T}_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}.$$

**Cas d'un échantillon d'effectif supérieur ou égal à 30 ( $n \geq 30$ ) :**

Dans ce cas, la moyenne d'un échantillon peut toujours être considérée comme une variable approximativement normale. L'intervalle de confiance est alors :

$$T \pm \mathcal{Z}_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}.$$

### Estimation de la variance $\sigma^2$

Si on s'intéresse à la variance  $\sigma^2$  d'une population normale, l'estimation par intervalle de confiance consiste à déterminer les bornes  $T_1$  et  $T_2$  d'un intervalle qui a un niveau de confiance  $(1 - \alpha)$  de contenir  $\sigma^2$ .

Les limites  $T_1$  et  $T_2$  sont telles que :

$$\mathbb{P}(T_1 \leq \sigma^2 \leq T_2) = 1 - \alpha.$$

Comme, pour une population normale, la variable aléatoire  $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2}$  possède une distribution de khi-deux à  $(n - 1)$  degré de liberté, on peut alors écrire :

$$\mathbb{P} \left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{T_2} \leq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} \leq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{T_1} \right) = 1 - \alpha.$$

ou encore :

$$\mathbb{P} \left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} < \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{T_2} \right) = \mathbb{P} \left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} > \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{T_1} \right) = \frac{\alpha}{2},$$

et :

$$\mathbb{P} \left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} \leq \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{T_1} \right) = 1 - \frac{\alpha}{2} \Rightarrow \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{T_1} = \chi_{1-\frac{\alpha}{2}}^2,$$

$$\mathbb{P} \left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sigma^2} < \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{T_2} \right) = \frac{\alpha}{2} \Rightarrow \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{T_2} = \chi_{\frac{\alpha}{2}}^2.$$

Les limites de confiances sont alors :

$$\sigma_1^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\chi_{1-\frac{\alpha}{2}}^2} \quad \text{et} \quad \sigma_2^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\chi_{\frac{\alpha}{2}}^2}.$$

Les valeurs de  $\chi_{\frac{\alpha}{2}}^2$  et  $\chi_{1-\frac{\alpha}{2}}^2$  sont à  $(n-1)$  degré de liberté.

Intervalle de confiance de la proportion

### Estimation de la proportion p

Si on s'intéresse à la proportion  $p$ , l'estimation par intervalle de confiance consiste à déterminer de part et d'autre de l'estimateur  $F_n$  les bornes  $p_1$  et  $p_2$  d'un intervalle qui a un niveau de confiance  $(1 - \alpha)$  de contenir  $p$ .

Les limites  $p_1$  et  $p_2$  sont telles que :

$$\mathbb{P}(T_1 \leq p \leq T_2) = 1 - \alpha,$$

ou d'une autre façon :

$$\mathbb{P}(p < T_1) = \mathbb{P}(p > T_2) = \frac{\alpha}{2},$$

les limites de confiance peuvent être écrites :

$$T_1 = f_n - \varepsilon_1 \quad \text{et} \quad T_2 = f_n + \varepsilon_2,$$

on peut alors écrire :

$$\begin{aligned}\mathbb{P}(p < f_n - \varepsilon_1) &= \mathbb{P}(p > f_n + \varepsilon_2) = \frac{\alpha}{2}, \\ \mathbb{P}(f_n - p > \varepsilon_1) &= \mathbb{P}(p - f_n > \varepsilon_2) = \frac{\alpha}{2}.\end{aligned}$$

Comme, la distribution de la proportion suit une loi normale de moyenne  $p$  et d'écart-type  $\sigma_{F_n} = \frac{\sqrt{pq}}{\sqrt{n}}$ , à condition que la taille de l'échantillon soit supérieure ou égale à 30 ( $n \geq 30$ ) et le produit  $np \geq 5$ ; on peut écrire :

$$\mathbb{P}\left(\frac{f_n - p}{\sqrt{p(1-p)}/\sqrt{n}} > \frac{\varepsilon_1}{\sqrt{p(1-p)}/\sqrt{n}}\right) = \mathbb{P}\left(\frac{p - f_n}{\sqrt{p(1-p)}/\sqrt{n}} > \frac{\varepsilon_2}{\sqrt{p(1-p)}/\sqrt{n}}\right) = \frac{\alpha}{2};$$

Équivalent à :

$$\begin{aligned}\mathbb{P}\left(\mathcal{Z}_1 > \frac{\varepsilon_1}{\sqrt{p(1-p)}/\sqrt{n}}\right) &= \mathbb{P}\left(\mathcal{Z}_2 > \frac{\varepsilon_2}{\sqrt{p(1-p)}/\sqrt{n}}\right) = \frac{\alpha}{2}, \\ \mathbb{P}\left(\mathcal{Z}_1 < \frac{\varepsilon_1}{\sqrt{p(1-p)}/\sqrt{n}}\right) &= \mathbb{P}\left(\mathcal{Z}_2 < \frac{\varepsilon_2}{\sqrt{p(1-p)}/\sqrt{n}}\right) = 1 - \frac{\alpha}{2}.\end{aligned}$$

Donc :

$$\prod \frac{\varepsilon_1}{\sqrt{p(1-p)}/\sqrt{n}} = \prod \frac{\varepsilon_2}{\sqrt{p(1-p)}/\sqrt{n}} = 1 - \frac{\alpha}{2}.$$

Si on désigne par  $\mathcal{Z}_{1-\frac{\alpha}{2}}$  la valeur de la variable normale réduite lue dans la table :

$$\frac{\varepsilon_1}{\sqrt{p(1-p)}/\sqrt{n}} = \frac{\varepsilon_2}{\sqrt{p(1-p)}/\sqrt{n}} = \mathcal{Z}_{1-\frac{\alpha}{2}},$$

il en résulte :

$$\varepsilon_1 = \varepsilon_2 = \mathcal{Z}_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}.$$

Les limites de confiances sont donc :

$$T_1 = f_n - \mathcal{Z}_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \quad \text{et} \quad T_2 = f_n + \mathcal{Z}_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}.$$

On notera l'intervalle de confiance :

$$f_n \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}.$$

La proportion  $p$  de la population sera estimée par la fréquence  $f_n$  de l'échantillon. On obtient ainsi un intervalle symétrique par rapport à la proportion.

**Exercice 2.2.1** Dans une entreprise produisant un article déterminé on veut estimer sa durée de vie en heures. À cette fin on a observé un échantillon aléatoire Table 2.1 et simple de 16 unités dont les résultats sont (en 1000 heures) :

1.10	1.05	1.25	1.08	1.35	1.15	1.30	1.25
1.30	1.35	1.15	1.32	1.05	1.25	1.10	1.15

TAB. 2.1 – La durée de vie en heures du produit.

- Estimer la durée de vie moyenne  $\mu$  et l'écart-type  $\sigma$  d'un article.
- Donner des estimations par intervalle de confiance pour  $\mu$  et  $\sigma$  ( $\alpha = 5\%$ ).

**Solution 2.2.1** L'estimation de la durée de vie moyenne  $\mu$  et de l'écart-type  $\sigma$  d'un article

L'estimation ponctuelle de la moyenne de la population est :

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{16} x_i}{16} = 1.2$$

L'estimation ponctuelle de l'écart-type  $\sigma$  de la population est :

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{16} (x_i - \bar{x})^2}{16 - 1}} = 0.11$$

intervalle de confiance de la moyenne  $\mu$  et de l'écart-type  $\sigma$

- L'intervalle de confiance de la moyenne  $\mu$  à un niveau de confiance de 95% ( $\alpha = 5\%$ ) :

La distribution de la population parent étant inconnue et la taille de l'échantillon inférieure à 30, l'intervalle de confiance de la moyenne est défini par :

$$\bar{X} \pm T_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}$$

La valeur de  $T_{1-\frac{\alpha}{2}}$  à 15 degrés de liberté est :  $t_{0.975} = 2.131$ , l'intervalle de confiance est :

$$\bar{X} \pm T_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} = 1.2 \pm 2.131 \frac{0.11}{\sqrt{16}}$$

Donc :

$$T_1 = 1.2 - 2.131 \frac{0.11}{\sqrt{16}} = 1.14 \text{ et } T_2 = 1.2 + 2.131 \frac{0.11}{\sqrt{16}} = 1.26$$

L'intervalle  $[1.14, 1.26]$  a une probabilité de 95% de contenir la vraie valeur de la moyenne de la population.

• L'intervalle de confiance de l'écart type  $\sigma$  à un niveau de confiance de 95% ( $\alpha = 5\%$ ) :

$$T_1 = \sqrt{\frac{\sum_{i=1}^{16} (x_i - \bar{x})^2}{\chi_{1-\frac{\alpha}{2}}^2}} = \sqrt{\frac{0.11^2 \times 15}{27.49}} = 0.08 \text{ et } T_2 = \sqrt{\frac{\sum_{i=1}^{16} (x_i - \bar{x})^2}{\chi_{\frac{\alpha}{2}}^2}} = \sqrt{\frac{0.11^2 \times 15}{6.26}} = 0.17.$$

**Exercice 2.2.2** On étudie le pourcentage d'utilisation d'une machine. 400 observations ont été effectuées qui ont donné le résultat suivant :

- $$\left\{ \begin{array}{l} \bullet \text{ Machine marche : 320 observations.} \\ \bullet \text{ Machine arrêtée : 80 observations.} \end{array} \right.$$

– Entre quelles limites peut-on fixer le taux d'utilisation de la machine avec un degré de confiance de 95% ?

– On fait un plus grand nombre d'observations. On obtient le même pourcentage d'utilisation ce qui permet, avec un risque d'erreur de 5%, de fixer les limites de confiance à  $[78.4\%, 81.6\%]$ . Combien a-t-on fait d'observations ?

**Solution 2.2.2** intervalle de confiance de la proportion  $p$ 

L'estimation ponctuelle de la proportion d'utilisation de la machine est :

$$\hat{p} = f_n = \frac{320}{400} = 0.8$$

Le taux d'utilisation de la machine est estimé à 80%.

L'intervalle de confiance de la proportion à un niveau de confiance de 95% est défini par :

$$f_n \pm \mathcal{Z}_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

La valeur de  $\mathcal{Z}_{1-\frac{\alpha}{2}}$  est :  $\mathcal{Z}_{0.975} = 1.96$ .

Les limites de confiances de la proportion sont :

$$T_1 = f_n - \mathcal{Z}_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} = 0.80 - 1.96 \sqrt{\frac{0.8(1-0.8)}{400}} = 0.76,$$

et :

$$T_2 = f_n + \mathcal{Z}_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} = 0.80 + 1.96 \sqrt{\frac{0.8(1-0.8)}{400}} = 0.84.$$

L'intervalle [76%, 84%] a une probabilité de 95% de contenir le vrai taux d'utilisation de la machine.

# Conclusion

Dans ce mémoire on a proposé des méthodes statistiques de calcul des estimateurs pour des paramètres d'échantillon (l'espérance, par exemple) qui permettent des procédures d'estimation et de choix de la valeur la plus proche à la valeur théorique de spécification de la forme des lois des probabilités. Ces méthodes sont facilement mises en œuvre à l'aide de logiciels modernes comme Gauss qui suit la réalité de beaucoup plus près que la méthode. L'exemple montre qu'elles peuvent donner des résultats très différents de ceux fournis par une méthode classique comme la méthode *MV*.

Il est intéressant de comparer les estimateurs par rapport à ses valeurs, et tels qu'ils sont de précision d'aspects théoriques, variances, résidus. Dans les cas simples, on peut tout estimer, et dans les cas complexes, il faut faire une sélection. On prise en compte l'information disponible, est représentée généralement dans les données et de l'information a priori, mais jusqu'à ce travail, presque nos paramètres sont estimés à partir des données, sans utiliser l'information a priori, aussi les méthodes Bayésiennes sont utiles pour estimer les paramètres à la fois à partir des données et de l'information a priori.

En particulier pour la méthode des moindres carrés, il y a plusieurs types de mesures et la variance des mesures dépendent de la date d'observation, et l'idée est de donner moins de poids aux mesures dont la variance est grande ; chaque mesure correspond à la moyenne de trois répétitions. On peut calculer les variances à partir des valeurs élémentaires obtenues sur les répétitions.

Dans certains cas, l'estimation par la méthode des moments est moins bonne que l'es-

timation par maximum de vraisemblance. Néanmoins, dans le cas de la loi Gamma par exemple, le calcul de la fonction de vraisemblance peut poser des problèmes (l'utilisation de l'ordinateur et d'algorithmes numériques est indispensable) tandis que l'estimation des moments est très facilement accessible.

On retrouve le fait, lorsque la taille de l'échantillon n'est pas suffisamment grande, la loi des grands nombres ne s'applique pas et par conséquent, les moments empiriques n'approchent pas suffisamment les moments théoriques.

# Bibliographie

- [1] Chouquet C. ( 2009 – 2010). M1IMAT : Modeles linéaires. Laboratoire de statistique et probabilités, université Paul Sabatier. Toulouse. (<https://www.math.univ-toulouse.fr/~bartheM1modlinpoly.pdf>).
- [2] Christian, P. R. (2006). Le choix Bayésien : Principes et pratique.
- [3] Delyon B. ( 29 mai 2018). Cours de master 2 : Estimation paramétrique. IRMAR, université Rennes I, Campus de Beaulieu, 35042 Rennes cédex, France. (<https://perso.univ-rennes1.fr/bernard.delyon/param.pdf>).
- [4] Dress, F. (2007). Les probabilités et la statistique de A à Z-500 définitions, formules et tests d’hypothèse : 500 définitions, formules et tests d’hypothèse. Hors collection. Dunod.
- [5] El Marhoum A. ( 2014 – 2015). Cours d’échantillonnage et estimations. Faculte des sciences juridiques, économiques et sociales, AGDAL. Université mohammed V, RABAT ( Maroc ) .
- [6] Makowski, D. (28 nov–01 déc 2005). Estimation des paramètres des modèles : principes généraux. UMR Agronomie INRA/INA-PG. ([http://www.modelia.org/html/050929\\_formationIntroModDec2005/supportsDesInterventions/M](http://www.modelia.org/html/050929_formationIntroModDec2005/supportsDesInterventions/M)
- [7] Méthode des moindres carrés : chapitre 5. (<https://math.unice.fr/~diener/MAB07/MCO.pdf>)
- [8] OLIVARES, P. (2005). ALEXANDER ALVAREZ. Journal de la société française de statistique, 146(4), 23 – 54.

- [9] Tugaut, J. Statistiques : quelque méthodes d'estimation. Télécom Saint-Etienne. (<http://tugaut.perso.math.cnrs.fr/pdf/enseignement/2017/SFI/CM03.pdf>)
- [10] Wikistat. ( 2016). Statistique niveau M1 : Estimation. Université de Toulouse; Université Toulouse *III* Paul Sabatier; INSA Toulouse; Institut de Mathématiques de Toulouse; Toulouse School of Economics). France. (<https://www.math.univ-toulouse.fr/~besse/Wikistat/>).
- [11] Yao J. Cours de master 1 : Statistique et logiciel. Université de Rennes 1. (<http://perso.univ-rennes1.fr/jian-feng.yao/pedago>).
- [12] Ycart, B. (2002). Estimation paramétrique, tests statistiques. Centre de Publication Universitaire.

# Annexe A : Définitions

**Définition 2.2.1** *Loi des grands nombres :*

*On définit les moyennes :*

$$\mu_i = \frac{\sum_{k=1}^i x_k}{i}, \forall i = \overline{1, n},$$

*Alors :*

$$\mathbb{P}(X_n \rightarrow \mu) = 1 \text{ quand } : n \rightarrow \infty.$$

**Définition 2.2.2** *Fraction ou taux :*

*C'est la proportion de l'action qui font partie de l'échantillon. C'est le rapport entre l'effectif partiel  $i$ , et l'effectif total de la population  $n$  :*

$$f_i = \frac{i}{n} \times 100.$$

**Définition 2.2.3** *Théorème de factorisation :*

*Si la vraisemblance de l'observation  $X$  s'écrit sous la forme :*

$$f(X, \theta) = h(X)g_\theta(T(X)),$$

*alors  $T(X)$  est une statistique exhaustive.*

**Définition 2.2.4** *Modèle exponentiel :*

*S'il existe une mesure de référence  $\delta$  telle que la loi  $P_\theta$  admette une densité par rapport*

à  $\delta$  de la forme :

$$f(\theta, x) = C(\theta) \exp(\langle T(x), Q(\theta) \rangle)$$

le modèle est dit exponentiel.

**Définition 2.2.5** *risque quadratique :*

Soit  $T$  une statistique exhaustif et  $U$  un estimateur de  $g(\theta)$  tel que :

$$\mathbb{E}_\theta(U^2) < +\infty, \forall \theta \in \Theta.$$

Et soit :

$$T^* = \mathbb{E}_\theta(U/T),$$

$T^* \Pi \theta$  et pour tout  $\theta \in \Theta$  :

$$\mathbb{E}_\theta((T^* - g(\theta))^2) \leq \mathbb{E}_\theta((U - g(\theta))^2),$$

ce qui signifie que le risque quadratique de  $T^*$  pour estimer  $g(\theta)$  est inférieur ou égal à celui de  $U$ .

**Définition 2.2.6** *Test du rapport de vraisemblance généralisé RVG*

Soit la famille paramétrique  $\{f(x, \theta), \theta \in \Theta \subseteq \mathbb{R}^d\}$ , et les hypothèses :

$$H_0 : \theta \in \Theta_0 \text{ vs } H_1 : \theta \in \Theta_0^c,$$

où  $\Theta_1 = \Theta - \Theta_0$  est le complémentaire de  $\Theta_0$  par rapport à  $\Theta$ . On appelle rapport de vraisemblance généralisé la fonction  $\lambda(x_1, x_2, \dots, x_n)$  telle que :

$$\lambda(x_1, x_2, \dots, x_n) = \frac{\sup_{\theta \in \Theta_0} L(\theta, x_1, x_2, \dots, x_n)}{\sup_{\theta \in \Theta} L(\theta, x_1, x_2, \dots, x_n)},$$

et le test du RVG définit par une région de rejet de la forme :

$$\lambda(x_1, x_2, \dots, x_n) < k \leq 1.$$

De plus, s'il existe une estimation du maximum de vraisemblance  $T$  tel que le dénominateur est la valeur de la fonction de vraisemblance en  $T$ , soit :  $L(T, x_1, x_2, \dots, x_n)$ .

**Définition 2.2.7**    *Quantité d'information de Fisher*

On appelle quantité d'information de Fisher  $I_n(\theta)$  apportée par un  $n$ -échantillon sur le paramètre la quantité suivante (si elle existe) :

$$I_n(\theta) = \mathbb{E} \left[ \left( \frac{\partial \ln L}{\partial \theta} \right)^2 \right].$$

**Remarque 2.2.1** *L'information de Fisher est utilisé pour trouver des estimateurs sans biais optimaux.*

# Annexe B : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous.

$(X_1, \dots, X_n)$	Échantillon de taille $n$ de variables aléatoire.
$(x_1, \dots, x_n)$	Échantillon de taille $n$ d'observations.
$\sum$	La somme finie lorsque $i = \overline{1, n}$ .
$n$	L'effectif totale.
$p.s$	convergence presque sûre.
$\Pi$	Signifie l'indépendance.
$C$	C'est une fonction à valeurs réel.
$Q$	Application mesurable.
$I$	Un intervalle supérieur à 0 sur $\Theta$ .
$C^1$	La classe des fonction vérifie la continuité de sa première dérivée.
$c$	Un valeur réel.
$\delta_c$	Un estimateur exprimé par valeur inconnu.
$\varepsilon$	Un valeur aléatoire appartenent à $\mathbb{R}_+$ , et proche plus en plus à 0.
$MV$	Maximum de vraisemblance.
$\xrightarrow{p}$	La convergence en probabilité.
$\theta_0$	Un estimateur $\in \Theta$ .

$p.p$	presque partout.
$\omega$	Un mesure.
$p$	Un estimateur de Pitman.
$p_\theta$	Un paramètre de translation : $p_\theta(x) = p_0(x - \theta)$ .
$k$	Un valeur réelle.
$\mathcal{Z}$	Une distribution normale réduite.
$\mathcal{X}$	Une distribution de Khi-deux.
$\mathcal{T}$	Une distribution de Student.