République Algériennne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed Khider, Biskra

Faculté des Sciences Exactes

Départementde Mathématiques

Mémoire présenté pour obtenir le diplôme de

Master en "**Mathematiques Appliquées**"

Option : Statistique

Par Benbouzid Yassine

Titre :

Normalité Asymptotique d'une Suite de Variables Aléatoires Indépendantes

Devant le Jury :

Mr	Benkhelifa Lazhar	Professeur	U. Biskra	Président
Mme	Dhiabi Samra	M.C.B	U. Biskra	Encadrant
Mme	Khemissi Zahia	M.C.B	U. Biskra	Examinatrice

Soutenu Publiquement le 03/06/2025

𝔇édicace

Je dédie ce travail à :

À mes très chers parents.

À mes très chers frères et soeurs.

À tous mes proches.

À tous mes amis.

À mes amis et mes collègues de la promotion 2025

‹‹Mathématiques ››.

Remerciements

Je veux dire que j'ai réussi à faire ce modeste travail grâce à Dieu, qui m'a donné la force, la santé, la volonté et le courage d'atteindre cette limite. À **mon cher père** et à **ma chère mère** pour leur patience et leur sacrifice afin d'atteindre cette étape avancée de l'étude et toute ma famille. Je voudrais remercier en particulier ma superviseure **Samra Dhiabi**, qui m'a honoré de sa supervision, de sa direction, de sa direction, de son humilité, de sa patience, de ses conseils et de toutes ses remarques constructives pour le bon déroulement de mon travail.

Je tiens également à exprimer mes remerciements à tous les amis du département de mathématiques et à l'étranger, et à tous ceux qui ont laissé leurs empreintes digitales loin ou à proximité sur invitation ou des conseils pour faire ce travail modeste travail. Je remercie beaucoup **Dr Ben atia Fatah,** qui a toujours été mon mentor et mon assistant dans ma condition la plus faible, une

partie de tout bien, et je lui souhaite beaucoup de succès avec la permission de Dieu. Et je remercie tous mes amis et mes proches qui m'ont encouragé tout au long de ma carrière scolaire. Merci du fond du coeur.

Notations et symbols

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

$v.\ a$	:	Variable aléatoire.
$(\Omega,\ \mathcal{F},\ P)$	:	Espace de probabilité.
X	:	Variable aléatoire réelles.
Ω	:	Un univers des possibles.
$\mathcal{F}$	:	Tribu.
$F_X(y)$	:	Fonction de répartition
A	:	événement
$i.\ i.\ d$	:	Indépendant et identiquement distribué.
$E(X)$	:	Espérance mathématique.
$V(X)$	:	Variance mathématique.
$Cov(X,\ Y)$	:	Covariance de corrélation.
σ	:	écart-type.
$f(X)$	:	Fonction de densité.
$\varphi_X(t)$	:	La fonction caractéristique X.
$B(n,\ p)$	:	La loi Binomial de taille et de paramètre p.

$P(\lambda)$	:	La loi de Poissan de paramètre λ.
$N(\mu,\ \sigma)$	:	La loi normale(ou La place Gauss) de paramètre μ et σ.
$N(0,\ 1)$	:	La loi normale centrée et réduite.
$(X_1, \ldots, X_n)$	:	échantillon de taille n de v. a.
$\xrightarrow{\mathcal{P}}$	:	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 dordre p.
TCL	:	Théorème Central Limite.

Table des matières

Table des figures

Introduction

Le sujet de la normalité asymptotique d'une suite de variables aléatoires indépendantes constitue l'un des axes fondamentaux de la théorie des probabilités, en raison de son rôle crucial dans l'explication du comportement des phénomènes aléatoires lorsque la taille de l'échantillon devient grande. Ce concept est étroitement lié au **Théorème Central Limite (TCL)**, qui affirme que la somme de variables aléatoires indépendantes et identiquement distribuées, une fois recentrée et normalisée, tend à suivre une loi normale lorsque le nombre de ces variables devient très élevé. Ce domaine a connu une évolution historique importante depuis le siècle avec les travaux **d'Abraham de Moivre** et **Laplace**, avant que sa formulation et ses conditions d'application ne soient précisées par des mathématiciens tels que Tchebychev, Markov, Lyapunov et Lindeberg. Partant de l'importance de ces résultats tant sur le plan théorique qu'applicatif, cette étude soulève la problématique suivante : **quelles sont les conditions nécessaires et suffisantes qui garantissent la convergence vers la loi normale pour une suite de variables aléatoires indépendantes ?** Ainsi, le principal objectif de ce travail est de présenter une étude rigoureuse des différentes formes du théorème central limite, d'exposer les conditions qui assurent la normalité asymptotique, et d'illustrer ces résultats à travers des applications classiques sur des lois connues comme la loi binomiale et la loi de poisson.

Chapitre 1. Ce travail débute par un rappel des concepts fondamentaux de la théorie des probabilités et des variables aléatoires. Nous y abordons la notion de loi de distribution ainsi que la fonction de répartition associée, illustrées par des exemples de lois discrètes et continues. Les principales propriétés des variables aléatoires, telles que l'espérance mathématique, la variance et les moments, y sont également présentées. Ensuite, l'étude se poursuit avec l'analyse des différents types de convergence des variables aléatoires, notamment la convergence en probabilité et la convergence en loi, avant de s'intéresser aux lois des grands nombres, à travers leurs formulations faibles et fortes.

Chapitre 2. Ce chapitre a pour objectif d'étudier le théorème central limite dans le cadre des variables aléatoires indépendantes et identiquement distribuées, en abordant également l'approximation des lois binomiale et de Poisson par la loi normale. Il traite ensuite des bases de l'estimation statistique, à travers l'estimation ponctuelle et l'étude des propriétés des estimateurs, en mettant l'accent sur les méthodes des moments et du maximum de vraisemblance, avant de présenter la construction des intervalles de confiance pour l'estimation des paramètres, tant dans le cas où l'écart-type est connu que lorsqu'il est inconnu, jusqu'à l'estimation de la variance.

Chapitre 3. Dans ce chapitre, nous présenterons différentes méthodes graphiques (histogramme , qqplot , etc.) ainsi que des tests statistiques de normalité (comme le test de Shapiro-Wilk, etc.). À l'aide du logiciel pour la simulation, nous appliquerons ces méthodes graphiques et tests de normalité. Nous illustrerons également certaines applications de l'approximation gaussienne pour quelques lois.

Chapitre 1

Fondements théoriques

Dans ce chapitre, nous introduirons quelques concepts et propriétés des variables aléatoires (espérance, variance, fonction de répartition), puis nous étudierons les lois de base qui régissent le comportement des variables aléatoires, nous présenterons ensuite les modes de convergence. Enfin nous décrivons les définitions de la loi des grands nombres.

1.1 Notions de base en probabilités

1.1.1 Espace de probabilité

Expérience aléatoire et événement

Une expérience aléatoire est une expérience dont l'issue ne peut pas être prédite à l'avance, même si elle est répétée dans les mêmes conditions. Elle peut produire différents résultats appelés issues.

Un événement aléatoire est un sous-ensemble de l'ensemble des issues possibles. On dit qu'un événement s'est réalisé si l'issue observée appartient à ce sous ensemble.

Langage des évènement	*Langage des ensembles*	*Notations*	*Exemples avec le Jet d'un dé*
A est un évènement	*A est une partie de E*	$A \subset E$	*A obtenir un Nombre pair* $A = \{2,\ 4,\ 6\}$
C est l'évènement ‹‹A ou B››	*C est la réunion de ‹‹A et B››*	$C = A \cup B$	*A obtenir* $5, B$: *obtenir* $2,\ 4,\ 5$ *ou* 6, C : *obtenir* $\{2,\ 4,\ 5,\ 6\}$
E est l'évènement ‹‹A et D››	*E est l'inter* sec *tion de A et D*	$E = A \cap D$	D : *btenir multiple de* $3, E = \{6\}$
A et F sont des évènements complémentaires	*A et Fsont complémentaires*	$F = \bar{A}$	A : *obtenir un nombre pair.* F : *obtenir un nombr impair.*

1.1.2 Univers des possibles

On représente le résultat de cette expérience comme un élément ω de l'ensemble Ω de tous les résultats possibles. On appelle l'ensemble Ω l'espace des éventualités, ou l'univers des possibles.

Un événement aléatoire (ou plus simplement un événement) est une assertion ou proposition logique relative au résultat de l'expérience (par exemple, la somme des points est paire).

On dira qu'un événement est réalisé ou non suivant que la proposition est vraie ou fausse une fois l'expérience accomplie. Donc, à un événement, on peut associer la partie de Ω constituée de tous les résultats réalisant l'événement.

1.1.3 Variables aléatoires

Définition 1.1.1 *Soit* $(\Omega,\ \mathcal{F},\ \mathbb{P})$ *un espace probabilisé. Une variable aléatoire réelle (v.a.r) est une fonction mesurable* $X : \Omega \to \mathbb{R}$*, telle que pour tout intervalle* $I \subset \mathbb{R}$*, l'image réciproque* $X^{-1}(I)$ *appartient à* $\mathcal{F}$ *.*

• Nous distinguons deux types de variable aléatoires : discrètes et continues.

Variables aléatoires discrète

Si une variable aléatoires X prend un nombre de valeurs fini ou dénombrable (son ensemble de définition est inclus dans $\mathbb{N}$), on parle de variable discrète.

On s'intéresse à définir l'ensemble des valeurs possibles et leurs probabilités associées.

Variables aléatoires continues

Une variable aléatoire est dite continue si elle peut prendre toutes les valeurs d'un intervalle de nombres réels. Autrement dit, elle prend ses valeurs dans un sous-ensemble de $\mathbb{R}$ (comme un intervalle).

Dans ce cas, la probabilité qu'elle prenne une valeur exacte est nulle. On s'intéresse donc à la probabilité qu'elle appartienne à un intervalle donné plutôt qu'à une valeur précise.

1.1.4 Loi de probabilité

Soit $(\Omega, \mathcal{F}, \ P)$ un espace probabilisé et X une variable aléatoire réelle.

On appelle loi de probabilité de X la probabilité, notée P_X , image de par X :

$$P_X\left(\left]a,\ b\right]\right) = P\left(X^{-1}\left(\left]a,\ b\right]\right)\right).$$

Variables indépendantes et identiquement distribuée

Définition 1.1.2 *Deux variables aléatoire X et Y sont indépendantes si pour tout couple $(I,\ J)$ d'intervalles de $\mathbb{R}$, on a :*

$$P\left((X \in I) \cap (Y \in J)\right) = P\left(X \in I\right) P\left(X \in J\right).$$

Définition 1.1.3 *On dira que deux éléments aléatoires X et Y ont la même loi, ou sont identiquement distribuées, si leurs loi P_X et P_Y sont des probabilités identiques. Cette relation sera notée*

$$X \rightsquigarrow Y,$$

L'élément aléatoire X est entièrement déterminé par sa loi P_X.

Définition 1.1.4 *En théorie des probabilités et en statistique, on dit que des variables aléatoires sont indépendantes et identiquement distribuées (notées i.i.d) lorsqu'elles sont toutes régies par la même loi de probabilité et qu'elles sont statistiquement indépendantes les unes des autres. Cela signifie que la connaissance de la réalisation de l'une de ces variables ne donne aucune information sur les autres, et que toutes suivent exactement la même distribution.*

1.1.5 Fonctions de répartition

Définition 1.1.5 *Soit* $(\Omega, \ \mathcal{F}, \ P)$ *un espace probabilisé et* X *une variable aléatoire réelle.*

On appelle fonction de répartition de X *la fonction qui définie sur* $\mathbb{R}$ *par :*

$$F_X : \begin{cases} \boldsymbol{R} & \rightarrow [0,1] \\ y & \rightarrow P_X\left(]-\infty, y]\right), \end{cases}$$

avec,

$$F_X\left(\mathrm{y}\right) = P\left(X \leq \mathrm{y}\right)$$

1. F est continue à gauche. la fonction F est croissante et continue à droite.
2. F est continue à gauche.
3. $\lim\limits_{x \to +\infty} F\left(x\right) = 1$ et $\lim\limits_{x \to -\infty} F = 0$.
4. Une fonction de répartition caractérise la loi.
5. $P\left(a \leq X \leq b\right) = F_X\left(b\right) - F_X\left(a\right)$.

Un exemple de fonction de répartition associée à une variable discrète est donné par la figure suivante1.1

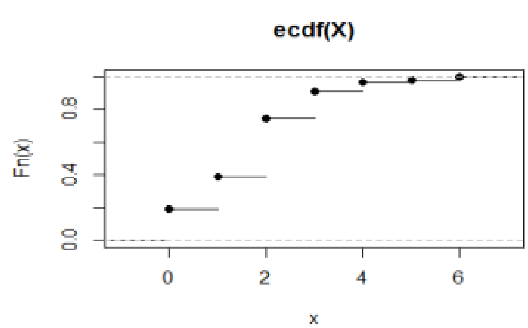

Fig. 1.1 – Fonction de répartition à une v.a.discrète

Un exemple de fonction de répartition associée à une variable continue est donné par la figure 1.2suivante

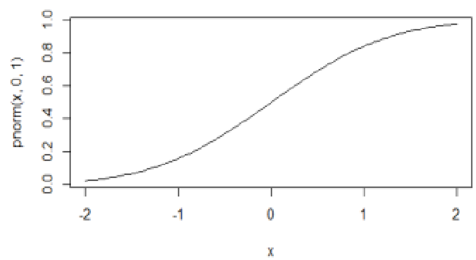

FIG. 1.2 – Fonction de répartition à une v.a.continue

1.1.6 Caractéristiques d'une variable aléatoire

Espérance

Définition 1.1.6 *Soit X une variable aléatoire réelle, l'espérance mathématique de X est définie par :*

- *Si X est une v.a.r discrète finie ou dénombrable :*

$$\mathbb{E}(X) = \sum_{x \in X(\Omega)} xP(X = x).$$

- *Si X est une v.a.r continue à densité f_X :*

$$\mathbb{E}(X) = \int_{\mathbb{R}} t f_X(t)\, dt.$$

Les propriétés de l'espérance valent aussi définies bien pour une variable aléatoire discrète ou une variable aléatoire absolument continue.

1. **Linéarité :** Pour deux variables aléatoires X et Y et deux constantes a et b, on a :
$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$
2. **Espérance d'une constante :** Si c est une constante, alors $\mathbb{E}\,(c) = c$.
3. **Indépendance :** Si X et Y sont indépendantes, alors $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Variance

Définition 1.1.7 *La variance d'une variable aléatoire $V(X)$ est effectivement définie comme l'espérance mathématique du carré de l'écart à l'espérance mathématique. Mathématiquement, cela s'exprime comme suit*

$$V(X) = E\left[(X - E\,[X])^2\right].$$

Où :

- $\mathbb{E}[X]$ *est l'espérance mathématique de X, c'est-à-dire la moyenne théorique de la variable aléatoire.*
- $\mathbb{E}(X - \mathbb{E}[X])^2$ *représente le carré de l'écart entre la variable aléatoire et son espérance.*

Propriété 1.1.1

1. **Non-négativité** : $V(X) \geq 0$.
2. **Variance d'une constante** : Si $X = c$ (une constante), alors $V(X) = 0$.
3. **Linéarité** : Pour toute constante a et b, $V(aX + b) = a^2 V(X)$.

L'écart-type

Définition 1.1.8 *soit X une variable aléatoire réelle admettant un moment d'ordre 2. L'écart-type de X est défini par :*

$$\sigma(X) = \sqrt{var(X)} = \sqrt{\mathbb{E}(X^2) - (\mathbb{E}(X))^2}.$$

Covariance

Définition 1.1.9 *On appelle covariance de deux variables aléatoires X et Y la quantité*

$$cov(X,\ Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))),$$

$$= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),$$

en particulier,

$$var(X+Y) = var(X) + var(Y) + 2cov(X,\ Y).$$

1. $cov(X,\ Y) = cov(Y,\ X)$.
2. $cov(aX + bY) = abcov(X,\ Y)$ et $cov(X+Z,\ Y) = cov(X,\ Y) + cov(Z,\ Y)$.

Corrélation

Définition 1.1.10 *On appelle coefficient de corrélation linéaire de X et de Y, que l'on note $\boldsymbol{corr}(X,\ Y)$, le nombre (s'il existe) :*

$$corr(X,Y) = \frac{cov(X,\ Y)}{\sigma_X\ \ \sigma_Y},$$

où

σ_X *est l'écart-type de* X *et* σ_Y *celui de* Y.

1. $corr\left(X\ ,Y\right) = corr\left(Y,\ X\right).$
2. $corr\left(aX + b,\ cY + d\right) = corr\left(X,\ Y\right).$
3. $-1 \leq corr\left(X,\ Y\right) \leq 1.$

1.2 Lois de probabilité

1.2.1 Lois discrètes

1. $Loi\ de\ Uniforme$

   Une distribution de probabilité suit une loi uniforme lorsque toutes les valeurs prises par la variable aléatoire équiprobables. Si n est le nombre valeurs différentes prises

   $$\forall i \in \mathbb{N}, P\left(X = x\right) = \begin{cases} \frac{1}{n} & \text{si} \quad x_i = 1, \ldots, n \\ 0 & \text{si non} \end{cases}.$$

   - $Espérance\ et\ Variance$

   L'espérance de la loi uniforme discrète par

   $$E\left(X\right) = \frac{n+1}{2}.$$

   La variance de la loi uniforme discrète par

   $$V\left(X\right) = \frac{n^2-1}{12}.$$

• *Fonction répartition*

$$F_X(x) = \left(\frac{x}{n}\right) 1_{[0,n]}(x) + 1_{]n,\infty[}(x)$$

• *Fonction caractéristique*

$$\varphi_X(t) = \frac{e^{it}}{n} \left(\frac{1 - e^{i\,t\,n}}{1 - e^{it}}\right)$$

2. *Loi de Bernoulli*

La loi de probabilité associée à la variable de Bernoulli X, $0 \leq p \leq 1$ telle que :

$$P(X = x_i) = \begin{cases} p;\, si\ x_i = 1 \\ q;\, si\ x_i = 0 \end{cases}$$

avec $P + q = 1$ est appelée loi de Bernoulli notée $\boldsymbol{B}(P)$.

• *Espérance et Variance*

L'espérance de la loi bernoulli $\mathbf{E}(X) = \mu = p.$

La variance de la loi bernoulli $V(X) = \sigma^2 = pq.$

• *Fonction répartition*

$$F_X(x) = (1 - p)\, 1_{[0,1[}(x) + 1_{[1,+\infty[}(x)$$

• *Fonction caractéristique*

$$\varphi_X(t) = 1 - p + pe^{it}$$

3. *Loi binomiale*

La loi binomiale de paramètres $n \in \mathbb{N}^*$ et $0 \leq P \leq 1$, est la loi de probabilité d'une variable aléatoire X d'une répétition de n épreuves de Bernoulli, notée $X \to \boldsymbol{B}\,(n,\ p)$, telle que :

$$P\left(X=x_i\right)=\begin{cases} C_n^{x_i} p^{x_i}\left(1-p\right)^{n-x_i} & \text{si} \quad x_i=0,1,\ldots,n \\ 0 & \text{si non} \end{cases}.$$

• *Espérance et Variance*

L'espérance de la loi binomiale $E\left(X\right)=\mu=np$

La variance de la loi binomiale $V\left(X\right)=\sigma^2=npq$

• *Fonction répartition*

$$F_X\left(x\right)=\begin{cases} \sum\limits_{K=0}^{x} C_k^n p^x\left(1-p\right)^{n-x} & \text{si} \quad x=0,1,\ldots,n \\ 0 & \text{si non} \end{cases}.$$

• *Fonction caractéristique*

$$\varphi_X\left(t\right)=\left(1-p+pe^{it}\right)^n$$

4. *Loi de Poisson*

Une variable aléatoire X à valeurs dans $\mathbb{R}$ suit une loi de Poisson de paramètre λ $(\lambda>0)$ si les réels p_{x_i} sont donnés par :

$$P\left(X=x_i\right)=\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}$$

On note : $X \to \mathcal{P}\left(\lambda\right)$

• *Espérance et Variance*

L'espérance de la loi poisson $E\left(X\right)=\mu=\lambda$.

La variance de la loi poisson $V\left(X\right)=\sigma^2=\lambda$.

• *Fonction répartition*

$$F_X(x) = \exp^{-\lambda} \sum_{x=0}^{n} \frac{\lambda^x}{x!}$$

• *Fonction caractéristique*

$$\varphi_X(t) = e^{\lambda\left(e^{it-1}\right)}$$

1.2.2 Lois continues

1. *Loi Uniforme*

La loi Uniforme est la loi exacte de phénomènes continus uniformément répartis sur un intervalle. La variable aléatoire X suit une loi uniforme sur le segment $[a,\ b]$ avec $a < b$ si sa densité de probabilité est donnée par

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si} \quad x \in [a,\ b]. \\ 0 & \text{si} \quad x \notin [a,\ b]. \end{cases}$$

• *Espérance et Variance*

L'espérance de la loi uniforme continue vaut : $\mathbb{E}(X) = \frac{b+a}{2}$.

La variance de la loi uniforme continue vaut : $V(X) = \frac{(b-a)^2}{12}$.

• *Fonction répartition*

$$F_X(x) = \int_{\infty}^{x} \frac{1}{b-a} 1_{[a,\ b]}(t)\, dt = \begin{cases} 0 & \text{si} \quad x \leq a \\ \frac{x-a}{b-a} & \text{si} \quad x \in [a,\ b] \\ 1 & \text{si} \quad x \geq b \end{cases}$$

- *Fonction caractéristique*

$$\varphi_X(t) = \frac{e^{ibt} - e^{iat}}{i(b-a)t}$$

2. *Loi normale ou loi de Laplace – Gauss*

Une variable aléatoire absolument continue X suit une loi normale de paramètres $(\mu,\ \sigma)$ si sa densité de probabilité est donnée par $f : \mathbb{R} \to \mathbb{R}$

$$x \to f(x) = \frac{1}{\sigma\sqrt{2\Pi}} e - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2.$$

Notation : $X \to N(\mu,\ \sigma)$.

- *Espérance et Variance*

L'espérance de la loi normale vaut : $E(X) = \mu$.

La variance de la loi normale vaut : $V(X) = \sigma^2$.

- *Fonction répartition*

$$F_X(x) = \int_{\infty}^{x} \frac{1}{\sigma\sqrt{2\Pi}} e - \frac{(t-\mu)^2}{2\sigma} dt.$$

- *Fonction caractéristique*

$$\varphi_X(t) = e^{i\mu t} - \frac{t^2\sigma^2}{2}.$$

- *Loi de exponentielle*

Un variable aléatoire X réelle positive suit une loi exponentielle de paramètre

λ positif, si sa densité de probabilité est définie par :

$$f(X) = \begin{cases} 1 - e^{-\lambda x} & ; \quad x \geq 0 \\ 0 & ; \quad x < 0 \end{cases}.$$

• *Espérance et Variance*

L'espérance de la loi exponentielle vaut :$E(X) = \frac{1}{\lambda}$.

La variance de la loi exponentielle vaut :$V(X) = \frac{1}{\lambda^2}$.

• *Fonction répartition*

$$F_X(x) = \int_{\infty}^{x} \lambda e^{-\lambda t} 1_{\mathbb{R}+}(t)\, dt = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ 1 - e^{-\lambda x} & \text{si} \quad x \geq 0 \end{cases}.$$

• *Fonction caractéristique*

$$\varphi_X(t) = \frac{\lambda}{\lambda - it}.$$

1.3 Les différents types de convergence

Une suite (X_n) de variables aléatoires étant une suite de fonctions de $\mathbb{R}$ dans Ω. Il existe diverses façons de définir la convergence de (X_n) dont certaines jouent un grand rôle en calcul des probabilités.

1.3.1 Convergence en loi

Bien quel est plus la faible, elle est très utilisée en pratique car elle permet d'approximer la fonction de répartition de X_n par celle de X.

Définition 1.3.1 *La suite (X_n) converge vers la variable X de fonction de répartition F si, en tout point de continuité de F, la suite (F_n) des fonctions de répartition des X_n converge vers F.*
On note :

$$X_n \xrightarrow{\mathcal{L}} X.$$

Un théorème dû à Polya établit que si F est continue alors la convergence est uniforme Pour des variables discrètes, la convergence en loi vers une v.a discrète s'exprime par :

$$P\left(X_n = X\right) \rightarrow P\left(X = x\right).$$

C'est ainsi qu'on a établi la convergence de la loi binomiale vers la loi de Poisson. Une suite de variables discrètes peut cependant converger en loi vers une variable continue.
On a également que, Si x_n est une suite de variables de densités. F_n et X une variable de densité f, alors :

$$X_n \xrightarrow{\mathcal{L}} X \Longrightarrow f_n\left(x\right) \rightarrow f\left(x\right) \;\; \forall x \in \mathbb{R}$$

La convergence en loi est intimement liée à la convergence des fonctions caractéristiques comme le précise le résultat fondamental suivant, que nous énoncerons sans démonstration :

Théorème 1.3.1 ***(Levy-cramer-dugué)*** *Si $X_n \xrightarrow{\mathcal{L}} X$ alors $\varphi_{X_n}\left(t\right) \rightarrow \varphi_X\left(t\right)$ uniformément dans tout intervalle ni. Si la suite des fonctions caractéristiques*

$\varphi_{X_n}(t)$ *,converge vers une fonction*

φ *dont la partie réelle est continue à l'origine, alors* φ *es tune fonction caractéristique et la suite* X_n *converge en loi vers une v.a* X *dont* φ *est la fonction caractéristique.*

La convergence en probabilité entraîne la convergence en loi. Nous présentons dans ce qui suit la convergence de la loi poisson et la loi binomiale vers la loi normale en détail dans le but de tester la stabilité de notre nouvel algorithme.

1.3.2 Convergence en probabilité

Définition 1.3.2 *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires réelles définies sur espace de probabilité* $(\Omega,\ \mathcal{A},\ P)$*.* X_n *converge vers* X *un autre variable aléatoires, en probabilité si :*

$$\forall \varepsilon > 0, \lim_{n\to\infty} \mathbf{P}\left(\mid X_n - X \mid > \varepsilon\right) = 0.$$

1.3.3 Convergence presque sûre

Définition 1.3.3 *La suite* (X_n) *converge presque sûrement vers* X *si :*

$$P\left(\{\omega : X_n(\omega) \neq X(\omega)\}\right) = 0.$$

et on note : $X_n \overset{Ps}{\to} X$*.*

En d'autres termes, l'ensemble des points de divergence est de probabilité nulle.

Remarque 1.3.1 *La convergence presque sûrement implique la convergence en probabilité.*

1.3.4 Convergence en moyenne d'ordre p

Définition 1.3.4 *La suite de variable aléatoires $(X_n)_{n\geq 1}$ converge vers X un autre variables aléatoires,en moyenne l'ordre $p \geq 1$ si :*

$$\lim_{n\to\infty} E\left(\mid X_n - X \mid^p\right) = 0$$

1.3.5 Loi des grands nombres

Définition 1.3.5 (Moyenne empirique) *Soient $X_1, X_{2,\ldots,}X_n$, n variables aléatoires indépendantes et de même loi de variance σ^2.On définit la somme de ces variables aléatoires par :*

$$S_n = \sum_{i=1}^{n} X_i,$$

et la moyenne empirique est donnée par :

$$\bar{X} = \frac{S_n}{n}.$$

Preuve.

1. $E\left[\bar{X}_n\right] = E\left[X_1\right].$
2. $Var\left[\bar{X}_n\right] = \frac{\sigma^2}{n}.$

■

Loi faible des grands nombres

Théorème 1.3.2 *Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles définies sur le même espace $(\Omega,\ \mathcal{F},\ P)$, indépendantes, d'espérance μ et de variance σ^2*

finie $\left(\lim\limits_{n\to\infty}\frac{1}{n^2}\sum\limits_{i=1}^{n}\sigma^2=0\right)$

Alors on a :

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \quad \underset{n\to\to\infty}{\overset{p}{\longrightarrow}} E\left(X_1\right).$$

Loi forte des grands nombres

Théorème 1.3.3 *Soit* $\left(X_n\right)_{n\in\mathbb{N}^*}$ *une suite de v.a. réelles i.i.d et intégrable* $\left(E\left[\mid X_1\mid<+\infty\right]\right)$

.

Alors on a :

$$\bar{X}_n \;=\frac{1}{n}\sum_{i=1}^{n}X_i \quad \underset{n\to\infty}{\overset{ps}{\longrightarrow}} E\left(X_1\right),$$

ou

$$\lim_{n\to\infty}E\left[\mid \bar{X}_n-E\left(X_1\right)\mid\right]=0.$$

Chapitre 2

Théorème central limite

Le théorème de la limite centrale est un résultat concernant la convergence de la loi d'une séquence de variables aléatoires. Ce résultat indique intuitivement que toute somme de variables aléatoires indépendantes et identiquement distribuées tend vers à une variable aléatoire gaussienne. L'une des approximations étudiées dans ce chapitre est l'approximation de la loi binomiale et de la distribution de Poisson par rapport à la distribution normale. Nous étudions également l'estimation, y compris l'estimation ponctuelle, l'estimation par intervalles ponctuels, les méthodes du maximum de vraisemblance et enfin, l'intervalle de confiance.

2.1 Formulation et interprétation du théorème central limite

Théorème 2.1.1 *Soit* $(X_n)_{n\geq 1}$ *une suite de variables aléatoires réelles indépendantes et identiquement distribuées i.i.d, de moyenne* $\mathbb{E}(X_n) = m$ *et de variance*

$Var(X_n) = \sigma^2 \leq \infty$. *Alors, lorsque* $n \to +\infty$, *la suite normalisée :*

$$\frac{S_n - nm}{\sqrt{n\sigma^2}},$$

converge en loi vers une variable aléatoire de loi normale centrée réduite $N(0,1)$, *où*

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Preuve du théorème central limite

Posons $Y_i = \frac{X_i - \mu}{\sigma\sqrt{n}}$. Alors

$$\varphi_{Y_i}(t) = \varphi \frac{x_i - \mu(t)}{\sigma\sqrt{n}} = \varphi_{X_i - \mu}\left(\frac{t}{\sigma\sqrt{n}}\right).$$

Pour t fixé, lorsque n tend vers l'infini, $\frac{t}{\sigma\sqrt{n}}$ est infiniment petit . Ecrivons le développement limité, au voisinage de 0. de la fonction caractéristique d'une variable aléatoire W :

$$\begin{aligned} \varphi_W(u) &= \varphi_W(0) + \mu\varphi'_W(0) + \frac{u^2}{2}\varphi''_W(0) + u^2\varepsilon(u). \\ &= 1 + iuE(W) - \frac{u^2}{2}E\left(W^2\right) + u^2\varepsilon(u). \end{aligned}$$

En posant $W = X_i - \mu, u = \frac{t}{\sigma\sqrt{n}}$, on a $E(W) = E(X_i - \mu) = 0$ et $E(W^2) = E((X_i - \mu)^2) = V(X_i) = \sigma^2$ d'où

$$\varphi_{X_i - \mu}\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2}{2\sigma^2 n}\sigma^2 + \frac{1}{n}\varepsilon\left(\frac{t^3}{\sigma^3\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + \left(\frac{1}{n}\varepsilon_i(n)\right),$$

avec $\lim_{n\to\infty} \varepsilon_i(n) = 0$.

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

L'indépendance des X_n entraîne cella des Y_i et ainsi

$$\begin{aligned}\varphi_{Z_n}(t) &= \prod_{i=1}^{n} \varphi_{Y_i}(t) \\ &= \exp\left(\sum_{i=1}^{n} \ln\left(1 - \frac{t^2}{2n} + \frac{1}{n}\varepsilon_i(n)\right)\right),\end{aligned}$$

et sachant que $\lim_{n\to\infty} \left(1 + \frac{a}{n}\right)^n = e^a$. On a :

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

qui est la fonction caractéristique $\mathcal{N}(0,1)$

Ce théorème établit une propriété générale, qui va justifier l'importance considérable de la loi normale, à la fois comme modèle pour décrire des situations pratiques, mais aussi comme outil théorique.

2.2 Application du théorème central limite

Dans la suite, nous étudierons deux applications classiques du théorème central limite : l'approximation de la loi binomiale et celle de la loi poisson par la loi normale.

2.2.1 Approximation de la loi binomiale par la loi normale

Théorème 2.2.1 *Soit* (X_n) *une suite de variables aléatoires suivant une loi binomiale* $\boldsymbol{B}(n,\ p)$. *En notant* $q = 1 - p$, *On a la convergence en loi suivante :*

$$\frac{X_n - np}{\sqrt{npq}} \xrightarrow{\mathcal{L}} N(0,1).$$

Cela signifie que la variable centrée et réduite suit asymptotiquement une loi normale standard.

La fonction caractéristique de X_n *vaut* $(p\exp(it) + 1 - p)^n$ *donc celle de* $\frac{X_n-np}{\sqrt{npq}}$ *vaut*

$$\varphi(t) = \left(p\exp\left(\frac{it}{\sqrt{npq}}\right) + 1 - p\right)^n \exp\left(-\frac{itnp}{\sqrt{npq}}\right)$$
$$\ln\ \varphi = n\ln\left(p\left(\exp\left(\frac{it}{\sqrt{npq}}\right) - 1\right)\right) - \frac{itnp}{\sqrt{npq}}.$$

Développons au deuxième ordre l'exponentielle, il vient

$$\ln\ \varphi = n\ln\left(1 + p\left(\exp\left(\frac{it}{\sqrt{npq}}\right) - \frac{t^2}{2npq}\right)\right) - \frac{itnp}{\sqrt{npq}},$$

puis le logarithme

$$\ln\ \varphi = n\left[\frac{pit}{\sqrt{npq}} - \frac{pt^2}{2npq} + \frac{p^2t^2}{2npq}\right] - \frac{itnp}{\sqrt{npq}},$$

soit

$$\ln\ \varphi = -\frac{t^2}{2q} + \frac{pt}{2q} = \frac{t^2}{2q}\left(P-1\right) = -\frac{t^2}{2},$$

car $p = 1 - q$.

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

qui est la fonction caractéristique de la loi normale centrée-réduite.

Lorsqu'on a une valeur suffisamment grande de n*, il est possible d'approximer la loi binomial par la loi normale. En général, cette approximation est considérée valide sous la condition* $np \geq 5$ *et* $nq \geq 5$*.*

Cependant, il est nécessaire d'appliquer ce qu'on appelle une correction de continuité. En effet, la convergence de la loi binomiale vers la loi normale implique que les extrémités du diagramme en bâtons de la loi binomiale $\boldsymbol{B}\left(n,\ p\right)$ *sont proches de la courbe de densité de la loi normale* $\mathcal{N}\left(np,\ \sqrt{npq}\right)$.

2.2.2 Approximation de la loi de Poisson par la loi normale

Théorème 2.2.2 *Soit* X_n *une famille de variables aléatoires de loi de poisson de paramètre* λ*. Lorsque* λ *tend vers l'infini, la variable centrée et réduite* $\frac{X_\lambda-\lambda}{\sqrt{\lambda}}$ *converge en loi vers une loi normale standard* $\mathcal{N}\left(0,1\right)$

La fonction caractéristique de X *est*

$$\varphi_X\left(t\right) = \exp\left(\lambda\left(\exp\left(\frac{it}{\sqrt{\lambda}}\right)-1\right)\right),$$

d'où

$$\varphi_{\frac{X-\lambda}{\sqrt{\lambda}}}(t) = \exp\left(\lambda\left(\exp\left(\frac{it}{\sqrt{\lambda}}\right)-1\right)\right)\exp\left(-\frac{it\lambda}{\sqrt{\lambda}}\right),$$
$$= \exp\left(\lambda\exp\left(\frac{it}{\sqrt{\lambda}}\right)-\lambda-it\sqrt{\lambda}\right),$$

comme

$$\exp\left(\frac{it}{\sqrt{\lambda}}\right)\simeq 1+\frac{it}{\sqrt{\lambda}}-\frac{t^2}{2\lambda},$$

il vient

$$\varphi_{\frac{X-\lambda}{\sqrt{\lambda}}}(t) = \exp\left(\lambda+it\sqrt{\lambda}-\frac{t^2}{2}-\lambda-it\sqrt{\lambda}\right) = \exp\left(-\frac{t^2}{2}\right).$$

La figure 2.1illustre l'approximation de la loi de Poisson $P(\lambda)$ par la loi de Gauss de même espérance λ et de même écart-type $\sqrt{\lambda}$.

L'approximation est très satisfaisante pour $\lambda > 18$. On trouvera en d'autres for mules d'approximation plus précises. On a, ici encore, intérêt à effectuer la correction de continuité.

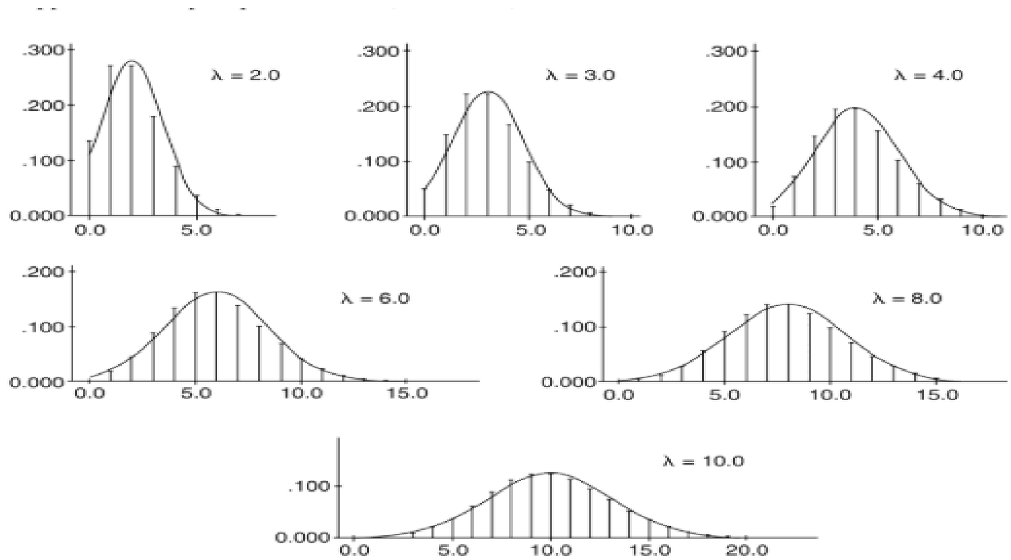

FIG. 2.1 – l'approximation de la loi de Poisson

2.3 Estimation

Après avoir étudié l'approximation des probabilités par la loi normale grâce au théorème central limite, nous allons maintenant nous intéresser à une application essentielle de l'analyse statistique : l'estimation des paramètres d'une population à partir d'un échantillon.

Estimer consiste à proposer des valeurs approchées pour ces paramètres inconnus en s'appuyant sur les données observées.

Pour réaliser cette estimation, différentes méthodes peuvent être employées. Nous présenterons principalement l'estimation ponctuelle.

2.3.1 Estimation ponctuelle

On cherche à estimer un paramètre inconnu θ d'une population, tel que la moyenne μ, l'écart-type σ ou une proportion p, Pour ce faire, on utilise un estimateur, noté

T, qui est une statistique, c'est-à-dire une fonction des variables aléatoires observées $(X_1, \ldots, X_2)$.La valeur calculée de cet estimateur à partir des données observées $(x_1,\ x_2,\ \ldots,\ x_3)$ fournit une estimation de θ. Cette estimation est considérée comme une «bonne approximation» de la vraie valeur du paramètre θ.

Qualité d'un estimateur

Définition 2.3.1 *On appelle biais de T pour θ la valeur*

$$b_\theta (T) = \mathbb{E} (T) - \theta.$$

Un estimateur T est dit sans biais si $\mathbb{E} (T) = \theta$.

2.3.2 Méthode du maximum de Vraisemblance

Définition 2.3.2 *La méthode du maximum de vraisemblance. Cette méthode consiste, étant donné un échantillon de valeur X_1, X_2, $\ldots$, X_n à prendre comme estimation de θ la valeur θ qui rend maximale la vraisemblance :*

$$L (x_1, \ldots\ ,\ x_n) = \prod_{i=1}^{n} f (x_i;\ \theta) .$$

En pratique on prend comme estimation de θ une solution de l'équation

$$\frac{\partial}{\partial \theta} \ln\ L (X;\ \theta) = 0,$$

dite équation de la vraisemblance.

2.3.3 Méthode des Moments

Définition 2.3.3 *La méthode des moments a été utilisée pour la première fois par **Karl Pearson**. Cette méthode repose sur l'idée d'égaliser les k premiers moments de la distribution en question avec les moments empiriques correspondants. Il faut résoudre un système de k équations, avec k le nombre de paramètres inconnues.*

Exemple 2.3.1 *On prélève un échantillon aléatoire $(X_1,\ X_2, \ldots,\ X_n)$ suit une loi $\mathcal{N}\,(0,\ 1)$ et on peut estimer les paramètres de cette loi par les deux méthodes $M.V$ et méthode des moments.*

· **Par la méthode(M. V) :**

On a : la densité de la loi normale

$$f_{x_i}\left(x_i\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\left(x_i - \mu\right)^2\right\}.$$

Alors, l'équation de vraisemblance

$$L\left(x_1, x_2, \ldots, x_n; \mu, \sigma^2\right) = \frac{1}{\left(2\pi\sigma^2\right)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2\right\},$$

$$\ln\ L\left(x_1, x_2, \ldots, x_n; \mu, \sigma^2\right) = -\frac{n}{2}\ln\left(2\pi\right) - \frac{n}{2}\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2.$$

$$\frac{\partial \ln\ L\left(x, \ldots, x_n; \mu, \sigma^2\right)}{\partial \mu} = \frac{2}{2\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right) = 0.$$

$$\frac{\partial \ln\ L\left(x_1, x_2, \ldots, x_n; \mu, \sigma^2\right)}{\partial \sigma^2} = -n + \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2 = 0.$$

Donc,

$$\Longrightarrow \sum_{i=1}^{n} x_i = n\mu \Longrightarrow \mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

$$\Longrightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2 = n.$$

$$\Longrightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu\right)^2.$$

$$\Longrightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2.$$

· **Par la méthode des moments :**

Nous avons vu précédent que $\mathbb{E}(X) = \mu$ et $\mathbb{V}(X) = \sigma^2$. Alors,

$$\mathbb{E}(X) = \mu = \bar{x}.$$

On a :

$$\begin{aligned}\mathbb{V}(X) = \sigma^2 &= \mathbb{E}(X) - (\mathbb{E}(X))^2 \\ &= \mathbb{E}(X) - \mu,\end{aligned}$$

alors,

$$\mathbb{E}\left(X^2\right) = \sigma^2 + \mu^2,$$

d'autre par on a :

$$E\left(X^2\right) = \frac{\sum X_i^2}{n}.$$

On en déduit :

$$\hat{\sigma}^2 = \frac{\sum X_i^2}{n} - \hat{\mu}^2 = \frac{\sum X_i^2}{n} - \bar{X}^2$$
$$= \frac{1}{n} \sum \left(X_i - \bar{X}\right)^2,$$

et enfin,

$$\hat{\mu} = \bar{X}$$
$$\hat{\sigma}^2 = s^2 \left(s^2 \text{ est un estimateur biaisé}\right)$$

2.4 Intervalles de confiance

Après l'estimation des paramètres de la loi normale on passe à l'intervalle de confiance.

Définition 2.4.1 *On appelle intervalle de confiance tout intervalle construit autour d'un estimateur ayant une certaine probabilité de contenir la valeur du paramètres correspondant de la population.*

2.4.1 Intervalle pour μ quand σ est connu

Supposons avoir des observation aléatoire i.i.d. $X_1, \ldots, X_n$ de la loi commune $N\ (\mu, \sigma^2)$ avec σ^2connu et μ à estimer.
On a la moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ qui suit la loi $N\ \left(\mu, \frac{\sigma^2}{n}\right)$. Alors,

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

En notant $z_{\frac{\alpha}{2}} = \left(-z_{1-\frac{\alpha}{2}}\right)$ et $z_{1-\frac{\alpha}{2}}$ les quantiles d'ordre $\frac{\alpha}{2}$ et $1-\frac{\alpha}{2}$ de la loi normale standard tel que

$$P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \ z_{1-\frac{\alpha}{2}}\right)$$

Donc, l'intervalle de confiance au niveau $1-\alpha$ pour le paramètre de position μ quand σ est connu correspond à

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

Représentation graphique :

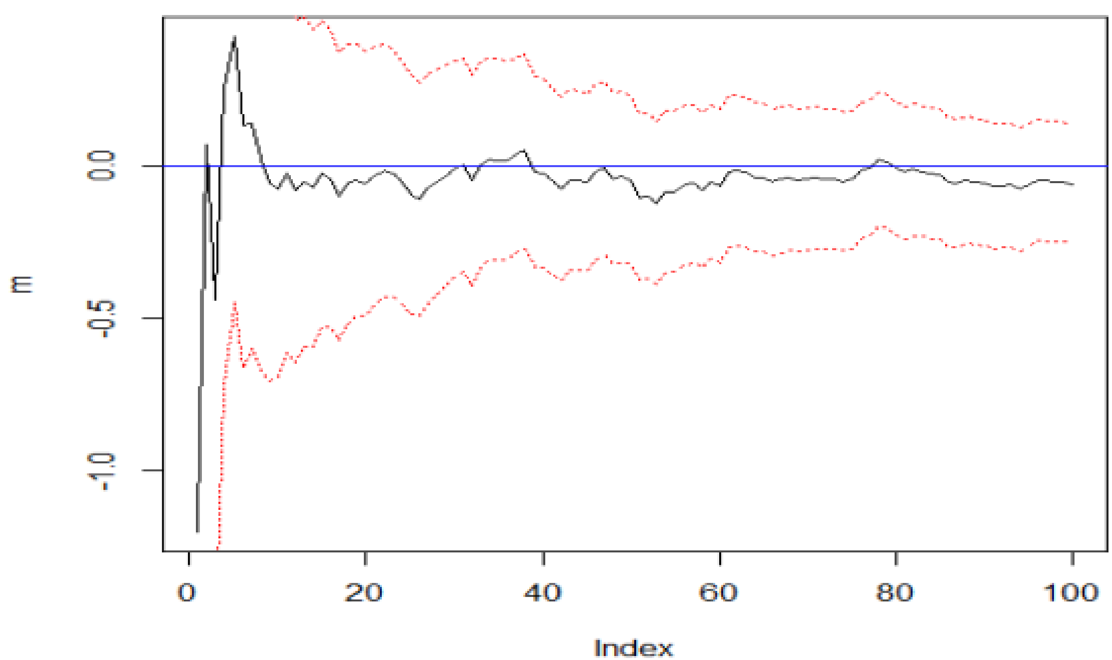

FIG. 2.2 – Intervalle pour μ quand σ est connu

Legend du graphique

- **Ligne rouge (pointillée) :** Représente les bornes supérieure et inférieure de l'intervalle de confiance.

• **Ligne bleue (continue) :** Représente la valeur réelle du paramètre (si elle est connue) ou la moyenne des estimations.

• **Ligne noire (continue) :** Représente l'estimation ponctuelle du paramètre (par exemple : la moyenne de l'échantillon $\bar{X}$ ou la variance de l'échantillon S^2) en fonction de l'augmentation de la taille de l'échantillon.

2.4.2 Intervalle pour μ quand σ est inconnu

Lorsque le paramètre d'échelle est inconnu on utilise leur estimateur s^2, tel que

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \text{ (estimateur sans bais).}$$

Donc, on obtient la statistique de test suivante :

$$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$$

Lemme 2.4.1 *Soient* $X_1, \ldots, X_n$ *des variables aléatoire i.i.d de loi commune* $\mathcal{N}(\mu, \sigma^2)$. *Alors nous avons*

· La moyenne empirique $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est de $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

· En notant $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$la variance empirique, il suit que $\frac{(n-1)\, S^2}{\sigma^2}$ suit une loi de khi-deux de degré $n-1$.

D'après le lemme précédent on obtient :

$$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} = \frac{\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}}{\sqrt{\frac{\frac{(n-1)\, S^2}{\sigma^2}}{\sqrt{(n-1)}}}} \sim t_{n-1}.$$

Par une simplification on a :

$$\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}.$$

Comme précédemment on déduire l'intervalle de confiance en partant de

$$P\left(-t_{n-1;1-\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \leq t_{n-1;1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

On aboutisse à l'intervalle de confiance

$$\left[\bar{X} - t_{n-1;1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}; \bar{X} + t_{n-1;1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$$

Représentation graphique :

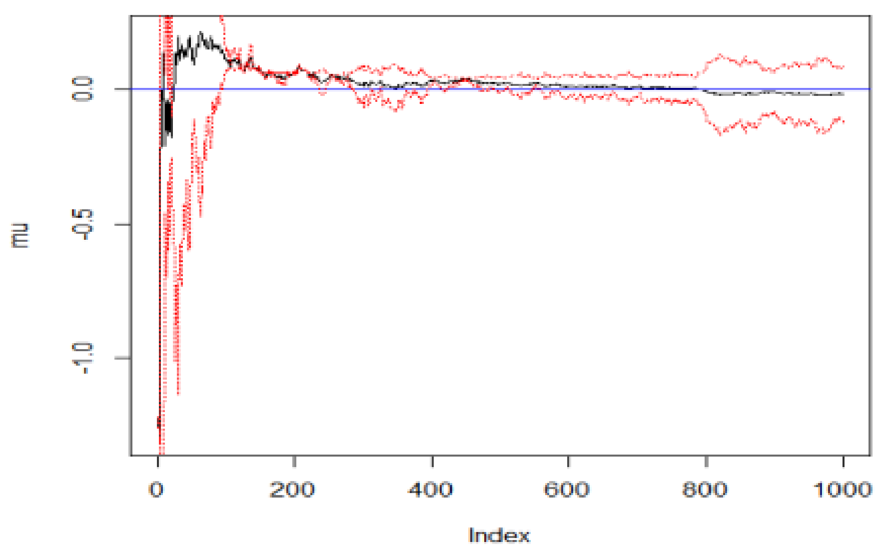

FIG. 2.3 – Intervalle pour μ quand σ est inconnu

Legend du graphique

(Intervalle de confiance pour μ)

• **Rouge :** bornes supérieure et inférieure de l'intervalle de confiance pour μ.

• **Bleu :** valeur réelle de μ (souvent 0).

2.4.3 Intervalle pour σ

On utilise le lemme de Fisher pour produire l'intervalle de confiance pour le paramètre d'échelle. Donc d'après les quantiles de la loi de $\mathcal{X}^2(n-1)$ on construire l'intervalle de confiance.

Premièrement on a :

$$\mathcal{P}\left(\mathcal{X}^2_{n-1;\frac{\alpha}{2}} \leq \frac{(n-1)\ s^2}{\sigma^2} \leq \mathcal{X}^2_{n-1;1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Par simples manipulations on aboutisse à l'intervalle de confiance

$$\left[\frac{(n-1)\ s^2}{\mathcal{X}^2_{n-1;1-\frac{\alpha}{2}}};\frac{(n-1)\ s^2}{\mathcal{X}^2_{n-1;\frac{\alpha}{2}}}\right].$$

Représentation graphique

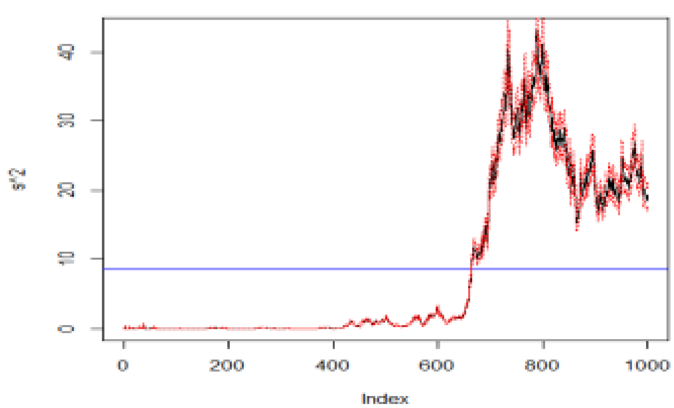

FIG. 2.4 – Intervalle pour σ

Legend du graphique

(Intervalle de confiance pour σ) :

- **Rouge :** bornes supérieure et inférieure de l'intervalle de confiance pour σ.
- **Bleu :** moyenne des estimations de la variance.

Chapitre 3

Test de normalité et application du TCL

Les tests de normalité en statistique permettent d'évaluer si un ensemble de données suit une distribution normale. Ils s'inscrivent dans la catégorie des tests d'adéquation, qui visent à comparer la distribution d'un échantillon à une distribution théorique. L'intérêt particulier pour la loi normale s'explique par son rôle central dans de nombreuses méthodes statistiques, mais aussi par son importance théorique, notamment à travers le théorème central limite.

Dans le cadre de ce chapitre, nous nous intéresserons aux différentes approches permettant d'évaluer la normalité d'un jeu de données. Nous présenterons d'abord les méthodes graphiques, telles que les histogrammes, les courbes de densité et les diagrammes quantile-quantile (Q-Q plots), qui offrent une première évaluation visuelle de la normalité. Nous introduirons ensuite des méthodes plus formelles, en particulier le test de Shapiro-Wilk, largement utilisé pour tester l'hypothèse de normalité.

3.1 Méthodes graphiques

3.1.1 Histogramme de fréquence

L'histogramme de fréquence constitue l'outil graphique le plus simple. Il consiste à diviser automatiquement l'intervalle de définition de la variable en k sous-intervalles de largeur égale, puis à représenter une série de barres dont la hauteur est proportionnelle à l'effectif correspondant à chaque sous-intervalle.

x = rnorm (5000)

hist(x, col= "4")

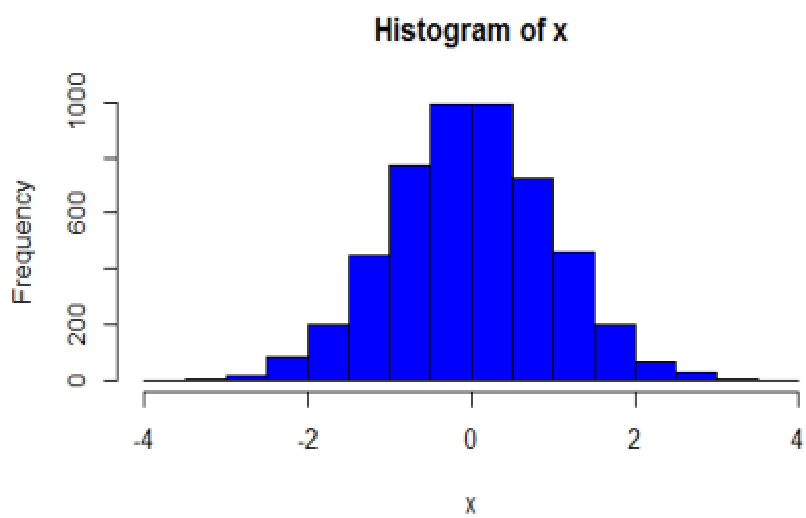

FIG. 3.1 – Histogramme de fréquence

Il est possible de visualiser la forme de la distribution des données à analyser en les représentants sous forme d'histogramme puis de comparer la forme de cet histogramme avec une courbe représentant une loi normale. Ce ci ne permet pas de conclure à la normalité des données mais peut donner une idée du type de loi.

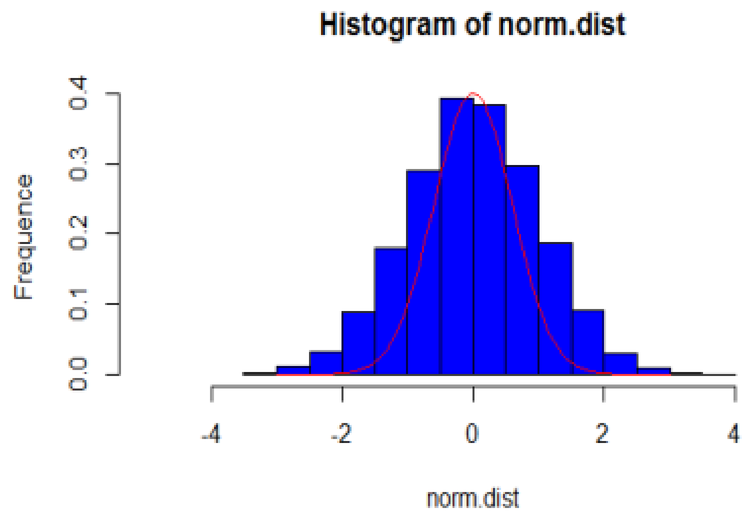

FIG. 3.2 – Histogramme de fréquence de distribution et densité de la loi normale

3.1.2 Le Q-Q(Quantile-Quantile) plot graphe

Le principe du QQ-plot général est de comparer la position de certains quantiles dans la population observée avec leur position dans la population théorique. Le QQ-plot est un outil graphique qui permet d'apprécier visuellement la normalité d'une v.a. La fonction à utiliser est la fonction qnorm ().

$x = rnorm(1000; 0; 1)$;

$qqnorm(x)$; # pour dessiner les quantile de loi normale $\mathcal{N}(0; 1)$ avec les quantiles de x

$qqline(x, col = 2)$; # ajouter une ligne au graphe qqnorm

Interprétation statistique Si les points sont alignés autour de la droite tracée par la fonction qqline (), c'est que la distribution de la variable étudiée est celle d'une loi normale.

Remarque 3.1.1 *Il existe également une autre méthode graphique des tests de*

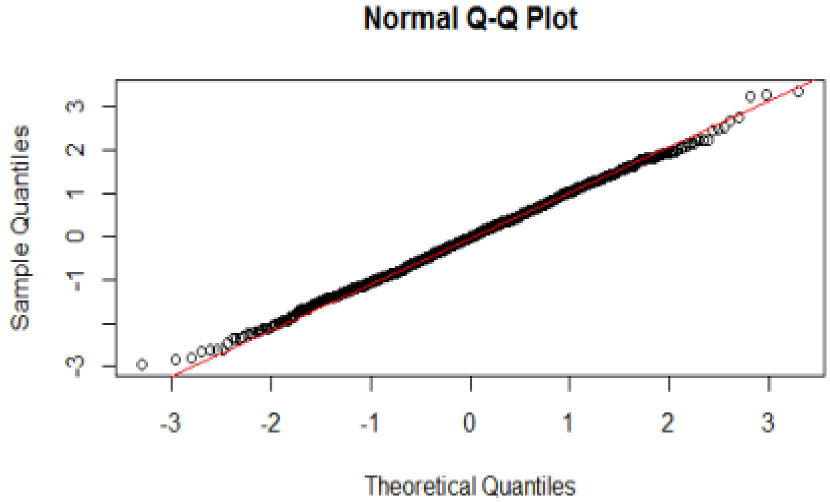

FIG. 3.3 – Graphe de quantiles de loi normale et les quantile de x

normalité comme la droite de Henry (est une méthode graphique pour ajuster une distribution gaussienne à celle d'une série d'observations).

3.2 Tests statistiques

3.2.1 Test de Shapiro-wilk

Le test de Shapiro-Wilk est conçu spécialement pour étudier la non-norrnalité d'une variable continue X. C'est le test le plus puissant pour tester la normalité d'une distribution. Les hypothèses sont : H_0 : X suit une loi normale donc si p-value $> 0,01$, l'hypothèse de normalité est compatible avec nos données et H_1 : X ne suit pas une loi normale.

La statistique de test est

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

où a_i sont des constantes générées à partir de la moyenne et de la matrice de variance covariance des quantiles d'un échantillon de taille n suivant la loi normale. Ces constantes sont fournies dans des tables spécifiques.

Instruction R : L'instruction à utiliser est : **shapiro. test ().**

Exemple 3.2.1

v<rnorm (1000, mean = 20, sd = 2)

Résultat : Le résultat du test statistique est :

(res.sh = shapiro.test (v))

Shapiro-Wilk normality test

data : v

W= 0.99884, p-value = 0.78

Exemple 3.2.2

u < runif(100, min = 2, max = 4)

Résultat : Le résultat du test statistique est :

(res.sh = shapiro.test(u))

Shapiro-Wilk normality test

data : u

W = 0,95295, p-value = 0,001304

Interprétation : Si la valeur-p est inférieure à une seuil α qu'on s'est donné (en général, 5%), alors on rejette H_0. Ici, la valeur-p dans exemple 3.2.1 est égale à $0,78$, ce qui nous

conduit (bien entendu) à accepter H_0. Et pour exemple 3.2.2 la valeur-p est égale à $0,001304$ donc on rejette H_0.

Remarque 3.2.1 *Il existe également un grand nombre de tests statistiques de normalité.*

3.3 Application du TCL sous R

3.3.1 Approximation gaussienne de la loi binomiale

```
n = 800
N = 1000
p = 0.3
size = 50
S = numeric(N)
M = numeric(N)
for(i in 1 :N){
X = rbinom(n, size, p)
M[i] = mean(X)
S[i] = sqrt(n)*(M[i] - size*p)/sqrt(size*p*(1-p))
}
par(mfrow = c(1, 2))
qqnorm(S, main = "QQ Plot")
qqline(S, col = "blue")
hist(S, prob = TRUE)
lines(density(S), col = "red")
abline(v = mean(S), lty = 2, lwd = 2, col = "blue")
```

shapiro.test(S)

print(shapiro_test)

Shapiro-Wilk normality test

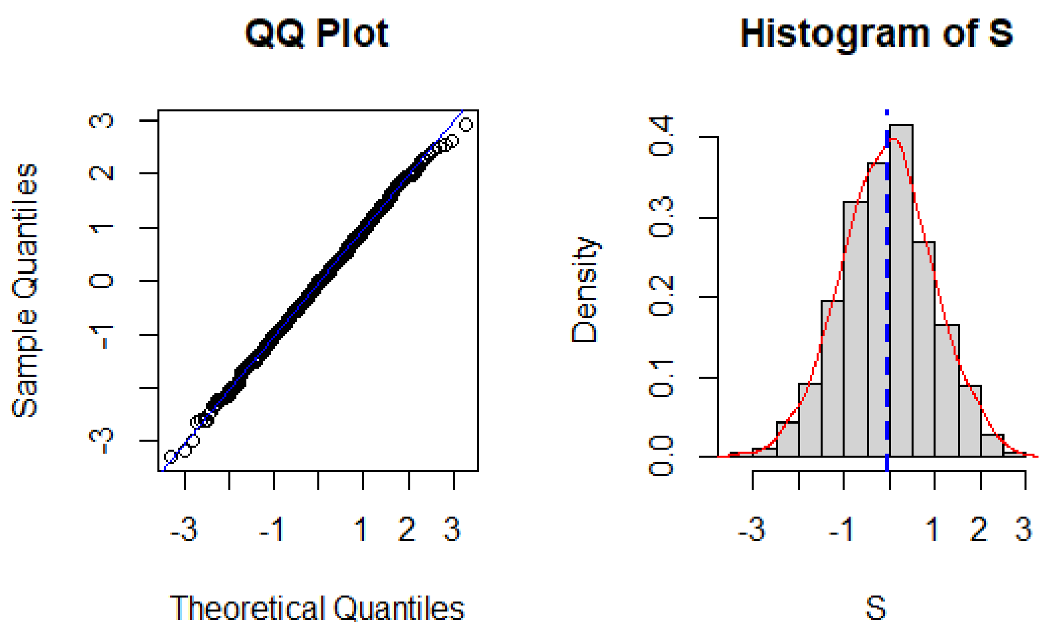

Fig. 3.4 – Graphe de quantile et d'histogramme de S

shapiro.test(S)

Shapiro-Wilk normality test

data : S

W = 0.99874, p-value = 0.7162

3.3.2 Approximation gaussienne de la loi de poisson

n = 5000

N = 3000

```
lambda =2
S = numeric(N)
M = numeric(N)
for(i in 1 :N) {
X = rpois(n, lambda)
M[i] = mean(X)
S[i] <- sqrt(n) * (M[i] - lambda) / sqrt(lambda)
}
par(mfrow = c(1, 2))
qqnorm(S)
hist(S, proba="T")
lines(density(S), col =2)
abline(v=mean(S), lty = 2, lwd = 2,col = 2)
shapiro.test(S)
```

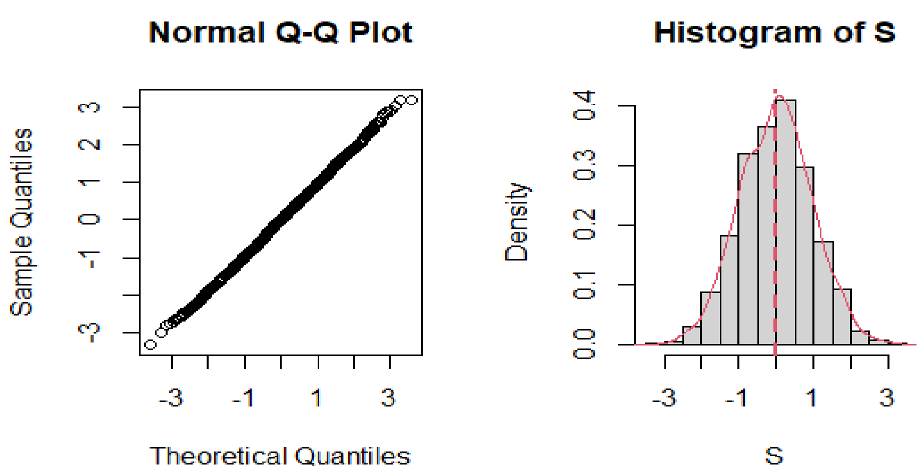

FIG. 3.5 – Graphe de quantile, et d'histogramme de S

Shapiro-Wilk normality test

data : S

W = 0.99952, p-value = 0.6838

3.3.3 Approximation gaussienne de la loi exponentielle

N = 1000

n = 90

S = numeric(N)

for(i in 1 :N){

X = rexp(n)

S[i] = sqrt(n)*(mean(X)- 1)}

qqnorm(S)

shapiro.test(S)

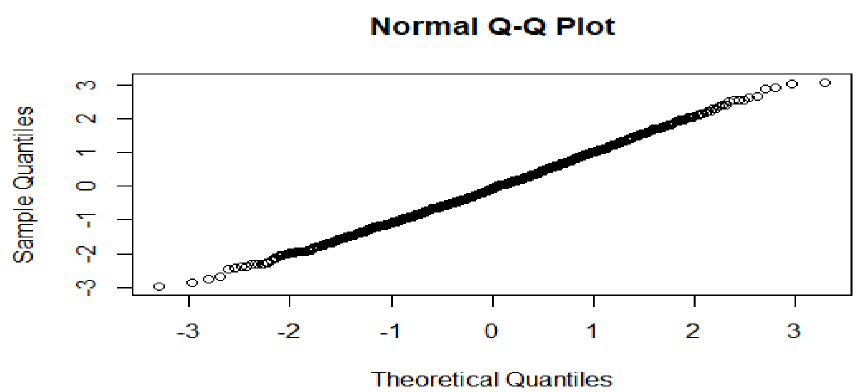

FIG. 3.6 – Graphe de quantile et de densités de s

Shapiro-Wilk normality test

data : S

W = 0.99764, p-value = 0.1631

Commentaire On observe dans chaque approximation que il existe une similarité entre l'histogramme de S et la courbe de densité de la loi normale et les points des quantiles sont alignées, le p-value de S est non significatif, alors S suit la loi normale centrée réduit.

Conclusion

Dans ce travail, nous avons étudié la normalité asymptotique d'une suite de variable aléatoires indépendantes à travers le prisme du Théorème Central limite (TCL), l'un des piliers fondamentaux de la théorie des probabilité. Après avoir rappelé les notions de base telles que les variables aléatoires, les fonctions caractéristiques, les lois de probabilité et les différents types de convergence, nous avons approfondi la compréhension théorique et pratique du TCL.

Nous avons mis en évidence comment cette propriété asymptotique permet d'approximer des lois complexes par la loi normale, facilitant ainsi les méthodes d'estimation, la construction d'intervalles de confiance, et diverses applications numériques, nous avons illustré la convergence en loi des sommes de variables aléatoires vers la distribution normale, confirmant ainsi les résultats théoriques de manière empirique.

Cette étude montre l'importance du TCL non seulement comme résultat théorique mais aussi comme outil pratique puissant en statistique et en sciences des données. Elle ouvre la voie à de nombreuses applications, notamment en inférence statistique, en finance, et en modélisation stochastique.

Bibliographie

[1] Admane, A. (2019). Types de convergences de suite variable aléatoire [Mémoire de master, Université Kasdi Merbah Ouargla, Faculté des Mathématiques et des Sciences de la Matière].

[2] Dusart, P. (2018). Cours de statistiques inférentielles.

[3] Ghiboub, D. (2018). Théorème de la limite centrale et applications [Mémoire de master, Université Mohamed Khider Biskra].

[4] Perrot, A. (2010,août). Cours de probabilités et statistiques [Stage ATSM, Département deMathématiques, IREM de Lyon, Université Claude Bernard Lyon 1].

[5] Racicot, F. É., & Théoret, R. (2006). Traité d'économétrie financière : Modélisation financière. Presses de l'Université du Québec.

[6] Rotambeleke, R. (2011). Tests de normalité [Université Lumière Lyon 2].

[7] Saporta, R. (2006). Probabilités, analyse des données et statistiques (2e éd. rév. et augm.). Éditions Technip.

[8] Vessery, R. (2006). Aide- mémoire statistique et probabilité pour l'ingénieur (3e éd.). Dunod.

[9] Zerrouki, O. (2023). Ajustement d'un échantillon de la loi inconnue : Modèle "loi. test" [Mémoire de master, Université Kasdi Merbah Ourgla]. Université

Kasdi Merbah-Ouargla.

الملخص

تتناول هذه المذكرة دراسة الانتماء التدريجي إلى التوزيع الطبيعي لسلسلة من المتغيرات العشوائية المستقلة، وذلك من خلال تحليل نظري وتطبيقي لمبرهنة الحد المركزي. في البداية، قمنا بتذكير بالمفاهيم الأساسية المتعلقة بالمتغيرات العشوائية، وأنواع التقارب. وكذا القوانين الاحتمالية. بعد ذلك تعمقنا في دراسة مبرهنة الحد المركزي. وفي الجزء الأخير، قمنا بمحاكاة رقمية لتوضيح ظاهرة التقارب نحو التوزيع الطبيعي. مما عزز النتائج النظرية بأدلة تجريبية.

Résumé

Ce mémoire traite de l’étude de l’appartenance asymptotique à la loi normale pour une suite de variables aléatoires indépendantes. L’étude se base sur l’analyse théorique et l’application du théorème central limite. Dans un premier temps, les concepts fondamentaux liés aux variables aléatoires, leurs types et lois de probabilité sont rappelés. Ensuite, une étude approfondie du théorème central limite est présentée. Enfin, la dernière partie comprend une simulation numérique illustrant le phénomène de convergence vers la loi normale, ce qui renforce les résultats théoriques par des expériences empiriques.

Abstract

This memory focuses on the asymptotic convergence to the normal distribution for a sequence of independent random variables. The study is based on theoretical analysis and the application of the Central Limit Theorem. Intially, it reviews the fundamental concepts related to random variables, their types, and probability laws. Then, a detailed examination of the Central Limit Theoreme is presented. Finally, the last part includes a numerical simulation illustrating the convergence phenomenon towards the normal distribution, thereby supporting the theoretical results with empirical experiments.