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

RAHAL Hafed

Titre :

Estimation des paramètres par échantillonnage

Membres du Comité d'Examen :

Dr.	**BENATIA Fateh**	UMKB	Président
Dr.	**TOUBA Sounia**	UMKB	Encadreur
Dr.	**DHIABI Samra**	UMKB	Examinateur

Juin 2018

Dédicace

Je dédie ce mémoire à :

Mes parents

Mon père, qui est parti de nous, mais Son Esprit est avec nous < Rahimallâhu abî rahmatan wâssi'a... >, qui peut être fier de moi et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie, l'éducation et le soutient permanent venu de toi.

Invocations pour mon père : que Dieu lui accorde Son Pardon complet et immédiat, et qu'Il fasse de sa tombe un jardin parmi les jardins du Paradis.

Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments.

A mes frères,

Et

Et toutes mes amies

REMERCIEMENTS

Je tiens à remercier en premier lieu, **Allah** Le tout-puissant, qui m'a donné la volonté, la force et les bonnes chances pour réussir.

Je veux exprimer ma profonde gratitude à mon encadreuse Madame **Dr.TOUBA Sounia** pour la connaissance qu'il m'a accordé en acceptant de m'encadrer. Je La remercier profondément, pour ses conseils avisés, leur encouragements, sa patience, ses multiples relectures et le meilleur encadrement possible qu'il m'a offert.

Avec un grand honneur, j'aimerais présenter mes remerciements et ma gratitude aux membres du jury, Monsieur **Dr.BENATIA Fateh**.et Madame **Dr.DHIABI Samra** , d'avoir examiné et évaluer mon travail.

Je n'oublie pas de remercier mes enseignants, surtout Le chef du département de mathématique Monsieur.

Je voudrais dire :

Merci **Papa** d'être pour moi le meilleur des père ,Qu' ALLAH lui fasse miséricorde.

Merci **Maman** pour m'avoir montré la beauté de la vie.

Je remercie chaleureusement, mes frères et sœurs, et je souhaite à leur tout la réussite.

Table des matières

Table des figures

Introduction

En statistique, un échantillon est un ensemble d'individu représentatif d'une population. L'objectif est d'obtenir une meilleure connaissance de la population par l'étude d'un seul échantillon. Le recours à un de l'échantillon répond en générale à une contrainte pratique (manque de temps, de place, évaluation destructive d'une production, cout financier... etc.) interdisant l'étude exhaustive de la population.

L'acte de sélection s'appelle l'échantillonnage. Pour garantir une bonne représentation, il faut choisir en général un échantillon, totalement ou partiellement aléatoire. La statistique sert donc intéressée aux principes d'échantillonnage, dans le but d'assurer ou au moins d'estimer la fiabilité de tirer des conclusions sure l'étude l'échantillon, mais études aux populations entières.

Quelque une des préoccupations de la théorie de l'échantillonnage sont :

- Capacité a capté la diversité du phénomène étudié.
- L'absence de biais ou d'erreur systématique.
- Le lien entre la taille de l'échantillon et la confiance que l'on peut accorder à la généralisation des résultats. Les erreurs connu sont généralement mesurées par des probabilités telles que l'espérance, la variance...

L'organisation de ce mémoire est la suivante :

Ce travail est organisé en trois chapitres, les deux premiers servent à une étude purement théorique du sujet, et l'autre sert à une application de ces résultats sous un environnement de traitement et d'analyse R.

Le premier chapitre rappelle les principaux concepts qu'ils convient de connaître dans le cadre de l'échantillonnage aléatoire simple, on a introduits le chapitre par des notions sur les variables aléatoires discrètes et continue, ensuite, on va parler sur les lois usuelles. Apre avoir les méthodes les lois fondamentales d'échantillonnages et nous terminons par les différents type de convergence.

Suivis du chapitre deux qui offre une présentation détaillé sur les caractéristique d'un échantillon, il constitue l'estimation en générale avec l'explication de la moyenne empirique et variance empirique. On se termine par une présentation sur les fonctions de distribution empirique.

Le dernier chapitre trois est consacré à l'application des différents résultats obtenus par le logiciel R

Nous souhaitons que toutes les personnes désireuses étudier l'échantillonnage trouveront la lecture de ce mémoire très utile et agréable.

Chapitre 1

Echantillonnage aléatoire simple

La technique des sondages (échantillonnages) permet de produire de l'information sur un domaine donné à partir de l'observation d'une partie de ce domaine. Elle s'applique particulièrement à l'étude des populations nombreuses. Avant d'aborder les différentes méthodes de sondage, il est mécessaire de présenter un certain nombre de notions qui seront utilisées tout au long de cet ouvrage.

Voici quelques définition et caracteristique de base :

Définition 1.0.1 (enquête = sondage) *Une ènquète est une activité organisée et méthodique de collecte de données sur des caractéristiques d'intérêt d'une partie ou de la totalité des unitès d'une population à l'aide de concepts, de méthodes et de procédures biens définis.*

Définition 1.0.2 (élèment) *Un élement est un objet sur lequel on va mesurer les caractéristiques d'intérêt.*

Définition 1.0.3 (population) *Une population notée E est un ensemble d'élèments.*

Définition 1.0.4 (population cible) *Une population cible est une population pour laquelle l'information est requise .*

Définition 1.0.5 (Unité) *Les unités d'échantillonnage sont des entités disjointes dont l'union est ègale à la population.*

Définition 1.0.6 (échantillon) *Un échantillon noté $\xi_n$de E est dit de taille n s'il est constitué de n unités distinctes ou non, tirées parmi les N élèments de la population.*

Définition 1.0.7 (base de sondage) *La base de sondage donne les moyens d'identifier les unités d'échantillonnage de la population cible et de communiquer avec elle. C'est à partir de la base de sondage que l'on va sélectionner les unités d'échantillonnage qui vont être enquêtées.*

On écrit :

$$E = \{e_1, ..., e_N\} : \text{base de sondage.}$$
$$\xi_n = \{e_{i_1}, ..., e_{i_n}\} = \{e_{i_k}\}_{1 \leq i_k \leq N} \text{ avec } 1 \leq k \leq n.$$

Le tirage d'un échantillon de n éléments à partir de la population E correspond aux données de cet échantillon qui comprend diverses valeurs décrites par le caractére X. Il est a noter que les calculs seront effectués sur l'échantillon $\{X(\xi_n)\}$, qu'on appel les réalisations X sur ξ_n.(voir R Clairin, P Brion [5])

Exemple 1.0.1 *On considére une population de* 40 *personnes.*

$$E = \{e_1, ..., e_{40}\}; N = 40.$$

Apartir de E, on extrait cinq personnes afin de construire un ensemble constituant l'échantillon de taille 5.

$$\xi_n = \{e_1, e_4, e_6, e_{30}, e_{36}\}; n = 5.$$

1.1 Variables aléatoires

Il existe deux types des variables aléatoires (v.as), les v.as discrètes et les v.as continues dont on va définir ci-dessous.

1.1.1 Variable aléatoire discrète

On appelle v.a. discrète définie sur l'espace probabilisable $(\Omega, \mathcal{A})$, l'application : $X : \Omega \to \mathbb{R}$, sachant que $X(\Omega)$ est dénombrable (fini) telle que

$$\forall x \in \mathbb{R} : X^{-1}(x) = \{w \in \Omega / X(w) = x\} \in \mathcal{A}.$$

Voici quelques caractéristiques de ce type de v.a :

1. Fonction de répartition (fd) de la v.a. X notée $\mathbb{P}(X = x_i)$ avec :

$$\sum_{i \in \mathbb{N}} \mathbb{P}(X = x_i) = 1.$$

2. Espérence mathématique :

$$E(X) = \sum_{i \in \mathbb{N}} x_i \mathbb{P}(X = x_i) = \mu.$$

3. Variance :

$$Var(X) = \sum_{i \in \mathbb{N}} (x_i - \mu)^2 \mathbb{P}(X = x_i).$$

4. Fonction caractèristique :

$$\phi_X(t) = E(e^{itX}) = \sum_{i \in \mathbb{N}} e^{itx_i} \mathbb{P}(X = x_i).$$

Exemple 1.1.1 *On lance une piéce de monnaie, on obtient les deux réalisations suivantes :*

$$X = \begin{cases} 0 & \text{si la pièce est pile,} \\ 1 & \text{si la pièce est face.} \end{cases}$$

On calcul l'spérence et la variance mathématique :

$E(X) = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2}$ avec $\mathbb{P}(X = x_i) = \frac{1}{2}$.

$Var(X) = \frac{1}{2}\left(0 - \frac{1}{2}\right)^2 + \frac{1}{2}\left(1 - \frac{1}{2}\right)^2 = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$.

1.1.2 Variable aléatoire continue

On dit que X est une v.a. continue si elle admet une fonction de densité notée f_X définie de $\mathbb{R}$ dans $\mathbb{R}$.Voici quelques caractéristiques de ce type de v.a. :

1. Fonction de répartition (fd) de la v.a. X :

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

2. Espérence mathématique :

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\, dx = \mu.$$

3. Variance :

$$\begin{aligned} Var(X) = E[X - E(X)]^2 &= \int_{-\infty}^{+\infty} [x - E(X)]^2 f_X(x)\, dx \\ &= E\left(X^2\right) - (E(X))^2 \\ &= \sigma^2. \end{aligned}$$

4. Fonction caractèristique :

$$\phi_X(t) = E(\exp(itX)) = \int_{-\infty}^{+\infty} \exp(itx) f_X(x)\, dx.$$

Proposition 1.1.1 *Soient X et Y deux v.as. réellesn, On a :*

- $\forall a \in \mathbb{R}, \forall b \in \mathbb{R} : E(aX + b) = aE(X) + b.$
- $\forall \lambda \in \mathbb{R}, \forall \mu \in \mathbb{R} : E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y).$
- $Var(aX + b) = a^2 Var(X).$
- $Var(X + a) = Var(X).$

1.2 Lois usuelles

Il existe deux types de les lois usuelles, les lois usuelles discrètes et les lois usuelles continues.

1.2.1 Lois usuelles discrètes

De nombreuses lois discrètes ont étaient élaborées dans la littérature parmi elles :

Loi de Bernoulli $\mathcal{B}(p)$

On dit que X suit une loi de Bernoulli de paramètre $p \in [0, 1]$ notée $X \sim \mathcal{B}(p)$ si :

$$X = \begin{cases} 1 & \text{si } \mathbb{P}(X = x_i) = p, \\ 0 & \text{si } \mathbb{P}(X = x_i) = q = 1 - p. \end{cases}$$

- $E(X) = 1 \times \mathbb{P}(X = x_1) + 0 \times \mathbb{P}(X = x_2) = \mathbb{P}(X = x_1) = p.$
- $E(X^2) = 1^2 \times \mathbb{P}(X = x_1) + 0^2 \times \mathbb{P}(X = x_2) = p.$
- $Var(X) = E(X^2) - (E(X))^2 = p(1 - p) = pq.$
- $\Phi_X(t) = E[\exp(itx)] = 1 - p + p\exp(it) = q + p\exp(it)$

Loi de binomiale $\mathcal{B}(n,p)$

Soit X une v.a caractérisée par $X(\Omega)=\{1,...n\}$ suit la loi binomiale de paramètres $n\geq 1$ et $p\in[0,1]$. Pour $k\in X(\Omega)$ on a :

$$\mathbb{P}_X(X=k)=C_n^k p^k q^{n-k},$$

où $C_n^k=n!/k!\,(n-k)!$ est le coefficient du binôme.

Remarque 1.2.1 *La loi binomiale est l'espérence de Bernoulli répétée n fois de manière indépendante c'est-à-dire (i.e) que X est la somme des n v.as. de Bernoulli indépendante de méme paramètre p*

On peut alors écrire :

$$X=\sum_{i=1}^{n}X_i.$$

D'où on déduit aisément que :

- $E(X)=E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E(X_i)=np.$
- $Var(X)=Var\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}Var(X_i)=np(1-p)=npq.$
- $\Phi_X(t)=E[\exp(itX)]=[q+p\exp(it)]^n.$

Exemple 1.2.1 *Le nombre des résultats pile apparus au cours de n jets d'une piéce de monnaie suit une loi $\mathcal{B}(n,1/2)$.*

$$\begin{aligned}P_X(X=k) &= C_n^k\left(\tfrac{1}{2}\right)^k\left(1-\tfrac{1}{2}\right)^{n-k}\\ &= C_n^k\left(\tfrac{1}{2}\right)^k\left(\tfrac{1}{2}\right)^{n-k}\\ &= C_n^k/2^n.\end{aligned}$$

avec :

$E(X)=np=n\times\frac{1}{2}=\frac{n}{2}.$

$Var(X)=npq=n\times\frac{1}{2}\left(1-\frac{1}{2}\right)=\frac{n}{4}.$

Loi de Poisson $\mathcal{P}(\lambda)$

Une v.a X suit une loi de Poisson de paramètre $\lambda > 0$ définie comme ci-dessous :

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

Le dévelopement en série entiére de l'exponentielle :

$$e^{\lambda} = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}.$$

Proposition 1.2.1 *Soit $X \sim \mathcal{P}(\lambda)$, alors :*

- $E(X) = \lambda$.
- $Var(X) = \lambda$.
- $\Phi_X(t) = \exp[\lambda(\exp(it) - 1)]$.

Preuve.

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty} k\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda}e^{\lambda} \\ &= \lambda. \end{aligned}$$

Afin de calculer la variance, on va d'abords calculer $E(X^2)$. Pour cela on calcule le moment factoriale $E(X(X-1)) = E(X^2) - E(X)$.

$$\begin{aligned} E(X(X-1)) &= e^{-\lambda}\sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!} \\ &= \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda}e^{\lambda} \\ &= \lambda^2. \end{aligned}$$

On déduit alors que :

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= E(X(X-1)) + E(X) - (E(X))^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda. \end{aligned}$$

Et enfin la fonction caractèristique :

$$\begin{aligned}\Phi_X(t) &= E[\exp(itX)] = \sum_{k=0}^{+\infty} e^{itk} e^{-\lambda}\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=0}^{+\infty}\frac{(\lambda e^{it})^k}{k!} = \exp[\lambda(\exp(it)-1)].\end{aligned}$$

■

Remarque 1.2.2 *Soient X et Y deux v.as. qui suivent la loi de poisson de paramètres λ et μ respectivements alors leur somme suit aussi une loi de Poisson de paramètre $\lambda+\mu$.*

Loi géométrique $G(p)$

On dit que X suit une loi géométrique de paramètre p, si la probabilité est donnée par

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

En utilisant la série entiére $\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$ pour $|x|<1$,puis en dérivant, on en déduit que $\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$; ce qui permer de vérifier que $\sum_{k=0}^{\infty}\mathbb{P}(X=k) = 1$.

Proposition 1.2.2 *Soit $X \sim G(p)$ alors :*

- $E(X) = 1/p$.
- $Var(X) = q/p^2$.
- $\Phi_X(x) = \dfrac{p\exp(it)}{1-q\exp(it)}$.

Preuve.

$$\begin{aligned}E(X) &= \sum_{k=1}^{\infty} kpq^{k-1} = p/(1-q)^2 \\ &= 1/p.\end{aligned}$$

Le calcule de la variance se fait apartir du moment factoriale :

$$\begin{aligned}E(X(X-1)) &= \sum_{k=1}^{\infty} k(k-1)pq^{k-1} \\ &= pq\sum_{k=2}^{\infty} k(k-1)q^{k-2} \\ &= 2pq/(1-q)^3 = 2q/p^2.\end{aligned}$$

D'ou en déduit :

$$\begin{aligned}Var(X) &= E(X(X-1)) + E(X) - (E(X))^2 \\ &= \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = q/p^2.\end{aligned}$$

Fonction caractéristique :

$$\begin{aligned}\Phi_X(x) &= \frac{p\exp(it)}{1-(1-p)\exp(it)} \\ &= \frac{p\exp(it)}{1-q\exp(it)}.\end{aligned}$$

voir J-P Lecoutre [3] ■

1.2.2 Lois usuelles continue :

De nombreuses lois continues ont étaient élaborées dans la littérature parmi elles :

Loi Uniforme :

soient $a, b \in \mathbb{R}$ avec $a < b$, on dit qu'une variable aléatoire X suit une loi uniforme sur $[a,b]$, $(X \sim U_{[a,b]})$ si sa fonction de densité f_X est donnée par

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

Proposition 1.2.3 *soit $X \sim U_{[a,b]}$, alors*

- $E(X) = \frac{a+b}{2}$.
- $Var(X) = \frac{(b-a)^2}{12}$.
- $\phi_X(t) = E(e^{itX}) = \frac{e^{itb}-e^{ita}}{it(b-a)}$.

Loi Exponentielle :

soit $\lambda > 0$, on dit qu'une v.a. X suit une loi exponentielle de paramètre λ, $(X \sim \varepsilon(\lambda))$ si sa fonction de densité f_X est donnée par :

$$f_X(x) = \lambda e^{-\lambda x} 1_{[0,+\infty]}(x).$$

Proposition 1.2.4 *soit $X \sim \varepsilon(\lambda)$, alors*

- $E(X) = \frac{1}{\lambda}$.
- $Var(X) = \frac{1}{\lambda^2}$.
- $\phi_X(t) = (1-\frac{it}{\lambda})^{-1} = \frac{\lambda}{\lambda - it}$.

Loi Gamma :

la fonction gamma est définie pour $a > 0, \lambda > 0$ par :

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} e^{-x} dx.$$

on a :$\forall n \in \mathbb{N}^*, \Gamma(n) = (n-1)!, \Gamma(1) = 1, \Gamma(\frac{1}{2}) = \sqrt{\pi}$

$$\forall a \in]1, +\infty[\,, \Gamma(a) = (a-1)\Gamma(a-1).$$

On dit qu'une v.a. X suit une loi gamma de paramètre a, λ , $(X \sim \Gamma(\lambda, a))$ si sa fonction de densité f_X est donnée par :

$$f_X(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} 1_{]0,+\infty[}(x).$$

Proposition 1.2.5 *soit* $X \sim \Gamma(\lambda, a)$*,alors*

- $E(X) \quad = \frac{a}{\lambda}$.
- $Var(X) \quad = \frac{a}{\lambda^2}$.
- $\phi_X(t) \quad = (1 - \frac{it}{\lambda})^{-a} = (\frac{\lambda}{\lambda - it})^a$.

Loi Normale :

soient $\mu \in \mathbb{R}, \sigma > 0$,on dit qu'une v.a. X suit une loi normale ou gaussienne de paramètre μ, δ , $(X \sim \mathcal{N}(\mu, \sigma^2))$ si sa fonction de densité f_X est donnée par :

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

Proposition 1.2.6 *soit* $X \sim \mathcal{N}(\mu, \sigma^2)$*,alors*

- $E(X) \quad = \mu$.
- $Var(X) \quad = \sigma^2$.
- $\phi_X(t) \quad = E(e^{itX}) = e^{it\mu - \frac{t^2\sigma^2}{2}}$.

Loi normale centrée rèduite

Définition 1.2.1 *on dit qu'une v.a. X suit une loi de normale centrée rèduite (ou gaussienne centrée rèduite) si sa fonction de densitéf_X est donnèe par*

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

Proposition 1.2.7 *soit $X \sim \mathcal{N}(0,1)$,alors*

- $E(X) = 0.$
- $Var(X) = 1.$
- $\phi_X(t) = e^{-\frac{t^2}{2}}.$

voir F,Boukhari [1]

1.2.3 statistique :

Les statistiques les plus coramment utilisées :

1. La moyenne empirique :
$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$
2. La variance empirique :
$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2.$$
3. La fonction de répartition empirique :
$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}}, x \in \mathbb{R}.$$

1.3 Les types de convergences

On considère une suite de variable alèatoire réelles $\{X_n\ , n \geq 1\}$ définies sur un espace de probabilité (Ω, A, P).

1.3.1 convergence en probabilité :

On dit que la suite de v.a. (X_n) converge en probabilité vers une v.a. X (on noté $X_n \xrightarrow{P} X$) si :

$$\forall \varepsilon > 0, \lim_{n \longrightarrow +\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

où,de façon équivalente :

$$\forall \varepsilon > 0, \lim_{n \longrightarrow +\infty} \mathbb{P}(|X_n - X| \leq \varepsilon) = 1.$$

1.3.2 convergence en loi :

On dit que la suite de v.a. (X_n) de fonction de répartition F_n converge en loi vers une v.a. X de fonction de répartition F si la suite $(F_n(x))$ converge vers $F(x)$ en tout point x où F est continue.

On écrit alors

$$X_n \xrightarrow{\mathcal{L}} X \text{ quand } n \longrightarrow +\infty$$

1.3.3 convergence presque sûre :

On dit que la suite de v.a. (X_n) converge en presque sûrement vers une v.a. X (on noté $X_n \xrightarrow{P.S} X$) si :

$$\forall \varepsilon > 0, \mathbb{P}(\sup\{|X_n - X| < \varepsilon\}) \underset{n \longrightarrow \infty}{\longrightarrow} 1.$$

1.3.4 convergence en moyenne d'ordre p :

On dit que la suite de v.a. (X_n) converge en moyenne d'ordre p (on noté $X_n \xrightarrow{m.p} X$) avec $0 < p < \infty$,vers une v.a. X si :

$$E\,|X_n - X|^P \underset{n \longrightarrow \infty}{\longrightarrow} 0.$$

Remarque 1.3.1 *dans le cas particulier* $p = 2$ *,la convergence en moyenne d'ordre* 2 *s'appelle convergence en moyenne quadratique*$(m.q)$.

1.3.5 La relation entre les types de convergences

La convergence presque sûre implique la convergence en probabilité.

La convergence en probabilité implique la convergence en loi.

La convergence en moyenne implique la convergence en probabilité.

Les implications réciproques sonten général fausses.

$$(X_n \xrightarrow{m.p} X) \Longrightarrow (X_n \xrightarrow{P} X) \Longrightarrow (X_n \xrightarrow{L} X)$$
$$\Uparrow$$
$$(X_n \xrightarrow{P.S} X)$$

1.4 Méthode d'échantillonnage

Il existe deux grands modes d'échantillonnage, le tirage d'un échantillon peut être aléatoire ou non aléatoire.

1.4.1 Méthodes non aléatoire :

L'échantillonnage non aléatoire (non probabiliste) repose sur un choix arbitraire des unités, c'est l'enquêteur qui choisit les unités et non le hasard.

Echantillonnage systématique

Cette méthode exige l'existence d'une liste de la population ou chaque individu est numéroté de 1 jusqu'à N. Notons n le nombre d'individus que doit comporter l'échantillon. L'entier voisin de $\frac{N}{n}$ sera noté r est appelé raison ou pas de sondage. Pour constituer l'échantillon il faut choisir au hasard un entier d entre 1 et r (le point de départ), l'individu dont le numéro correspond à d est le première individu, pour sélectionner les autres, il

suffit d'ajouter à d le pas de sondage, alors les individus choisis seront dont les numéros correspondent à : $d+r, d+2r, d+3r, \ldots, d+(n-1)r$.

L'échantillonnage unité type

La méthode consiste à choisir un individu moyen dite unité type, que l'on déclare représentatif d'un groupe d'individus possédant les mêmes caractéristiques.

L'échantillonnage par quotas

C'est la méthode la plus fréquement utilisée et la plus connu, essentiellement dans les enquêtes, d'opinion et les études de marchés. Elle consiste à partitionner la population suivant un certain nombre de critère (sexe, classe d'âge, ...) cette méthode se base sur l'hypothèse que l'échantillon reproduit fidèlement les caractéristiques sur les quelles va porter l'enquête.

1.4.2 Méthodes aléatoires

L'échantillonnage aléatoire (ou probabiliste) repose sur un choix au hasard d'unités dans la population, i.e, chaque unité de la population à une probabilité mesurable d'etre choisie.

L'échantillonnage aléatoire simple

Cette méthode consiste à choisir dans la population des unités au hasard, tous les unités ont la même probabilité d'être choisie. Ce choix peut se faire avec remise ou sans remise.

-**Tirage avec remise** : chaque unité tirée est examinée puis remise dans la population ; chaque unité peut donc figurer plus d'une fois dans l'échantillon et la composition de la base de sondage ne varie pas au cours de ce processus de tirage.

-**Tirage sans remise** : les unités tirées n'étant pas remises dans la population, chaque unité figure au plus une fois dans la population mais la composition de la base de sondage varie à chaque tirage.(voir Ph,Tassi [4])

1.5 Lois fondamentales d'échantillonnage

On peut définie deux théorèmes fondamentaux sont d'une importance considérable dans l'étude de la probabilité et de la statistique :

- Lois des grands nombres

- Théorème centrale limite

1.5.1 Lois des grands nombres

Soit $\{X_n\ , n \geq 1\}$ une suite de variables aléatoire indépendantes, dans ce paragraphe on s'intéresse au comportement des moyennes arithmétiques

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Lorsque n devient de plus en plus grands. les résultats concernant ce problème sont appelés : **Lois des Grands Nombres** (LGN).

Ces résultats ce décomposent en deux parties :

– **Lois fortes des grands nombres** :

$$\bar{X}_n \xrightarrow{P.S} E(X) \text{ quand } n \to +\infty.$$

– **Lois faibles des grands nombres :**

$$\bar{X}_n \xrightarrow{P} E(X) \text{ quand } n \to +\infty.$$

La différence entre ces deux familles de résultats réside dans le mode de comportement qu'on étuide,dans la première on s'intéresse au comportement ponctuel (convergence presque sûre), alors que dans la seconde on regarde le comportement en probabilité(convergence en probabilité) qui est plus faible.

1.5.2 Théorème centrale limite

Le théorème centrale limite (TCL) est un théorème fondamental, non seulement en calcul des probabilités mais également en statistique mathématique, il affirme gross modo qu'une somme finie de v.as. indépendantes , centrées de même lois et de variance finie, se comporte en loi (lorsqu'elle est bien normalisée et lorsque n devient très grand)comme une v.a. de loi normale centrée réduite. Ainsi on a pas besoin de connaitre la loi des v.a. considérées.

Théorème 1.5.1 *Soit* $\{X_n, n \geq 1\}$ *une suite des v.as. i.i.d , de même loi et de carré intégrable.*

Posons $\mu = E(X_1)$ *et* $\sigma^2 = Var(X_1)$ *, alors :*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1) \text{ lorsque } n \to +\infty. \text{ où : } S = \sum_{i=1}^{n} X_i.$$

Par conséquent

$$\forall x \in \mathbb{R}, \lim_{n \longrightarrow +\infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-t^2}{2}} dt.$$

Preuve. Pour $i \geq 1$, posons $Y_i = \dfrac{X_i - \mu}{\sigma}$.Les variables Y_i ainsi définies sont indépendantes, de même loi, centrées et réduites $i.e$ $E(Y_i) = 0$ et $Var(Y_i) = 1$ .La fonction Caractéristique de Y admet un dévellopement limite de Taylor :

$$\phi_Y(t) = 1 - \frac{t^2}{2} + o(t^2), \ t \to 0.$$

On note :

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}} = \sum_{i=1}^{n} \frac{Y_i}{\sqrt{n}}.$$

Alors,

$$\phi_{\frac{Y_i}{\sqrt{n}}}(t) = E\left(e^{iY_i \frac{t}{\sqrt{n}}}\right) = \phi_{Y_i}\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o(\frac{t^2}{n}).$$

D'après les propriétés des fonctions caractéristiques (la fonction caratéristique d'une somme de v.a. $i.i.d$)

$$\phi_{S_n}(t) = [\phi(t)]^n ,$$

On a

$$\lim_{n\to+\infty} \phi_{Z_n}(t) = \lim_{n\to+\infty} \left[1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right]^n = \lim_{n\to+\infty} \left[1 - \frac{t^2}{2n}\right]^n + o(\frac{t^2}{n}).$$

Sachant que :

$$\lim_{n\to+\infty} \left(1 + \frac{\lambda}{n}\right)^n = e^{\lambda}.$$

Nous obtenons :

$$\lim_{n\to+\infty} \phi_{Z_n}(t) = e^{-\frac{t^2}{2}}.$$

■

Cette limite est la fonction caractéristique de la loi normale centrée réduite $\mathcal{N}(0,1)$d'où l'on déduit le TCL grâce au théorème de continuité de **Lévy** qui affirme que la convergence de fonction caractéristique implique la convergence en loi.(voir G, Saporta [2])

Chapitre 2

Les caractéristiques d'un échantillon

Le problème central de l'inférence statistique est rppelons le suivant : disposant d'observations sur un échantillon de taille n on désire en déduire les propriétés de la population dont il est issu. Ainsi on cherchera à estimer, par exemple **la moyenne** μ de la population à partir de la moyenne $\bar{X}$ d'un échantillon.

2.1 Estimation

Définition 2.1.1 *un estimateur de* θ *est une application de* T_n *de* E^n *dans* F *qui à un échantillon* $(X_1, \ldots, X_n)$ *de la loi* P_θ *associe une v.a.r (ou plusieurs dans le cas d'un paramètre multidimensionnel) dont on peut déterminer la loi de probabilité.*

2.1.1 Biais d'un estimateur

Comme T_n est une v.a. on ne peut pas imposer une condition qu'à sa valeur moyenne, ce qui nous amène à dèfinir le *biais* d'un estimateur comme l'écart entre sa moyenne et la vraie valeur du paramètre :

$$b_\theta(T_n) = E_\theta(T_n) - \theta.$$

D'où une propriété relative au *biais* d'un estimateur.

Définition 2.1.2 *un estimateur T_n de θ est dit* ***sans biais*** *si pour tout θ de Θ et tout entier positif n :*

$$E_\theta(T_n) = \theta.$$

Définition 2.1.3 *un estimateur T_n de θ est dit* ***asymptotiquement sans biais*** *si pour tout θ de Θ :*

$$E_\theta(T_n) \to \theta \text{ quand } n \to +\infty.$$

Exemple 2.1.1 *1. Si le paramètre à estimer est la moyenne théorique, i.e : $\theta = E(X)$, l'estimateur est la moyenne empirique $T_n = \bar{X}_n$.*

$$b_\theta(T_n) = b_\theta\left(\bar{X}_n\right) = E_\theta\left(\bar{X}_n\right) - \theta = \theta - \theta = 0.$$

donc $\bar{X}_n$ la moyenne empirique est un estimateur sans biais de la moyenne théorique.

2. S_n^2 est un estimateur biaisé de σ^2 car $E(S_n^2) = \frac{n-1}{n}\sigma^2$.

le biais est :

$$E(S_n^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-1}{n}\sigma^2.$$

S_n^2 est donc **asymptotiquement sans biais.**

L'inégalité de Bienaymé-Tchebychev :

Soit X une v.a. d'éspérance $E(X)$ et de variance finie $Var(X) = \sigma^2 < \infty$.

L'inégalité de Bienaymé-Tchebychev est :

$$\forall \varepsilon \geq 0, \mathbb{P}(|X - E(X)| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

2.1.2 Convergence d'un estimateur

Un estimateur T_n d'un paramètre θ est dit **convergent** si $E_\theta(T_n)$ tend vers θ lorsque n tend vers l'infini.$\left(E_\theta(T_n) \underset{n\longrightarrow\infty}{\longrightarrow} \theta\right)$.

Il sera dit **consistant** si T_n converge en probabilité vers θ lorsque n tend vers l'infini.

$$T_n \xrightarrow{p} \theta \iff \mathbb{P}\left(|T_n - \theta| < \varepsilon\right) \to 1 \quad \forall \varepsilon > 0, \quad n \to \infty$$
$$\iff \mathbb{P}\left(|T_n - \theta| > \varepsilon\right) \to 0.$$

Théorème 2.1.1 *Si T_n est convergent et de variance tendant vers 0 lorsque n tend vers l'infini alors T_n est* ***consistant.***

Preuve. On a, pour tous réels θ et $\varepsilon > 0$,

$$|T_n - \theta| > \varepsilon \implies |T_n - E(T_n)| > \varepsilon - |\theta - E(T_n)|.$$

Si $\lim_{n\to+\infty} E(T_n) = \theta$, alors à partir d'un certain rang N, on a $|\theta - E(T_n)| < \frac{\varepsilon}{2}$
Ainsi :

$$\begin{aligned}
\mathbb{P}\left(|T_n - \theta| > \varepsilon\right) &\leq \mathbb{P}\left(|T_n - E(T_n)| > \varepsilon - |\theta - E(T_n)|\right) \\
&\leq \mathbb{P}\left(|T_n - E(T_n)| > \varepsilon - \frac{\varepsilon}{2}\right) \\
&\leq \mathbb{P}\left(|T_n - E(T_n)| > \frac{\varepsilon}{2}\right) \\
&\leq \tfrac{4}{\varepsilon^2} Var(T_n) \quad \text{(par Bienaymé-Tchebychev)}
\end{aligned}$$

borne supérieure qui tend vers 0 lorsque $n \to \infty$. ■

2.1.3 Qualité d'un estimateur

La qualité d'un estimateur se mesure également par **l'erreur quadratique moyenne** (ou **risque quadratique**) definie par $E\left[(T_n - \theta)^2\right]$.

Théorème 2.1.2 *Soit T_n un estimateur du paramètre θ à étudier.on a :*

$$E\left[(T_n - \theta)^2\right] = Var(T_n) + \left[E(T_n) - \theta\right]^2.$$

Preuve. On peut écrire

$$T_n - \theta = (T_n - E(T_n)) + (E(T_n) - \theta)$$

$$\begin{aligned} E\left[(T_n-\theta)^2\right] &= E\left[(T_n-E(T_n)+E(T_n)-\theta)^2\right] \\ &= E[(T_n-E(T_n))^2]+E\left[(E(T_n)-\theta)^2\right]+2E\left[(T_n-E(T_n)\left(E(T_n)-\theta\right)\right] \end{aligned}$$

comme $E(T_n)-\theta$ est une constante et que : $E(T_n-E(T_n))=0$.

$$E\left[(T_n-\theta)^2\right]=Var(T_n)+\left[E(T_n)-\theta\right]^2.$$

■

Remarque 2.1.1 *Entre deux estimateurs sans biais, le" meilleur" sera celui dont la variance est minimale (On parle d'efficacité).*

$$Var(T_n)\leq Var(T_m).$$

voir J-P,Lecoutre [3]

2.2 La moyenne empirique

Définition 2.2.1 *la statistique $\bar{X}_n$ ou moyenne empirique d' un échantillon aléatoire $(X_1,\ldots,\ X_n)$ est définie par :*

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i.$$

Propriété 2.2.1 *Soit $\mu=E(X)$ et $\sigma^2=Var(X)$ l'espérance et variance d'une v.a. X ; On a alors*

$$E(\bar{X}_n)\ =\ \mu\ \text{ et }\ Var(\bar{X}_n)\ =\ \tfrac{\sigma^2}{n}.$$

Preuve. $E(\bar{X}_n)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\sum_{i=1}^{n}\mu=\frac{1}{n}n\mu=\mu.$
$Var(\bar{X}_n)=Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}.$

D'aprés l'indépendance des X_i (tirage avec remise d'une population finie) donc :

• La moyenne empirique $\bar{X}_n$ est un estimateur sans biais pour l'espérance mathématique μ.

• Lorsque $n \to +\infty$, $Var(\bar{X}_n) \to 0$,il s'ensuit que $\bar{X}_n$ converge en moyenne quadratique vers μ puisque $E\left[\left(\bar{X}_n - \mu\right)^2\right] \to 0$. ■

Remarque 2.2.1 *Si l'échantillon tiré sans remise d'une population finie, les variables ne sont pas indépendante. Dans ce cas, on a toujours* $E(\bar{X}_n) = \mu$ *;mais On trouve un autre résultat pour* $Var(\bar{X}_n)$:

$$Var(\bar{X}_n) = \left(\tfrac{N-n}{N-1}\right)\tfrac{\sigma^2}{n}.$$

En effet :

$$Var(\bar{X}_n) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\left[\sum_{i=1}^{n} Var(X_i) + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} cov(X_i\ ,\ X_j)\right]$$

avec : $\sum_{i=1}^{n} Var(X_i) = n\sigma^2$ *et*

$$\begin{aligned} cov(X_i\ ,\ X_j) &= E\left[(X_i-\mu)(X_j-\mu)\right] \\ &= \sum_{l=1}^{N}\sum_{k=1}^{N}(x_l-\mu)(x_k-\mu)\,\mathbb{P}\left(X_i = x_l\ ,\ X_j = x_k\right) \\ &= \sum_{l=1}^{N}\sum_{k=1}^{N}(x_l-\mu)(x_k-\mu)\,\mathbb{P}\left(X_i = x_l\right) P\left(X_j = x_k \mid X_i = x_l\right) \\ &= \sum_{l=1}^{N}\sum_{k=1}^{N}(x_l-\mu)(x_k-\mu)\,\frac{1}{N}\mathbb{P}\left(X_j = x_k \mid X_i = x_l\right). \end{aligned}$$

$$cov(X_i\ ,\ X_j) = \begin{cases} \sum_{l=1}^{N}\sum_{k=1}^{N}(x_l-\mu)(x_k-\mu)\,\frac{1}{N}\frac{1}{N-1} & \text{si} \quad k \neq l \\ 0 & \text{si} \quad k = l \end{cases}$$

donc :

$$cov(X_i, X_j) = \frac{1}{N}\frac{1}{N-1}\sum_{\substack{l,k=1 \\ k\neq l}}^{N}(x_l-\mu)(x_k-\mu)$$

comme

$$\left[\sum_{l=1}^{N}(x_l-\mu)\right]^2 = \sum_{l=1}^{N}(x_l-\mu)^2 + \sum_{\substack{l,k=1 \\ k\neq l}}^{N}(x_l-\mu)(x_k-\mu)$$

$$\left[\sum_{l=1}^{N}(x_l-\mu)\right]^2 = (N\mu - N\mu)^2 = 0$$

et $\sum_{l=1}^{N}(x_l-\mu)^2 = N\sigma^2$.

On obtient :

$$\sum_{\substack{l,k=1 \\ k\neq l}}^{N}(x_l-\mu)(x_k-\mu) = -N\sigma^2.$$

et

$$cov(X_i, X_j) = \frac{1}{N}\frac{1}{N-1}(-N\sigma^2) = \frac{-\sigma^2}{N-1}.$$

Donc :

$$\begin{aligned} Var(\bar{X}_n) &= \frac{1}{n^2}\left[n\sigma^2 + \sum_{\substack{i,j=1 \\ i\neq j}}^{N}\frac{-\sigma^2}{N-1}\right] \\ &= \frac{1}{n^2}\left[n\sigma^2 + n(n-1)(\frac{-\sigma^2}{N-1})\right] = \frac{\sigma^2}{n}\left[1-\frac{n-1}{N-1}\right] \\ &= \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right]. \end{aligned}$$

telle que $\frac{N-n}{N-1}$ s'appelle **le facteur d'exaustivité .**

Si $N \to \infty$, avec n fixée, $\frac{N-n}{N-1} \to 1$. Alors il n'y a pas de différence entre entre les deux mode de tirage.

Moment centrée d'ordre k

Le moment centrée de X d'ordre k ($k \in \mathbb{N}^*$) définie parμ_k

$$\mu_k = E\left(X-\mu\right)^k.$$

Alors

$$\begin{aligned}\mu_3(\bar{X}_n) &= E\left(\bar{X}_n-\mu\right)^3 = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right)^3\\ &= \frac{1}{n^3}E\left[\sum_{i=1}^{n}(X_i-\mu)\right]^3 = \frac{1}{n^3}\sum_{i=1}^{n}E\left[X_i-\mu\right]^3\\ &= \frac{\mu_3}{n^2}.\end{aligned}$$

et

$$\begin{aligned}\mu_4(\bar{X}_n) &= E\left(\bar{X}_n-\mu\right)^4 = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right)^4\\ &= \frac{1}{n^4}E\left[\sum_{i=1}^{n}(X_i-\mu)\right]^4 = \frac{1}{n^4}\sum_{i=1}^{n}E(X_i-\mu)^4+\begin{pmatrix}2\\4\end{pmatrix}\frac{1}{n^4}\sum_{i<j}E\left[(X_i-\mu)^2(X_j-\mu)^2\right]\\ &= \frac{\mu_4+3(n-1)\sigma^4}{n^3}.\end{aligned}$$

Théorème 2.2.1 (Comportement asymptotique) *Si $\bar{X}_n$ est la moyenne empirique d'un échantillon $X_1,\ldots,X_n$ tel que $E(X^2)<\infty$*
d'aprés le TCL , $\sqrt{n}\dfrac{\bar{X}_n-\mu}{\sigma} \xrightarrow{L} \mathcal{N}(0,1)$; $(n\to+\infty)$.
De même, la loi des grands nombres, n'implique que $\bar{X}_n \xrightarrow{P} \mu$ et $\bar{X}_n \xrightarrow{P.S} \mu$ quand $n\to+\infty$.

2.3 La variance empirique

Définition 2.3.1 *La statistique S_n^2 ou **variance empirique** d'un échantillon aléatoire $(X_1,\ldots,X_n)$ est définie par :*

$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2.$$

Propriété 2.3.1 *Soit $\mu=E(X)$ et $\sigma^2=Var(X)$ l'ésperance et variance d'une v.a.X ;*

On a alors :

$$\begin{aligned} E(S_n^2) &= \frac{n-1}{n}\sigma^2. \\ Var(S_n^2) &= \frac{n-1}{n^3}\left[(n-1)\,\mu_4 - (n-1)\sigma^4\right]. \end{aligned}$$

Preuve.

1. $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu) - \left(\bar{X}_n - \mu\right)\right]^2$
$= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 + \frac{1}{n}\sum_{i=1}^{n}\left(\bar{X}_n-\mu\right)^2 - \frac{2}{n}\left(\bar{X}_n-\mu\right)\sum_{i=1}^{n}(X_i-\mu).$

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \left(\bar{X}_n-\mu\right)^2 \tag{2.1}$$

$$\begin{aligned} E(S_n^2) &= \frac{1}{n}\sum_{i=1}^{n}E\,(X_i-\mu)^2 - E\left(\bar{X}_n-\mu\right)^2 \\ &= Var(X_i) - Var(\bar{X}_n) \\ &= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2. \end{aligned}$$

2. $\mu_4 = E\,(X-\mu)^4$ le moment centré d'ordre 4 ,

d'aprés (2.1) on a $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - \left(\bar{X}_n-\mu\right)^2$.
on pose $Y_i = (X_i-\mu)^2$, donc :

$$\begin{aligned} Var(S_n^2) &= \frac{1}{n^2}Var\left(\sum_{i=1}^{n}Y_i\right) - \frac{2}{n}\sum_{i=1}^{n}Cov\left(Y_i, \left(\bar{X}_n-\mu\right)^2\right) + Var\left[\left(\bar{X}_n-\mu\right)^2\right] \\ &= \frac{1}{n}Var(Y_1) - 2Cov\left(Y_1, \left(\bar{X}_n-\mu\right)^2\right) + Var\left[\left(\bar{X}_n-\mu\right)^2\right] \\ &= U_n - 2V_n + W_n. \end{aligned}$$

on a d'abord

$$U_n = \frac{1}{n}\left(E\left[(X_1-\mu)^4\right] - E^2\left[(X_1-\mu)^2\right]\right) = \frac{\mu_4-\sigma^4}{n} \tag{2.2}$$

D'autre part

$$\begin{aligned} V_n &= Cov\left[(X_1-\mu)^2, \left(\bar{X}_n-\mu\right)^2\right] \\ &= E\left[(X_1-\mu)^2\left(\bar{X}_n-\mu\right)^2\right] - E\left[(X_1-\mu)^2\right] E\left[\left(\bar{X}_n-\mu\right)^2\right] \\ &= E\left[(X_1-\mu)^2\left(\bar{X}_n-\mu\right)^2\right] - \frac{\sigma^4}{n}. \end{aligned}$$

avec

$$\begin{aligned} E\left[(X_1-\mu)^2\left(\bar{X}_n-\mu\right)^2\right] &= \frac{1}{n^2}\left(\sum_{i=1}^{n} E\left[(X_1-\mu)^2(X_i-\mu)^2\right] + \sum_{j\neq k} E\left[(X_1-\mu)^2(X_j-\mu)(X_k-\mu)\right]\right) \\ &= \frac{1}{n^2}\left(E\left[(X_1-\mu)^4\right] + \sum_{i=2}^{n} E\left[(X_1-\mu)^2(X_i-\mu)^2\right] + 0\right) \\ &= \frac{\mu_4+(n-1)\sigma^4}{n^2}. \end{aligned}$$

D'où

$$V_n = \frac{\mu_4+(n-1)\sigma^4}{n^2} - \frac{\sigma^4}{n} = \frac{\mu_4-\sigma^4}{n^2}. \tag{2.3}$$

Enfin

$$W_n = Var\left[\left(\bar{X}_n-\mu\right)^2\right] = E\left[\left(\bar{X}_n-\mu\right)^4\right] - E^2\left[\left(\bar{X}_n-\mu\right)^2\right] = E\left[\left(\bar{X}_n-\mu\right)^4\right] - \frac{\sigma^4}{n^2}.$$

et on a

$$E\left[\left(\bar{X}_n-\mu\right)^4\right] = \frac{\mu_4-3\sigma^4}{n^3} + \frac{3\sigma^4}{n^2}.$$

Alors

$$W_n = \frac{\mu_4-3\sigma^4}{n^3} + \frac{2\sigma^4}{n^2}. \tag{2.4}$$

Finalement , d'aprés (2.2)(2.3) et (2.4) on obtient

$$\begin{aligned} Var(S_n^2) &= \frac{\mu_4-\sigma^4}{n} - 2\frac{\mu_4-\sigma^4}{n^2} + \frac{\mu_4-3\sigma^4}{n^3} + \frac{2\sigma^4}{n^2} \\ &= \frac{\mu_4-\sigma^4}{n} - \frac{2(\mu_4-2\sigma^4)}{n^2} + \frac{\mu_4-3\sigma^4}{n^3} \\ Var(S_n^2) &= \frac{n-1}{n^3}\left[(n-1)\mu_4 - (n-3)\sigma^4\right]. \end{aligned}$$

et si $n \to \infty \quad Var(S_n^2) \approx \dfrac{\mu_4 - \sigma^4}{n}$

la variance S_n^2 étant biaisée et ayant donc tendance à sous-estimer σ^2 , on utilise fréquemment **la varaince corrigée** dont l'espérance vaut exactement σ^2 :

$$\begin{aligned} S_n^{2*} &= \frac{n}{n-1} S_n^2 &&= \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^2. \\ E(S_n^{2*}) &= \frac{n}{n-1} E(S_n^2) &&= \sigma^2. \end{aligned}$$

■

Théorème 2.3.1 (Comportement asymptotique) *Soit* $X_1, \ldots, X_n \sim X$ *telle que* $E\left(X^2\right) < \infty$, *alors on a*

$$S_n^2 \xrightarrow{P.S} \sigma^2 \quad , \quad S_n^{2*} \xrightarrow{P.S} \sigma^2.$$

Preuve. On a $S_n^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2$

donc : $\bar{X}_n \xrightarrow{P.S} E(X) \implies \bar{X}_n^2 \xrightarrow{P.S} E^2(X)$.

et $\overline{X_n^2} - \bar{X}_n^2 \xrightarrow{P.S} E(X^2) - E^2(X)$

D'ou $S_n^2 \xrightarrow{P.S} \sigma^2$.

De même façon $S_n^{2*} \xrightarrow{P.S} \sigma^2$. ■

Théorème 2.3.2 *Soit* $(X_1, ..., X_n) \sim X$ *telle que* $(E(X^4) < \infty$, *alors ona :*

$$\frac{\sqrt{n}\left(S_n^2 - \sigma^2\right)}{\sqrt{\mu_4 - \sigma^4}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1), \text{ quand } n \to \infty$$

Preuve. Daprés la décomposition (2.1) de $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 - \left(\bar{X}_n - \mu\right)^2$ ona :

$$\begin{aligned} \sqrt{n}\left(S_n^2 - \sigma^2\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 - \left(\bar{X}_n - \mu\right)^2 - \sigma^2\right) \\ &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)^2 - \sigma^2\right) - \sqrt{n}\left(\bar{X}_n - \mu\right)^2 \\ &= Z_n - U_n \end{aligned}$$

On commence par Z_n ,d'aprés le TCL aux vecteur $(X_i-\mu)^2$ $i.i.d$, d'espérance σ^2 et de variance :

$$Var\left(X_i-\mu\right)^2=E\left(X_i-\mu\right)^4-\left[E\left(X_i-\mu\right)^2\right]^2=\mu_4-\sigma^4$$

On obtient donc

$$\frac{Z_n}{\sqrt{\mu_4-\sigma^4}}\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,1\right).$$

Il reste à preuve que U_n est négligeable. D'aprés le TCL, LGN et le théorème de Slutsky :

$$\bar{X}_n\xrightarrow{P}\mu \text{ et } \sqrt{n}\left(\bar{X}_n-\mu\right)^2\xrightarrow{P}0.$$

Donc $U_n\xrightarrow{P}0$, d'ou le résultat. ■

Remarque 2.3.1 *De méme façons pour* S_n^2

$$\sqrt{n}\left(S_n^{2*}-\sigma^2\right)\xrightarrow{l}\mathcal{N}\left(0,\mu_4-\sigma^4\right), \text{ quand } n\to\infty$$

2.3.1 Corrélation entre la moyenne et la variance empirique

Théorème 2.3.3 *Soit* X *(centrée) telque* $E\left(X^3\right)<\infty$ *et* $E(X)=0$ *, alors*

$$Cov(\bar{X}_n,S_n^2)=\frac{n-1}{n^2}\mu_3.$$

Preuve. On a

$$\begin{aligned}
Cov(\bar{X}_n,S_n^2) &= E\left(\bar{X}_nS_n^2\right)-E\left(\bar{X}_n\right)E\left(S_n^2\right)=E(\bar{X}_n,S_n^2)\\
&= E\left[\bar{X}_n\left(\overline{X_n^2}-\bar{X}_n^2\right)\right]=E\left[\bar{X}_n\overline{X_n^2}\right]-E\left[\left(\bar{X}_n\right)^3\right]\\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)\left(\frac{1}{n}\sum_{j=1}^{n}X_j^2\right)\right]-\frac{\mu_3}{n^2}\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}E\left(X_iX_j^2\right)-\frac{\mu_3}{n^2}
\end{aligned}$$

et

$$E\left(X_iX_j^2\right) = \begin{cases} E\left(X_i^3\right) = \mu_3 & \text{si} \quad i = j \\ E\left(X_i\right)E(X_j^2\big) = 0 & \text{si} \quad i \neq j \end{cases}$$

$$Cov(\bar{X}_n, S_n^2) = \frac{1}{n^2}\sum_{i=1}^{n}\mu_3 - \frac{\mu_3}{n^2} = \frac{\mu_3}{n} - \frac{\mu_3}{n^2}$$

$$Cov(\bar{X}_n, S_n^2) = \frac{n-1}{n^2}\mu_3. \ \blacksquare$$

2.4 Fonction de répartition empirique

Définition 2.4.1 *la fonction de répartition empirique d'un échantillon* $(X_1, \ldots, X_n)$ *est définie par*

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i \leq x\}} \ , \ x \in \mathbb{R}.$$

posons $Y_i = \mathbf{1}_{\{X_i \leq x\}}$, $\forall x \in \mathbb{R}$, *on obtient* $F_n(x) = \bar{Y}_n$ *et* $Y = \mathbf{1}_{\{X \leq x\}}$ *est une v.a. valant soit* 0 *et* 1 , $Y \sim \mathcal{B}(p)$ *avec* $p = \mathbb{P}(Y = 1) = E(Y)$, *d"où*

$$E(Y) = E(\mathbf{1}_{\{X \leq x\}}) = \mathbb{P}(X \leq x) = F(x)$$

on déduit que :

$$nF_n(x) \sim \mathcal{B}(n, F(x))$$

De plus,

$$\begin{array}{lllllll} E\left(F_n(x)\right) & = & E(\bar{Y}_n) & = & E(Y) & = & F(x). \\ Var\left(F_n(x)\right) & = & Var(\bar{Y}_n) & = & \frac{Var(Y)}{n} & = & \dfrac{F(x)(1-F(x))}{n}. \end{array}$$

Alors $F_n(x)$ *est un estimateur sans biais de* $F(x)$.

Théorème 2.4.1 (Comportement asymptotiuque) *la fonction de distribution empirique* $F_n(x)$ *converge p.s vers la distribution* $F(x)$:

$$F_n(x) \xrightarrow{P.S} F(x) \ , \ quand \ n \to \infty, \forall x \in \mathbb{R}.$$

Preuve. D'aprés la loi forte des grands nombres , on a

$$F_n(x) \xrightarrow{P.S} E\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}_{\{X_i\leq x\}}\right) = E\left(\mathbf{1}_{\{X_i\leq x\}}\right) = \mathbb{P}(X \leq x) = F(x).$$

■

Chapitre 3

Application sous R

Dans ce chapitre, nous allons essayer d'appliquer ce qu'on a déja étudié dans les chapitres précédents et ceci sur des exemples concrets, pour cela on va donc utiliser le logiciel statistique **R** pour simuler l'étude de l'échantillonnage.

3.1 La moyenne empirique

Nous avons étudiée le comportement asymptotique de la moyenne empirique, prenant un échantillon de la loi exponentielle de paramétre 1 de taille $n = 1000$. La figure (3.1) représente la convergence de la moyenne empirique, dans le graphe à gauche nous présentons la moyenne empirique en fonction de la taille de l'échantillon et au millieu figure l'histograme qui confirme les résultats obtenus, et dans le graphe à droite nous avons comparé la disribution de $\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}$ avec la distribution normale.

Programme en code R :

```
par(mfrow=c(1,3))
n<-1000
X<-rexp(n)
Y<-cumsum(X)
N<-seq(1,n, by=1)
```

plot(N, Y/N,xlab="Taille d'échantillon", ylab="Moyenne empirique",main=" La M.e vs M.t")

M.e :La moyenne empirique, M.t : Moyenne théorique.

abline(h=1,col="red")

text(600, 1.5,expression(mu==0.5),clo=2)

n=1000 ;r=800 ;M=numeric(r) ;m=numeric(r)

for(i in 1 :r){

x=rexp(n) ;m[i]=mean(x)

M[i]=sqrt(n)*(m[i]-1)}

hist(M,col=3)

qqnorm(M,col=3)

qqline(M,lwd='2')

Représentation graphique :

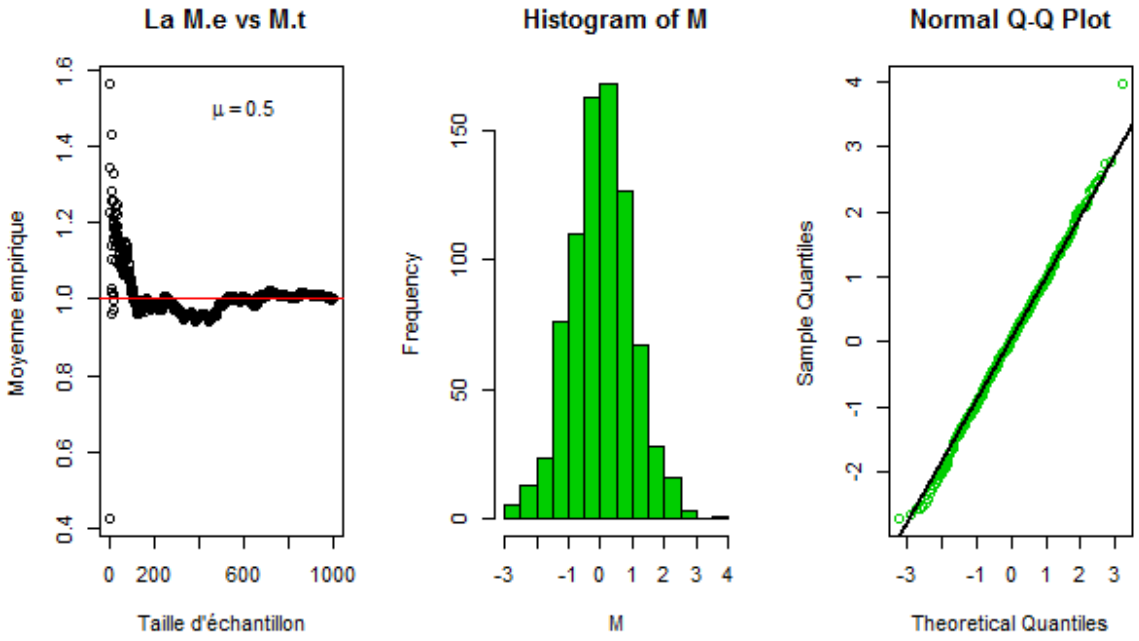

FIG. 3.1 – La convergence de la moyenne empirique.

On remarque dans le graphe à gauche que pour n assez grand la moyenne empirique

converge vers la moyenne théorique ($\mu = 0.5$) réprésentée par la droite horizontale rouge. Ce rèsultat est confirmé par l'histograme du millieu. Enfin à droite, on remarque que les points sont bien alignés, ceci confirme que la distribution d'échantillonnage des moyennes s' approche d'une distribution normale quand n est assez grand.

3.2 La variance empirique

Pour étudier le comportement asymptotique de la variance empirique, prenant un échantillon de la loi exponentielle de paramétre 2 de taille $n = 1000$. La figure (3.2) représente la convergence de la variance empirique, dans le graphe à gauche nous avons tracé la variance empirique en fonction de la taille d'échantillon et au millieu on a mit un histograme qui confirme les résultats obtenus, et dans le graphe à droite nous avons comparé la disribution de $\sqrt{n}\frac{(S_n^2-\sigma^2)}{\sqrt{\mu_4-\sigma^4}}$ avec la distribution normale.

Programme en code R :

```
par(mfrow=c(1,3))
n=1000 ;d=1 :n ;S=numeric(n)
for(i in 1 :n){
x=rexp(i,2) ;S[i]=var(x)}
plot(d,S,xlab="Taille d'échantillon",ylab="Variance empirique",main=" V.e vs V.t ",col='4')
# V.e : La Variance empirique, V.t :La Variance théorique.
abline(h=1/4,lwd=3,col="red")
text(800, 0.45,expression(sigma^2==0.25),col=3)
n=10000 ;r=900
s=numeric(r) ;v=numeric(r)
for(i in 1 :r){
z<-rexp(n,2) ;v[i]<-var(z)
s[i]<-sqrt(n)*((v[i]-(1/4))/sqrt(0.5))
```

}

hist(s,col=4)

qqnorm(s,col=3)

qqline(s,lwd='2')

Représentation graphique :

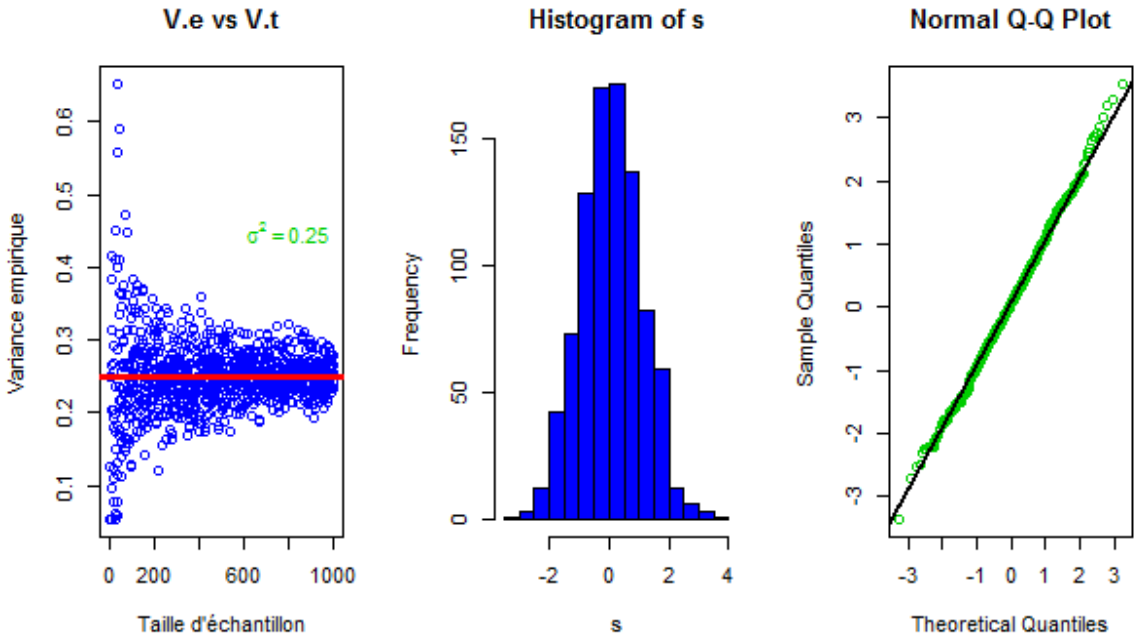

FIG. 3.2 – La convergence de la variance empirique.

On observe dans le graphe à gauche que si n est assez grand la variance empirique converge vers la variance théorique ($\sigma^2 = 0.25$) réprésentée par une droite horizontale rouge. On confirme par la suite ces rèsultats grace à histograme du millieu. Enfin à droite,on remarque que les points sont bien alignés, ceci confirme que la distribution d'échantillonnage des variances est proche d'une distribution normale quand n est assez grand.

Conclusion

En conclusion pour un chercheur la selection de l'echantillon obeit à certains criteres specefiques La sélectionne de cet échantillon se fait aussi selon certaines méthodes appropriées aux expériences scientifiques Ce choix dépend aussi du but de la recherche, et de l'identification du milieu de recherche pour obtenir de meilleures resultats.

Dans notre travail, nous avons étudié la théorie de l'échantillonnage les caractéristiques des distributions statistiques de l'échantillon. Dans le premier et le deuxième chapitres, nous avons présentés les méthodes et les lois, des distributions d'échantillonnage ainsi que leurs caractéristiques, en particulier des caractéristiques de comportement asymptotiques des statistiques mathématiques.Le troisième chapitre est un processus de simulation numérique pour vérifier et confirmer les résultats théoriques obtenus dans les deux chapitres précédents.

Bibliographie

[1] F. Boukhari. (2017). Probabilités. Université Abou Bekr Belkaid Tlemcen.

[2] G. Saporta. (2006). Probabilités,Analyse des Données et Statistique. $2^{éme}$ édition.Edition Technip,Paris.

[3] J-P. Lecoutre. (2009). Statistique et Probabilités. $4^{éme}$ édition.Dunod,paris.

[4] Ph. Tassi. (1989). Methodes Statistiques. $2^{éme}$ édition.Economica,Paris.

[5] R. Clairin, P. Brion. (1997). Manuel de Sondages - Application aux pays en développement. $2^{éme}$édition.-EHESS-INED-INSEE-ORSTOM-Université Paris VI.

[6] R. Ihaka,.R. Gentleman. (1996) *R : A Language for Data Analysis and Graphics*. Journal of Computational and Graphical Statistics **5** : $299 - 314$.

Annexe A : Logiciel *R*

R est systéme, communément appelé langage et logiciel qui permet de réaliser des analyses statistique, plus particuliérement , il comporte des moyens qui rendent possible la manipulation des données , les calculs et les représentations graphiques, R a aussi la possibilité d'exécuter des programmes stockés dans des fichiers textes et comporte un grand nombre de procédures statistiques appelées paquets, Ces derniers permettent de traiter assez rapidement des sujets aussi varariés que les modéles linéaires(simples et généralisés), la régression(linéairel et non linéaire), les séries chronologique, les tests paramétriques et non paramétriques classiques, les différentes méthodes d'analyse des donnés,...plusieurs paquets, tels *ade4, FactoMineR, MASS, multiVarariate, scatterplot3d et rgl* entre autres sont destinés à l'analyse des donnés statistique multidimensionnelles.

Il a été initialement créé, en 1996 par Robert Gentleman et Ross Ihaka du département de statistique de l'Université d'Auckland en NouVarelle Zélande. Depuis 1997 il s'est forme une équipe "*R Core Team*" que dévareloppe *R*.Il est conçu pour pouvarior étre utilisé a Varec les systémes d'expoloitation *Unix, Lunix, Windows et MacOS.*

Un élément clé dans la missionde dévareloppement de *R* est le *ComprehensiVare R Archi-Vare Network* (CORAN) qui est un ensemble de sites qui fournit tout ce qui est nécessaire à la distribution de *R*, ses extensions, sa documentation, ses fichiers sources et ses fichiers binaires. Le site maître du CORAN est situé Autriche à Varinne, on peut y accéder par l'URL :"*http ://cran.r-project.org/".* Les autres site de CORAN, appelés site miroirs, sont répandus partout dans le monde.

R est un logiciel libre distribué sous les termes de la "GNU Public Licence" Il fait partie intégrante du projet GNU et posséde un site officiel à liadresse "*http ://www.R-project.org*". Il est souvarent présenté comme un clone de S qui est un langage de haut niVareau, dévareloppé par les *ATUT Bell Laboratories* et plus particuliérement par *Rick Becker, John Chambers* et *Allan Wilks*. S est utilisable à traVarers le logiciel *S-plus* qui est commercialisé par la société Insightful (*http ://www.splus.com/*).

Annexe B : Abréviations et Notations

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

E	population.
ξ_n	échantillon de E.
e_i	unité de la population E.
e_{i_k}	unité de l'échantillon ξ_n.
$\mathcal{B}(p)$	loi de bernoulli.
$\mathcal{B}(n,p)$	loi de binomiale.
$\mathcal{P}(\lambda)$	loi de poisson.
$G(p)$	loi de géométrique.
$U_{[a,b]}$	loi uniforme.
$\varepsilon(\lambda)$	loi exponentielle.
$\Gamma(\lambda,a)$	loi gamma.
$\mathcal{N}(\mu,\sigma)$	loi normale d'espérence μ et d'écart type σ.
$(X_1,...,X_n)$	échantillon de taille n de v.a's.
$\xrightarrow{\mathcal{P}}$	convergence en probabilité.
$\xrightarrow{\mathcal{L}}, \xrightarrow{\mathcal{D}}$	convergence en loi, convergence en distribution.

$\xrightarrow{p.s.}$	convergence presque sûre.
$\xrightarrow{m.p}$	convergence en moyenne d'ordre p.
(Ω, A, P)	espace de probabilité.
LGN	lois des grands nombres.
TCL	théorème centrale limite.
$b_\theta(T_n)$	biais de l'estimateur T_n.
Θ	ensemble des paramétres.
$v.a$	variable aléatoire.
$i.i.d$	indépendantes et identiquement distribuées.
$E(.), \mu$	espérance mathématique(moyenne).
$Var(.), \sigma^2$	variance mathématique.
$\bar{X}_n$	moyenne empirique.
S_n^2	variance empirique.
S_n^{2*}	variance empirique corrigée.
$Cov(\bar{X}_n, S_n^2)$	covariance de $\bar{X}_n$ et S_n^2
$F(x)$	fonction de répartition
$F_n(x)$	fonction de répartition empirique.
Φ_X	fonction caractéristique de X.
f_X	fonction de densité.
Ω	ensemble fondamental.
μ_k	moment centré d'ordre k.
$\sim$	équivalance en loi de probabilité.
$1_{\{X_i \leq x\}}$	fonction indicatrice de l'ensemble $\{X_i \leq x\}$.