

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 Melle. **NAIMI Taouss**

Titre :

---

**Intervalle de confiance pour la moyenne et la variance  
dans le cas d'un échantillon Gaussien**

---

Membres du Comité d'Examen :

Pr. SAYAH Abdallah	Université de Biskra	Président
Pr. YAHIA Djabrane	Université de Biskra	Encadreur
Dr. KHEREDDINE Souraya	Université de Biskra	Examinatrice

Juin 2024

## *Dédicace*

*Je dédie ce humble travail*

*A mon très cher père **BRAHIM** et ma très chère mère, aimable **FATMA** : aucune dédicace ne saurait exprimer l'amour, le dévouement et le respect que j'ai toujours eu pour vous, ce travail est le fruit de vos sacrifices consentis pour mon éducation et ma formation.*

*Je vous dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, vous préserver et accorde santé, longue vie et bonheur. Vos prières et bénédiction m'ont été un grand secours pour mener à bien mes études. Aucune dédicace ne saurait être assez éloquente pour exprimer ce que vous méritez pour tous les sacrifices depuis ma naissance, durant mon enfance et même à l'âge adulte.*

*A mes chers frères **SADDAM** "allah yarhmou" et **ABDOU**, et à mes chers sœurs **EMITELLI** et **ASMA** en témoignage de l'attachement, de l'amour et de l'affection que je porte pour vous. Je vous remercie d'être l'épaule sur laquelle je peux toujours compter. Je ne peux trouver les mots justes et sincères pour vous exprimer mon affection et mes pensées.*

*Vous êtes toujours dans mon cœur, je vous aime beaucoup.*

*A mes petites princesses **ILINE** et **MARIA**, et mon petit prince **IYED** (les enfants de ma sœur) que Dieu vous protège.*

*A mes chers amis.*

*Je vous dédie ce travail avec tous mes vœux de bonheur, de santé et de réussite.*

## *REMERCIEMENTS*

*Je commence par exprimer ma gratitude envers "ALLAH" le Tout-Puissant, qui m'a béni de la santé, du courage, de la patience et de la détermination indispensables pour mener à bien ce travail.*

*Je tiens tout d'abord à remercier mon encadrant, Professeur YAHIA DJABRANE, pour ses précieux conseils, ses orientations pertinentes jusqu'à la réalisation de ce travail. Aucun mot ne vous rendra justice.*

*Je tiens aussi à remercier les membres de jury, Professeur SAYAH ABDALLAH et Docteur KHEREDDINE SOURAYA, d'avoir accepté d'examiner et d'évaluer ce travail.*

*Je souhaite exprimer également ma gratitude envers tous mes enseignants(es) qui m'ont accompagné tout au long de mon parcours éducatif, surtout les enseignants(es) du département de Mathématiques à l'Université de Mohamed Khider.*

*Je ne peux pas non plus manquer de remercier mes collègues de la promotion 2024, notamment ceux de l'option : Statistique.*

# Table des matières

<b>Remerciements</b>	ii
<b>Table des matières</b>	iii
<b>Table des figures</b>	v
<b>Introduction</b>	1
<b>1 Généralités</b>	<b>3</b>
<b>1.1 Quelques lois usuelles</b>	3
1.1.1 La loi gaussienne ou normale	3
1.1.2 La loi du Khi-deux	8
1.1.3 La loi de Student	10
1.1.4 La loi de Fisher	12
<b>1.2 Les types de convergence</b>	13
1.2.1 Convergence en probabilités	13
1.2.2 Convergence en loi	13
1.2.3 Convergence presque sûre	13
1.2.4 Convergence en moyenne quadratique	14
<b>1.3 Echantillonnage</b>	16
1.3.1 Moyenne empirique	16

1.3.2	Variance empirique	17
1.3.3	Cas particulier d'échantillons gaussiens	18
1.4	Méthode de Box-Muller	19
<b>2</b>	<b>Intervalle de Confiance : cas Gaussien</b>	<b>21</b>
2.1	Intervalle de confiance et fonction pivot	21
2.2	Construction des IC d'une loi mère gaussienne	24
2.2.1	IC pour la moyenne	25
2.2.2	IC pour la variance	30
2.2.3	IC pour la différence des moyennes	33
2.2.4	Intervalle de confiance sur le rapport des variances	36
	<b>Conclusion</b>	<b>38</b>
	<b>Bibliographie</b>	<b>39</b>
	<b>Annexe : Abréviations et Notations</b>	<b>40</b>

# Table des figures

1.1	Densité d'une loi normale : centrée avec différentes variances (panneau de gauche) et réduite avec différentes moyennes (panneau de droite).	7
1.2	Fonction de répartition d'une loi normale : centrée avec différentes variances (panneau de gauche) et réduite avec différentes moyennes (panneau de droite).	8
1.3	Fonction de densité de la loi du Khi-deux avec différents degrés de liberté.	9
1.4	Approximation de la loi de Student, avec différents ddl, par la distribution de la loi normale $\mathcal{N}(0; 1)$ .	11
1.5	Fonction de densité de la loi du Fisher avec différents degrés de liberté.	12
1.6	Algorithme de Box-Muller : Histogrammes de $X$ , (pour $n = 100$ panneau de gauche), (pour $n = 500$ (panneau de droite)).	20
2.1	IC de niveau 95% pour la moyenne dans le cas $\sigma^2$ connue d'une population normale $N(0; 1)$ .	27
2.2	IC $\mu$ en fonction du niveau de confiance $\gamma$ .	29
2.3	IC de niveau $\gamma = 0.95$ pour la moyenne dans le cas $\sigma^2$ inconnue.	30
2.4	IC de niveau $\gamma = 0.95$ pour la variance dans le cas $\mu$ connue.	32
2.5	IC de niveau $\gamma = 0.95$ pour la variance dans le cas $\mu$ inconnue.	33
2.6	IC de niveau $\gamma = 0.95$ pour la différence entre deux moyennes.	36
2.7	IC de niveau $\gamma = 0.95$ pour le rapport entre deux variances.	37

# Introduction

**E**n statistique, l'intervalle de confiance (IC) fournit une estimation de l'incertitude associée à un échantillon, et la loi normale joue un rôle crucial dans la construction de ces intervalles lorsque les conditions d'application sont réunies. L'utilisation conjointe de ces concepts permet de faire des inférences statistiques sur les paramètres de la population à partir des données échantillonnées. De ce point de vue, l'objet principal de ce mémoire concerne ces deux concepts fondamentaux en statistique, souvent utilisés ensemble pour estimer des paramètres inconnus d'une population : L'intervalle de confiance et la loi normale.

Historiquement, la loi normale est marquée par les contributions de nombreux mathématiciens et scientifiques et elle est souvent associée à Carl Friedrich Gauss, un mathématicien et physicien allemand. Gauss a utilisé cette distribution pour analyser des erreurs de mesure dans l'astronomie. Il a montré en (1809) que les erreurs de mesure pouvaient être modélisées par cette distribution, qui a pris son nom en raison de ses contributions significatives. Pierre-Simon Laplace, un mathématicien et statisticien français, a également contribué à l'élaboration de la loi normale. Dans son ouvrage "Théorie analytique des probabilités" (1812), Laplace a utilisé cette distribution dans son analyse des erreurs et a montré que la somme de nombreuses variables aléatoires indépendantes suit une distribution normale, ce qui a plus tard été formalisé dans le théorème central limite.

Au 19ème siècle, la loi normale a été de plus en plus adoptée dans divers domaines de la science et de l'ingénierie pour modéliser des phénomènes naturels et des erreurs de mesure. Son

utilisation s'est étendue à l'économie, à la biologie, à la physique, et à bien d'autres disciplines. Le théorème central limite, formalisé au début du 20ème siècle, a renforcé l'importance de la loi normale. Ce théorème stipule que, sous certaines conditions, la somme de nombreuses variables aléatoires indépendantes et identiquement distribuées tend vers une distribution normale, quel que soit la distribution d'origine de ces variables.

Aujourd'hui, la loi normale est une pierre angulaire de la théorie des probabilités et statistiques. Elle est utilisée pour modéliser une grande variété de phénomènes naturels et humains, et sert de base pour de nombreux tests statistiques et techniques d'inférence. Ce mémoire explore la construction des intervalles de confiance pour la moyenne et la variance dans le cadre d'un échantillon gaussien. Nous utiliserons les propriétés de la distribution normale, l'estimation de ces paramètres et leurs comportement asymptotiques pour établir des intervalles de confiance pour la moyenne, la variance, la différence deux moyennes et le rapport de deux variances d'une population gaussienne.

Le mémoire est organisé en deux chapitres, comme suit : Le premier chapitre est consacré à un rappel de quelques lois usuelles et leurs caractéristiques, ainsi qu'aux propriétés fondamentales de l'échantillonnage relatives à la moyenne et la variance empiriques. Des résultats importants, tels que les lois des grands nombres et le théorème central limite. Puisque la loi normale joue un rôle important dans la détermination des intervalles de confiance, la dernière partie de ce chapitre est réservée aux méthodes de génération de variables aléatoires gaussiennes via la méthode de Box-Muller. Le deuxième chapitre commence par un rappel sur l'intervalle de confiance : définitions, usage et construction simple. Puisqu'en s'intéresse à la moyenne et la variance de la population normale, nous allons étudier en détail le problème de construction des intervalles de confiance de ces deux paramètres, ainsi que la différence et le rapport entre eux dans le cas où la moyenne est connue (respectivement inconnue) et de même pour la variance. Une étude de simulation à l'aide du logiciel de traitement statistique **R** est donnée pour illustrer par des exemples, les résultats théoriques obtenus.



# Chapitre 1

## Généralités

Ce chapitre est consacré à un rappel de quelques lois usuelles et leurs caractéristiques, ainsi qu'aux propriétés fondamentales de l'échantillonnage relatives à la moyenne et la variance empiriques. Des résultats importants, tels que les lois des grands nombres et le théorème central limite, ainsi qu'aux méthode de génération de variables aléatoires gaussiennes.

### 1.1 Quelques lois usuelles

#### 1.1.1 La loi gaussienne ou normale

- **L'origine de la loi :**

L'émergence de ce que l'on nomme aujourd'hui la loi normale est le résultat des travaux de plusieurs scientifiques éminents. Bien que ses prémices remontent aux années 1730, son établissement officiel est souvent attribué au début du 19<sup>e</sup> siècle, grâce aux contributions indépendantes de deux mathématiciens : le Français Pierre-Simon de Laplace et l'Allemand Carl Friedrich Gauss. Laplace a développé son approche en se penchant sur les probabilités, tandis que Gauss a privilégié une approche plus statistique dans son étude de cette loi.

**Définition 1.1.1 (Fonction de densité)** *Soit  $X$  une variable aléatoire (va) de densité  $f$  : On dit que  $X$  suit une loi normale de paramètres réels, d'espérance  $\mu$  et de variance  $\sigma^2$  et*

d'écart type  $\sigma$  ( $\mu \in \mathbb{R}, \sigma > 0$ ), si elle admet la densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \forall x \in \mathbb{R},$$

on note  $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$ .

**Proposition 1.1.1** *La fonction  $f$  définie bien une densité de probabilité car :*

- $\forall x \in \mathbb{R}, f(x) \geq 0$ .
- La fonction  $f$  est continue sur  $\mathbb{R}$ .
- L'intégrale de la fonction de densité  $f(x)$  égale 1, on a :

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

Par le changement de variable

$$t = \left(\frac{x-\mu}{\sigma}\right) \iff x = (t\sigma + \mu) \quad \text{avec } dx = \sigma dt$$

On obtient :

$$\int_{-\infty}^{+\infty} f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt = 1 \quad (\text{car : } \int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi} \text{ intégrale de Gauss})$$

**Définition 1.1.2 (Fonction de répartition)** *On appelle fonction de répartition d'une va  $X$  qui suit la loi normale  $\mathcal{N}(\mu, \sigma^2)$ , la fonction définie par :*

$$\mathbb{F}(x) = \mathbb{P}(Y \leq x) = \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy.$$

**Proposition 1.1.2** *L'espérance et la variance de la loi normale  $\mathcal{N}(\mu, \sigma^2)$  sont respectivement :*

$$\mathbb{E}[X] = \mu \quad \text{et} \quad \mathbb{V}(X) = \sigma^2.$$

**Preuve.** Calculons d'abord l'espérance :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x.f(x)dx = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Posons :  $z = \frac{x-\mu}{\sigma} \implies x = \sigma z + \mu$  et  $dx = \sigma dz$ , alors :

$$\begin{aligned} \mathbb{E}[X] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma z + \mu) e^{-\frac{z^2}{2}} \sigma dz \\ &= \frac{1}{\sqrt{2\pi}} \sigma \int_{-\infty}^{+\infty} z e^{-\frac{z^2}{2}} dz + \frac{\mu}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz}_{=\sqrt{2\pi}} = 0 + \mu = \mu. \end{aligned}$$

La variance de  $X$  s'écrit :

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

et on a :

$$\mathbb{E}(X^2) = \int_{-\infty}^{+\infty} x^2.f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Par changement de variable  $z = \frac{x-\mu}{\sigma}$ , nous trouvons :

$$\begin{aligned} \mathbb{E}(X^2) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma z + \mu)^2 e^{-\frac{z^2}{2}} \sigma dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma^2 z^2 e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mu^2 e^{-\frac{z^2}{2}} dz + \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \sigma z \mu e^{-\frac{z^2}{2}} dz \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-\frac{z^2}{2}} dz + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz + \frac{\sigma\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z e^{-\frac{z^2}{2}} dz \\ &= \sigma^2 + \mu^2. \end{aligned}$$

Finalement on trouve

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \sigma^2 + \mu^2 - (\mu)^2 = \sigma^2$$

Passons ensuite à la fonction caractéristique et à la fonction génératrice des moments d'une loi normale  $\mathcal{N}(\mu, \sigma^2)$ . ■

**Proposition 1.1.3 (Fonction caractéristique)** Soit  $X$  une va de loi  $\mathcal{N}(\mu, \sigma^2)$ , la fonction caractéristique de loi normale notée  $\varphi_X(t)$ , définie sur  $\mathbb{R}$  à valeurs complexes par :

$$\varphi_X(t) = \exp(it\mu) \exp\left(-\frac{t^2\sigma^2}{2}\right).$$

**Preuve.** On a :

$$\varphi_X(t) = \mathbb{E}[e^{it\mu}] = \int_{-\infty}^{+\infty} e^{itx} f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Posons  $l = \frac{x-\mu}{\sigma}$ , alors  $x = \sigma l + \mu$  et  $dx = \sigma dl$ , donc :

$$\begin{aligned} \varphi_X(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{l^2}{2}} e^{it(\sigma l + \mu)} \sigma dl = \frac{e^{it\mu}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(l-\sigma t)^2} e^{-\frac{1}{2}\sigma^2 t^2} dl \\ &= \frac{e^{it\mu} e^{-\frac{1}{2}\sigma^2 t^2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(l-\sigma t)^2} dl = e^{it\mu} e^{-\frac{1}{2}\sigma^2 t^2}. \end{aligned}$$

■

**Proposition 1.1.4 (Fonction génératrice des moments)** La fonction génératrice des moments d'une loi  $\mathcal{N}(\mu, \sigma^2)$  est donnée par :

$$g(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

**Preuve.** Par la définition de la fonction génératrice des moments, on a :

$$\begin{aligned} g(t) &= \mathbb{E}[e^{tx}] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2} + tx} dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2} + t(x-\mu+\mu)} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t\mu} e^{-\frac{(x-\mu)^2}{2\sigma^2} + t(x-\mu)} dx = \frac{e^{t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2} + t(x-\mu)} dx \end{aligned}$$

Posons  $z = x - \mu$ , on obtient :

$$\begin{aligned} g(t) &= \frac{e^{t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2\sigma^2} + tz} dz = \frac{e^{t\mu}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}(z^2 - 2\sigma^2 z + (\sigma^2 z)^2 - (z\sigma^2))} dz \\ &= \frac{e^{t\mu} e^{\frac{\sigma^2 z^2}{2}}}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}(z - \sigma^2 t)^2} dz = e^{t\mu} e^{\frac{\sigma^2 z^2}{2}}. \end{aligned}$$

■

**Définition 1.1.3 (Loi normale standard)** *La loi normale «standard» ou loi «de référence» est un cas particulier de la loi gaussienne de paramètre  $\mu = 0$  et  $\sigma^2 = 1$ , c'est ce qu'on appelle la loi normale centré-réduite  $\mathcal{N}(0; 1)$ . Elle admet la fonction densité défini par :*

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

et en déduit par transformée : si  $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2) \implies Z = \frac{X - \mu}{\sigma} \rightsquigarrow \mathcal{N}(0; 1)$ .

Nous donnons dans la sortie graphique [1.1](#) suivante, l'évolution des courbes de la fonction de densité d'une va  $\mathcal{N}(\mu, \sigma^2)$ , pour différentes valeurs des paramètres  $\mu$  et  $\sigma^2$ .

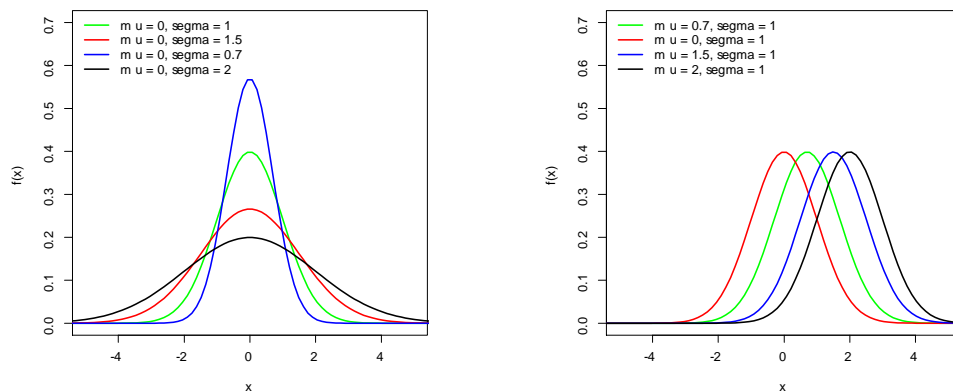


FIG. 1.1 – Densité d'une loi normale : centrée avec différentes variances (panneau de gauche) et réduite avec différentes moyennes (panneau de droite).

**Remarque 1.1.1** *La fonction de répartition d'une va  $Z$  de loi normale standard, définie*

par :

$$\phi(x) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz.$$

La figure [1.2](#) suivante, présente l'évolution des courbes de la fonction de distribution d'une va  $\mathcal{N}(\mu, \sigma^2)$ , pour différentes valeurs des paramètres  $\mu$  et  $\sigma^2$ .

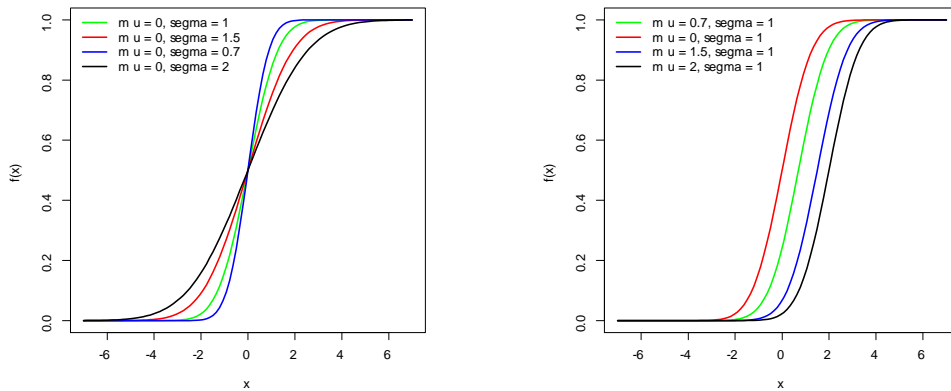


FIG. 1.2 – Fonction de répartition d'une loi normale : centrée avec différentes variances (panneau de gauche) et réduite avec différentes moyennes (panneau de droite).

**Remarque 1.1.2** La fonction caractéristique et la fonction génératrice d'une v.a  $X$  de la loi normale standard sont données respectivement par :

$$\varphi(x) = e^{-\frac{x^2}{2}}, \quad g(t) = e^{\frac{t^2}{2}}, \quad t \in \mathbb{R}.$$

**Proposition 1.1.5** Toute combinaison linéaire de va gaussiennes indépendantes est une va gaussienne.

**Preuve.** Pour la démonstration voir [\[5\]](#), page 54]. ■

## 1.1.2 La loi du Khi-deux

**Définition 1.1.4** Soit  $Z_1, Z_2, \dots, Z_n$  une suite de va indépendantes et identiquement distribué (i.i.d) de loi  $\mathcal{N}(0,1)$ . Alors, la va  $\sum_{i=1}^n Z_i^2$  suit une loi appelée loi du khi-deux à  $n$  degrés de liberté (d.d.l), notée  $\chi^2(n)$ .

**Propriétés 1.1.1** Si  $X$  suit une loi de  $\chi^2(n)$ , alors

1. sa fonction caractéristique est  $(1 - 2it)^{-n/2}$ .

2. sa densité est :

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \text{ pour } x > 0 \\ 0 \text{ sinon.} \end{cases}$$

où  $\Gamma$  est la fonction Gamma d'Euler définie par :  $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$ .

3. l'espérance de  $X$  est égale au nombre  $n$  d.d.l et sa variance est  $2n$ .

4. La somme de deux va indépendantes qui suivent respectivement  $\chi^2(n_1)$  et  $\chi^2(n_2)$  suit aussi une loi du  $\chi^2$  avec  $n_1 + n_2$  d.d.l.

Le graphe [1.3](#) suivant, donne une illustration de la fonction de densité de la loi du Khi-deux pour différentes valeurs du degré de liberté.

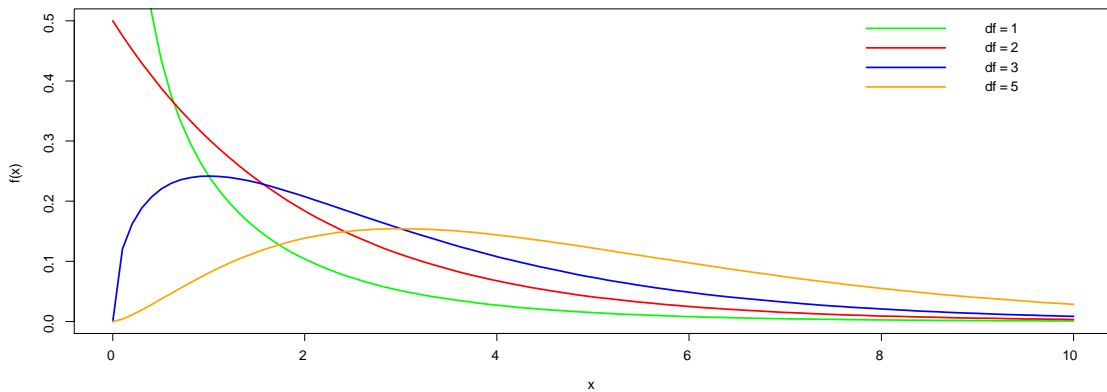


FIG. 1.3 – Fonction de densité de la loi du Khi-deux avec différents degrés de liberté.

**Théorème 1.1.1** Soit un  $n$ -échantillon  $X_1, X_2, \dots, X_n$  de loi  $\mathcal{N}(\mu, \sigma^2)$ , on a :

$$\frac{(n-1)}{\sigma^2} S^2 = \frac{(n-1)}{\sigma^2} \frac{1}{(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2 \rightsquigarrow \chi^2(n-1),$$

avec  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  et  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  sont respectivement la moyenne empirique et la variance empirique corrigée.

**Preuve.** On a :

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

Soit

$$\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} + \left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2.$$

Les deux termes de droite sont respectivement  $\frac{(n-1)S^2}{\sigma^2}$  et  $\left( \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2$  qui est une gaussienne centrée réduite. Ces termes aléatoires étant indépendants (car si la loi mère est gaussienne,  $\bar{X}$  et  $S^2$  sont des v.a indépendantes) et les v.a  $\frac{X_i - \mu}{\sigma}$  étant indépendantes de loi  $\mathcal{N}(0; 1)$ , en termes de fonctions caractéristique, on a :

$$\left( \frac{1}{1 - 2it} \right)^{\frac{n}{2}} = \varphi_{\frac{(n-1)S^2}{\sigma^2}}(t) \cdot \left( \frac{1}{1 - 2it} \right)^{1/2} \quad (\text{si } t < \frac{1}{2})$$

Donc,

$$\varphi_{\frac{(n-1)S^2}{\sigma^2}}(t) = \left( \frac{1}{1 - 2it} \right)^{\frac{n-1}{2}}$$

Sachant que l'espérance d'une loi  $\chi^2(n-1)$  est  $n-1$  et sa variance  $2(n-1)$ , on voit que :

$$\mathbb{E}(S^2) = \sigma^2 \quad \text{et} \quad \mathbb{V}(S^2) = \frac{\sigma^2}{n-1}.$$

Ce qui prouve le théorème. ■

### 1.1.3 La loi de Student

**Définition 1.1.5** Soit  $Z$  et  $Q$  deux va indépendantes telles que  $Z \rightsquigarrow \mathcal{N}(0, 1)$  et  $Q \rightsquigarrow \chi^2(n)$ , alors la va

$$\frac{Z}{\sqrt{\frac{Q}{n}}}$$

suit une loi appelée loi de student à  $n$  d.d.l, notée  $t(n)$ .

**Propriétés 1.1.2** Pour la loi de Student à  $n$  degrés de liberté, on a



1. La densité est :

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}$$

2. L'espérance n'est pas définie pour  $n = 1$  et vaut 0 si  $n \geq 2$ . Sa variance n'existe pas pour  $n \leq 2$  et vaut  $n/(n - 2)$  pour  $n \geq 3$ .

Le graphe [1.4](#) suivant, donne une illustration de la fonction de densité de la loi du Student pour différentes valeurs du degré de liberté.

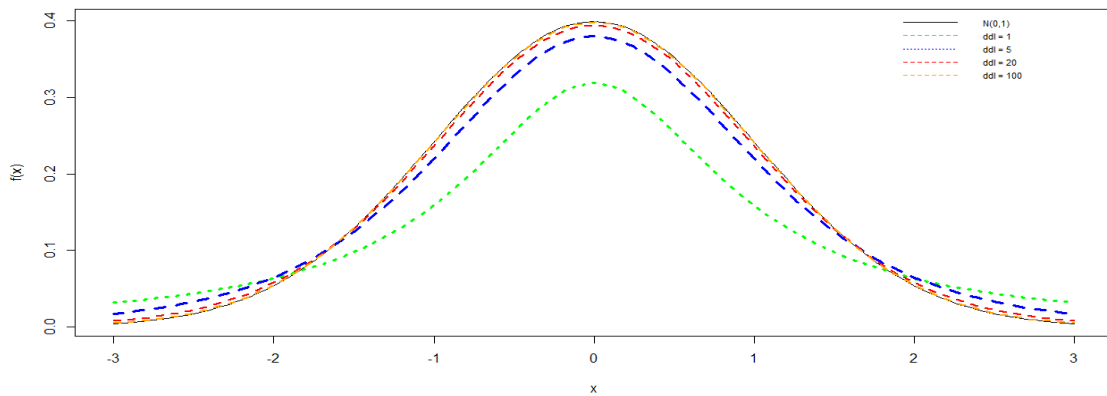


FIG. 1.4 – Approximation de la loi de Student, avec différents ddl, par la distribution de la loi normale  $\mathcal{N}(0; 1)$ .

**Remarque 1.1.3** Pour  $n = 1$ , la loi de Student s'appelle loi de Cauchy.

**Théorème 1.1.2** Soit  $X_1, X_2, \dots, X_n$  un  $n$ -échantillon de loi mère  $\mathcal{N}(\mu, \sigma^2)$ . Alors :

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \rightsquigarrow t(n - 1)$$

**Preuve.** Il suffit d'appliquer la définition de la loi de student aux va's :

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad \text{et} \quad Q = \frac{(n - 1)S^2}{\sigma^2}.$$

les calculs sont admis. ■

**Remarque 1.1.4** Les applications de la loi de student se rencontrent souvent en statistique dès lors qu'on est appelé à remplacer  $\sigma$ , en général inconnu par son «estimateur» naturel  $S$ .

### 1.1.4 La loi de Fisher

**Définition 1.1.6** Soit  $Q_1$  et  $Q_2$  deux v.a indépendantes telles que :  $Q_1 \rightsquigarrow \chi^2(n_1)$  et  $Q_2 \rightsquigarrow \chi^2(n_2)$ , alors la va

$$F = \frac{Q_1/n_1}{Q_2/n_2}$$

suit une loi de Fisher-Snedecor à  $(n_1, n_2)$  d.d.l, on note  $F \rightsquigarrow F(n_1, n_2)$ .

**Propriété 1.1.1** La fonction de densité de la loi  $F(n_1, n_2)$  est :

$$f(x) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} \left(\frac{n_1}{n_2}\right)^{\frac{n_1}{n_2}} \frac{x^{\frac{n_1-2}{2}}}{(1 + \frac{n_1}{n_2}x)^{\frac{n_1+n_2}{2}}} \quad \text{si } x > 0 \quad (0, \quad \text{sinon})$$

Son espérance existe si  $n_2 \geq 3$  et est égale à  $\frac{n_2}{n_2-2}$ . Sa variance existe si  $n_2 \geq 5$  et est égale à  $\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ .

**Remarque 1.1.5** Si  $H$  suit une loi de Fisher  $F(n_1, n_2)$ , alors  $\frac{1}{H} \rightsquigarrow F(n_2, n_1)$ .

Si de plus,  $T$  suit une loi de Student à  $n$  d.d.l, alors  $T^2 \rightsquigarrow F(1, n)$ .

Le graphe [1.5](#) suivant, donne une illustration de la fonction de densité de la loi du Fisher avec différentes valeurs du degrés de libertés.

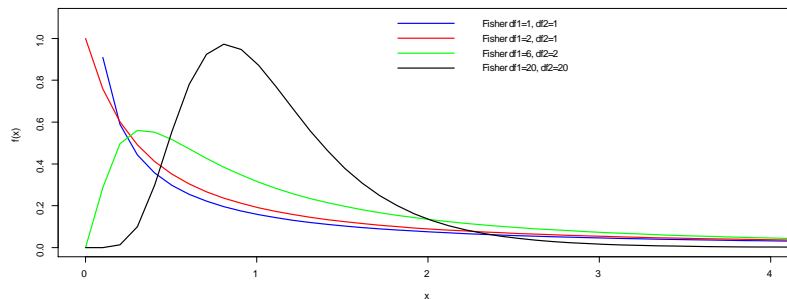


FIG. 1.5 – Fonction de densité de la loi du Fisher avec différents degrés de liberté.

## 1.2 Les types de convergence

On considère une suite  $(X_n)_{n \geq 1}$  de v.a's, étant une suite de fonction de  $\Omega$  dans  $\mathbb{R}$ , il existe diverses façon de définir la convergence de  $(X_n)_{n \geq 1}$  dont certaines jouent un grand rôle en calcul des probabilités et dans l'estimation et l'inférence statistique.

### 1.2.1 Convergence en probabilités

**Définition 1.2.1** *On dit que la suite de va  $(X_n)_{n \geq 1}$  converge en probabilité vers une va  $X$ , si pour tout  $\epsilon > 0$  :*

$$P(|X_n - X| < \epsilon) \xrightarrow{n \rightarrow \infty} 1$$

où, de façon équivalente :

$$P(|X_n - X| > \epsilon) \xrightarrow{n \rightarrow \infty} 0$$

et on écrit :  $X_n \xrightarrow{\mathcal{P}} X$ .

### 1.2.2 Convergence en loi

**Définition 1.2.2** *On dit que la suite de va  $(X_n)_{n \geq 1}$  de fonction de répartition  $F_n$ , converge en loi vers une va  $X$  de fonction de répartition  $F$ , si :*

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

on écrit alors :  $X_n \xrightarrow{\mathcal{L}} X$ .

### 1.2.3 Convergence presque sûre

**Définition 1.2.3** *On dit que la suite de v.a  $(X_n)_{n \geq 1}$  converge presque sûrement vers une v.a  $X$ , si :*

$$P(\lim_{n \rightarrow \infty} X_n \neq X) = 0$$

et on écrit :  $X_n \xrightarrow{p.s.} X$ .

## 1.2.4 Convergence en moyenne quadratique

**Définition 1.2.4** On dit que la suite de v.a  $(X_n)_{n \geq 1}$  converge en moyenne ou d'ordre  $p$  vers la va  $X$ , avec  $0 < p < \infty$ , si :

$$\mathbb{E} [|X_n - X|^p] \xrightarrow{n \rightarrow \infty} 0$$

et on écrit pour  $p = 2$  (convergence en moyenne quadratique) :  $X_n \xrightarrow{\mathcal{M}q} X$ .

**Remarque 1.2.1** Les implications suivantes permettent le passage entre certains types de convergence :

$$\begin{aligned} X_n \xrightarrow[n \rightarrow \infty]{\mathcal{M}q} X &\implies X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} X \\ X_n \xrightarrow[n \rightarrow \infty]{p.s.} X &\implies X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} X \\ X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} X &\implies X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X. \end{aligned}$$

**Remarque 1.2.2** Les relations entre les types de convergences précédentes sont vrais, si on remplace la va  $X$  par une constante réelle  $a$ .

**Théorème 1.2.1 (Inégalité de Markov)** Soit  $Y$  une va réelle,  $g$  une fonction croissante et positive ou nulle sur l'ensemble des réels, vérifiant  $g(a) > 0$ , alors :

$$\forall a > 0, \quad P(Y \geq a) \leq \frac{\mathbb{E}(g(Y))}{g(a)}.$$

**Preuve.** On a :

$$\begin{aligned} \mathbb{E}(g(Y)) &= \int_{\Omega} g(y)f(y)dy = \int_{Y < a} g(y)f(y)dy + \int_{Y \geq a} g(y)f(y)dy \\ &\geq \int_{Y \geq a} g(y)f(y)dy \quad (\text{car } g \text{ est positive ou nulle}) \\ &\geq g(a) \int_{Y \geq a} f(y)dy \quad (\text{car } g \text{ est croissante}) \\ &= g(a)P(Y \geq a). \end{aligned}$$

Il est clair que :  $\mathbb{E}(g(Y)) \geq g(a)P(Y \geq a)$ . ■

**Théorème 1.2.2 (Inégalité de Bienaymé-Tchebychev)** Soit  $X$  une va admettant une espérance  $\mathbb{E}(X)$  et de variance finie  $\sigma^2$ , alors pour tout réel  $\epsilon$  strictement positif, on a :

$$P(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

**Preuve.** Pour montrer l'inégalité de Bienaymé-Tchebychev, il suffit de prendre  $Y = |X - \mathbb{E}(X)|$ ,  $a = \epsilon$  et  $g(t) = t^2$  dans l'inégalité de Markov. ■

**Théorème 1.2.3 (Loi faible des grand nombres)** Soit  $(X_n)$  une suite de va iid d'espérance et de variance finies, alors :

$$X_n \xrightarrow[n \rightarrow \infty]{\mathcal{P}} E(X) = \mu.$$

**Théorème 1.2.4 (Loi forte des grand nombres)** Soit  $(X_n)$  une suite de va iid d'espérance et de variance finies, alors :

$$X_n \xrightarrow[n \rightarrow \infty]{p.s} E(X) = \mu.$$

**Théorème 1.2.5 (Centrale limite)** Soit  $(X_n)$  une suite de va i.i.d admettant une moyenne  $\mu$  et une variance  $\sigma^2$  finies. Alors, la suite  $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$  converge en loi vers la va de loi  $\mathcal{N}(0, 1)$ , et

en écrit :

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

**Preuve.** La démonstration est basée sur les fonctions caractéristiques. Pour les détails, voir [5, page 83] et [2, page 14]. ■

## 1.3 Echantillonnage

**Définition 1.3.1** Soit  $X$  une va sur un référentiel  $\Omega$ . Un échantillon de  $X$  de taille  $n$  est un  $n$ -uplet  $(X_1, \dots, X_n)$  de va's indépendantes de même loi que  $X$ . La loi de  $X$  sera appelée loi mère. Une réalisation de cet échantillon est un  $n$ -uplet de réels  $(x_1, \dots, x_n)$  où  $X_i(\omega) = x_i$ .

### 1.3.1 Moyenne empirique

**Définition 1.3.2** On appelle moyenne de l'échantillon ou **moyenne empirique**, la statistique notée  $\bar{X}$  définie par :

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

**Proposition 1.3.1** Soit  $\mu$  et  $\sigma^2$ , respectivement la moyenne et la variance de la v.a  $X$  de la loi mère. On a :

$$\mathbb{E}(\bar{X}) = \mu \quad \text{et} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

En effet,

$$\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \mu.$$

Puis, en raison de l'indépendance des  $X_i$  :

$$\mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

**Remarque 1.3.1** 1. Par le théorème centrale limite  $\bar{X}$  converge en loi vers  $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$  lorsque  $n$  tend vers l'infini.

2. Les moments centrés d'ordre 3 et 4 de  $X$  sont respectivement :

$$\mu_3(\bar{X}) = \frac{\mu_3}{n^2} \quad \text{et} \quad \mu_4(\bar{X}) = \frac{\mu_4 + 3\sigma^2(n-1)}{n^3}.$$

### 1.3.2 Variance empirique

**Définition 1.3.3** On appelle *variance empirique*, la statistique notée  $S_n^2$  définie par :

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (1.1)$$

**Proposition 1.3.2** L'espérance et la variance de la variance empirique sont :

$$\mathbb{E}(S_n^2) = \frac{n-1}{n} \sigma^2 \quad \text{et} \quad \mathbb{V}(S_n^2) = \frac{n-1}{n^3} [(n-1)\mu_4 - (n-3)\sigma^4],$$

$\mu_k$  est le moment centré d'ordre  $k$ .

**Preuve.** Supposons que  $\mathbb{E}(X) = \mu$ , on a alors :

$$\begin{aligned} \mathbb{E}(S_n^2) &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i - \mu + \mu - \bar{X})^2 \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i - \mu)^2 - \mathbb{E}(\bar{X} - \mu)^2 \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{V}ar(X_i) - \mathbb{V}ar(\bar{X}) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2. \end{aligned}$$

Pour la variance  $\mathbb{V}(S_n^2)$ , voir [9] ou [8]. ■

**Définition 1.3.4** On appelle *variance empirique corrigée* la statistique :

$$S^{*2} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

**Remarque 1.3.2** 1) L'espérance de la variance corrigée est égale  $\sigma^2$ . Cependant l'écart-type

corrigé reste biaisé pour  $\sigma$  car :

$$\mathbb{E}(\sqrt{S^{*2}}) \neq \sqrt{\mathbb{E}(S^{*2})},$$

mais il est asymptotiquement sans biais.

2) La variance empirique lorsque  $n$  tend vers l'infini est :  $V(S_n^2) \xrightarrow[n \rightarrow \infty]{} \frac{\mu_4 - \sigma^4}{n}$ , et d'après le

Théorème [1.2.5](#) :

$$\frac{S_n^2 - \frac{n-1}{n}\sigma^2}{\sqrt{V(S_n^2)}} \xrightarrow{\mathcal{L}} N(0; 1).$$

Par approximation, on trouve :

$$\sqrt{n} \frac{S_n^2 - \sigma^2}{\sqrt{\mu_4 - \sigma^4}} \xrightarrow{\mathcal{L}} \mathcal{N}(0; 1).$$

**Théorème 1.3.1** La covariance entre  $\bar{X}$  et  $S_n^2$  est :

$$\text{Cov}(\bar{X}, S_n^2) = \frac{n-1}{n^2} \mu_3. \tag{1.2}$$

**Preuve.** Pour la démonstration voir [\[9\]](#), page 281. ■

### 1.3.3 Cas particulier d'échantillons gaussiens

– Si la loi mère est gaussienne alors  $\bar{X}$  est gaussienne, en tant que combinaison linéaire de gaussiennes indépendantes d'après la Proposition [1.1.5](#). Alors,

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

– La moyenne et la variance empiriques sont indépendantes. En effet, d'après [\(1.2\)](#)

$$\text{Cov}(\bar{X}, S_n^2) = \frac{n-1}{n^2} \mu_3 = 0, \quad \text{car } \mu_3 = 0 \quad (\text{dans le cas gaussien}).$$



– La variance empirique vérifie :

$$\frac{nS_n^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2.$$

pour plus d'informations et détails voir [9], page 282] et [8], page 54].

## 1.4 Méthode de Box-Muller

Dans cette dernière section, nous présentons une méthode de génération de va's gaussiennes. La méthode Box-Muller est introduite par les statisticiens G.E.P. Box et M.E. Muller en 1958. Cette méthode permet de générer des v.a normales à distribution centrée réduite :

Soient  $U$  et  $V$  deux v.a i.i.d de loi uniforme  $\mathcal{U}[0, 1]$ . Alors  $X$  et  $Y$  définies par :

$$\begin{cases} X = \cos(2\pi U)(-2 \log V)^{\frac{1}{2}} \\ Y = \sin(2\pi U)(-2 \log V)^{\frac{1}{2}} \end{cases}$$

sont deux va's normales centrées réduites et indépendantes.

Maintenant nous essayons de donner un argument du fait que  $X$  et  $Y$  soient de loi normale, mais ce n'est pas une preuve rigoureuse. Posons  $\rho^2 = X^2 + Y^2$ , tel que  $\rho^2 \rightsquigarrow \chi^2(2)$ , et  $\theta = \arctan\left(\frac{Y}{X}\right)$  est l'angle formé par le vecteur  $(X, Y)'$  de la loi uniforme sur  $[0, 2\pi]$ .

La construction de Box-Muller permet de retrouver ces propriétés. Alors on a :

$$\arctan\left(\frac{Y}{X}\right) = \arctan\left(\frac{\sin(2\pi U)(-2 \log V)^{\frac{1}{2}}}{\cos(2\pi U)(-2 \log V)^{\frac{1}{2}}}\right) = \arctan(\tan(2\pi U)) = 2\pi U.$$

Comme  $U$  est uniforme sur  $\mathcal{U}[0, 1]$ , et  $2\pi U$  est uniforme sur  $[0, 2\pi]$ . D'autre part, on a :

$$\rho^2 = X^2 + Y^2 = (\cos^2(2\pi U) + \sin^2(2\pi U))(-2 \log V) = -2 \log V$$

Donc il reste à déterminer la loi de  $-2 \log V$  :

$$\begin{aligned} P(-2 \log V \leq x) &= P\left(\log V \geq \frac{-x}{2}\right) = P\left(V \geq \exp\left(\frac{-x}{2}\right)\right) \\ &= 1 - P\left(V \leq \exp\left(\frac{-x}{2}\right)\right) = 1 - \exp\left(\frac{-x}{2}\right) \quad (\text{car } V \rightsquigarrow \mathcal{U}[0, 1]). \end{aligned}$$

Ce qui correspond à la fonction de répartition de la loi du Khi deux à 2 d.d.l.

Comme application, par l'algorithme de Box-Muller nous avons réussi à simuler et générer des données distribuées selon une loi normale  $\mathcal{N}(0, 1)$  à l'aide du logiciel R, pour deux tailles d'échantillon distincte ( $n = 100$  et  $n = 500$ ). Dans la figure [1.6](#) suivante, nous donnons l'allure ressemblant à la courbe de Gauss des histogrammes des données simulés par l'algorithme de Box-Muller, pour  $n = 100$  (panneau de gauche) et pour  $n = 500$  (panneau de droite).

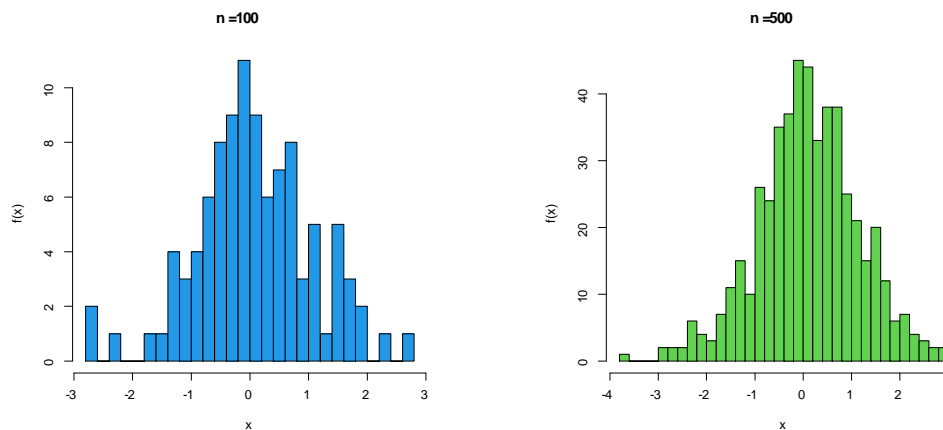


FIG. 1.6 – Algorithme de Box-Muller : Histogrammes de  $X$ , (pour  $n = 100$  panneau de gauche), (pour  $n = 500$  (panneau de droite)).

# Chapitre 2

## Intervalle de Confiance : cas Gaussien

L'intervalle de confiance (IC) est une plage de valeurs probable pour un paramètre statistique, calculé à partir d'un échantillon de données. En utilisant des méthodes telles que les fonctions pivots, les lois limites et les distributions connues, on peut construire ces intervalles pour divers paramètres, tels que la moyenne ou la variance, offrant ainsi une mesure de fiabilité de l'estimation.

Dans ce chapitre, nous allons étudier en détail le problème de construction des intervalles de confiance dans le cas gaussien de la moyenne et la variance, ainsi que la différence et le rapport entre eux dans le cas où la moyenne est connue (respectivement inconnue) et de même pour la variance. Une étude de simulation à l'aide du logiciel de traitement statistique **R** est donnée pour illustrer par des exemples, les résultats théoriques obtenus.

### 2.1 Intervalle de confiance et fonction pivot

**Définition 2.1.1** Soit  $X_1, X_2, \dots, X_n$  un échantillon issu d'une va  $X$  de loi de densité (ou fonction de probabilité)  $f(x; \theta)$  où  $\theta \in \Theta$  est un paramètre inconnu de dimension 1. On appelle procédure d'intervalle de confiance de niveau  $\gamma \in (0, 1)$  tout couple de statistiques  $(T_1, T_2)$  tel que  $\forall \theta \in \Theta$  :

$$P_{\theta}(T_1 \leq \theta \leq T_2) \geq \gamma.$$

Pratiquement, on choisira  $\gamma$  assez élevé : couramment  $\gamma = 0.95$  (95%). Ainsi, il y a une forte probabilité pour que l'intervalle à bornes aléatoires  $[T_1, T_2]$  contienne la vraie valeur de  $\theta$ . Autrement dit, dans l'univers des échantillons possibles d'une proportion, au moins  $\gamma$  d'entre eux, on obtient un intervalle qui contient  $\theta$ .

**Remarque 2.1.1** *Dans certaines situations, on peut n'intéresser qu'à établir une borne inférieure ou une borne supérieure pour  $\theta$ ,  $T_1$  ou  $T_2$  étant rejeté à l'infini. On parle alors d'intervalle de confiance unilatéral (par opposition à «bilatéral»).*

**Exemple 2.1.1** *Considérons l'exemple artificiel d'une loi mère gaussienne avec une variance connue supposée égale à 1. On a*

$$\frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0; 1)$$

d'où :

$$P(-1,96 < \frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}}} < 1,96) = P(\frac{-1,96}{\sqrt{n}} < \bar{X} - \mu < \frac{1,96}{\sqrt{n}}) = 0.95.$$

1.96 est le quantile d'ordre 97.5% de la loi  $N(0; 1)$ . Donc,

$$P(\bar{X} - \frac{1.96}{\sqrt{n}} < \mu < \bar{X} + \frac{1.96}{\sqrt{n}}) = 0.95$$

d'où, à 95% :

$$\bar{X} - \frac{1.96}{\sqrt{n}} < \mu < \bar{X} + \frac{1.96}{\sqrt{n}}$$

et ceci  $\forall \mu$  ce qui prouve que  $[\bar{X} - 1.96/\sqrt{n}; \bar{X} + 1.96/\sqrt{n}]$  constitue une procédure d'intervalle de confiance (IC) de niveau 95%. On voit sur cet exemple que la largeur de l'intervalle est proportionnelle à l'écart type  $1/\sqrt{n}$  de l'estimateur ponctuel  $\bar{X}$  pour  $\mu$ .

**Remarque 2.1.2** *Pour les cas continus, comme dans l'exemple précédent, on peut espérer atteindre exactement le niveau  $\gamma$  que l'on s'est fixé, du fait de la continuité des fonctions de répartition. Pour les cas discrets, cependant un niveau de probabilité donné peut ne pas être atteint en raison des sauts de discontinuité. Nous donnerons plus loin un exemple illustrant cela. On se devra alors d'avoir une attitude conservatrice, c'est-à-dire d'utiliser une procédure*

garantissant que  $[T_1, T_2]$  ait une probabilité de couvrir  $\theta$  qui soit au moins égale au niveau nominale  $\gamma$ . C'est pourquoi il est nécessaire que  $\gamma$  apparaisse comme une borne minimale de probabilité dans la Définition 2.1.1. Notons encore au vu de l'exemple 2.1.1, que le choix de l'intervalle n'est pas unique. On aurait également pu prendre :

$$P \left( z_{0.03} < \frac{\bar{X} - \mu}{\frac{1}{\sqrt{n}}} < z_{0.98} \right) = 0.95$$

comme point de départ ou tout autre couple de quantiles  $(z_\alpha, z_{0.95+\alpha})$  avec  $\alpha \in [0; 0.05]$ .

Dans le contexte de la Définition 2.1.1, soit  $x_1, x_2, \dots, x_n$  une réalisation de  $X_1, X_2, \dots, X_n$  conduisant à la réalisation  $(t_1, t_2)$  de  $(T_1, T_2)$ . Alors l'intervalle  $[t_1, t_2]$  est appelé intervalle de confiance de niveau  $\gamma$  pour  $\theta$  et l'on note :

$$IC_\gamma(\theta) = [t_1, t_2].$$

l'intervalle de confiance est donc l'application numérique de la procédure suite à la réalisation de l'échantillon.

**Remarque 2.1.3** 1. Lorsque l'intervalle est symétrique par rapport à l'estimation ponctuelle on pourra aussi noter comme pour l'application ci-dessus :

$$IC_{0.95}(\mu) = \bar{X} \pm \frac{1.96}{\sqrt{n}}.$$

2. On remarquera que le fait d'augmenter le niveau de confiance accroît la largeur de l'intervalle et qu'il n'est pas possible de donner un intervalle certain autre que  $\Theta$  dans sa totalité.
3. S'il s'agit d'estimer une fonction  $h(\theta)$  bijective, par exemple strictement croissante du paramètre  $\theta$ , il suffit de prendre l'intervalle  $[h(t_1), h(t_2)]$ .

**Définition 2.1.2** Une fonction  $g(X_1, X_2, \dots, X_n; \theta)$  est appelée fonction pivot si :

1. la loi de  $g(X_1, X_2, \dots, X_n; \theta)$  est connue et ne dépend pas de  $\theta$ .

2. pour tous réels  $u_1$  et  $u_2$  tels que  $u_1 \leq u_2$  et tout  $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ , la double inégalité :

$$u_1 \leq g(X_1, X_2, \dots, X_n; \theta) \leq u_2$$

peut se résoudre (ou «pivoter») en  $\theta$  selon :

$$t_1(x_1, x_2, \dots, x_n) \leq \theta \leq t_2(x_1, x_2, \dots, x_n)$$

Dans l'exemple 2.1.1 la va  $\frac{\bar{X} - \mu}{1/\sqrt{n}}$  était une fonction pivot car pour toute valeur  $\bar{X}$  on peut résoudre l'inégalité :

$$u_1 < \frac{\bar{X} - \mu}{1/\sqrt{n}} < u_2 \Rightarrow \bar{x} - \frac{u_2}{\sqrt{n}} < \mu < \bar{x} - \frac{u_1}{\sqrt{n}}.$$

Notons que dans cette dernière définition, on peut évidemment se restreindre aux valeurs  $(x_1, x_2, \dots, x_n)$  appartenent à l'ensemble des réalisations possibles pour  $\theta$  quelconque. Remarquons aussi qu'une fonction pivot n'est pas une statistique car elle contient le paramètre inconnu  $\theta$ .

## 2.2 Construction des IC d'une loi mère gaussienne

Nous étudions dans la suite, le problème de construction des IC de la moyenne et de la variance d'une loi mère gaussienne. Nous verrons que le résultat pour la moyenne d'une gaussienne peut servir d'approximation pour une loi mère quelconque. Nous rencontrerons également des situations nouvelles de comparaisons entre deux lois ou en pratique deux populations distinctes. Pour simplifier les écritures nous prendrons comme c'est l'usage des IC le niveau de confiance  $\gamma = 0.95$  faisant donc intervenir les quantiles d'ordre 0.025 et 0.975, le passage à une autre valeur de  $\gamma$  étant évident.

## 2.2.1 IC pour la moyenne

### • IC pour la moyenne dans le cas $\sigma^2$ inconnue :

Nous abordons d'emblée le cas où  $(\mu, \sigma^2)$  est un paramètre de dimension 2 inconnu, mais nous intéresserons ici uniquement à un encadrement pour  $\mu$  indépendamment de  $\sigma^2$ . nous reviendrons ensuite brièvement sur le cas plus simple mais peu réaliste où  $\sigma^2$  est supposé connu. Rappelons le résultat du (1.1.2) :

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \underset{\text{approx}}{\rightsquigarrow} t(n-1).$$

Cette va est de toute évidence une fonction pivot pour  $\mu$  et nous obtenons un IC comme suit, les développements étant de même nature que dans l'exemple 2.1.1 :

$$P\left(-t_{1-\frac{\alpha}{2}}(n-1) < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{1-\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha = \gamma$$

$$P\left(-t_{0.975}(n-1) < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{0.975}(n-1)\right) = 0.95 \quad (\alpha = 5\%)$$

ou  $t_\alpha(n-1)$  est la notation adoptée pour le quantile d'ordre  $\alpha$  de la loi de student à  $n-1$  d.d.l, dont nous rappelons que comme la loi de gauss elle est symétrique par rapport à 0. Il s'ensuit que :

$$P\left(\bar{X} - t_{0.975}^{(n-1)} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{0.975}^{(n-1)} \frac{S}{\sqrt{n}}\right) = 0.95$$

et le résultat très classique :

$$IC_{0.95}(\mu) = \left[ \bar{X} - t_{0.975}^{(n-1)} \frac{S}{\sqrt{n}}, \bar{X} + t_{0.975}^{(n-1)} \frac{S}{\sqrt{n}} \right] \quad (2.1)$$

Notons que les quantiles des lois de student sont donnés dans toutes les tables statistiques usuelles.

**Remarque 2.2.1** La largeur de cet IC dont on peut montrer qu'elle est minimale par rapport à d'autres éventuelles procédures, dépend d'une part de la taille d'échantillon et d'autre part

de la dispersion même de la loi mère à travers l'estimation  $S$  de son écart type  $\sigma$ . Plus la population est homogène et plus la taille d'échantillon est élevée, plus l'estimation sera précise.

**Remarque 2.2.2** Les praticiens utilisent cette formule sans se soucier de la «normalité» de la loi mère. Ceci est de fait justifié d'une part grâce au théorème central limite qui assure avec des conditions généralement réalistes que  $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$  soit à peu près gaussien dès lors que  $n$  est assez grand ( $n > 30$  est en pratique bien suffisant), et d'autre part grâce à la convergence de la variance d'échantillon  $S_n^2$  en tant qu'estimateur de  $\sigma^2$ .

• **IC pour la moyenne dans le cas  $\sigma^2$  connue :**

Revenons maintenant sur la situation où  $\sigma^2$  est connu qui bien que présentée dans tous les ouvrages est un cas très simple, car rares sont les situations pratiques de ce type.

En fait le cas où  $\sigma^2$  est connu a été traité dans l'exemple [2.1.1](#) où par commodité, on a supposé  $\sigma^2 = 1$ . L'IC obtenu est donc :

$$IC_{0,95}(\mu) = \left[ \bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} \right] \quad (2.2)$$

Dans ce cas, les quantiles sont à lire sur le tableau de loi de Gauss du fait que  $\sigma^2$  n'a pas à être estimé.

**Remarque 2.2.3** *Pratiquement, nous rencontrons que le cas où la variance est inconnue. Par exemple, dans un problème de régression simple*

$$Y_i = a + bX_i + \varepsilon_i, \quad i = 1, \dots, n$$

les erreurs  $\varepsilon$  sont supposés souvent de loi mère normale, centrée et de variance résiduelle  $\sigma_\varepsilon^2$  inconnue :

$$\varepsilon \rightsquigarrow N(0; \sigma_\varepsilon^2).$$

Nous donnons maintenant une application sur des données simulées, en utilisant le logiciel statistique R, des IC de la moyenne dans le cas gaussien.



**Exemple 2.2.1 (IC pour la moyenne dans le cas  $\sigma^2$  connue)** *Considérons des échantillons de tailles  $n$  variée de 30 à 300 d'une population normale  $\mathcal{N}(\mu; \sigma^2)$ . Fixons le niveau de confiance à 95% ( $\alpha = 0.05$ ), donc  $z_{1-\frac{\alpha}{2}} = 1.96$ . On trouve du résultat (2.2) précédent l'IC de la moyenne :*

$$IC_{0.95}(\mu) = \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} = \bar{X} \pm \frac{1}{\sqrt{n}} 1.96.$$

*Nous prenons dans cet exemple :  $\mu = 0$  et  $\sigma^2 = 1$ . L'IC pour la moyenne pour  $\sigma^2$  connue est illustré par la figure 2.1, qui donne l' $IC_{0.95}(\mu)$  en fonction de  $n$  ( $n = 30, 31, \dots, 300$ ) :*

*La courbe en noir représente la moyenne empirique  $\bar{X}$ , qui converge vers la vraie valeur  $\mu = 0$  (la ligne horizontale),*

*Les lignes en rouge et en bleu sont respectivement les bornes inf et sup de l' $IC_{0.95}(\mu)$ .*

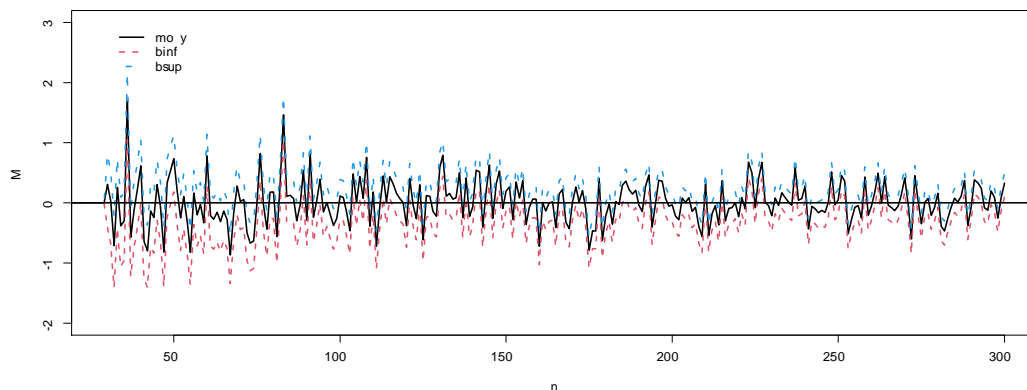


FIG. 2.1 – IC de niveau 95% pour la moyenne dans le cas  $\sigma^2$  connue d'une population normale  $N(0; 1)$ .

**Code R utilisé pour l'exemple précédent :**

```
# IC mu pour sigma connue
ICmoyenne <- fonction(x,sd ,alpha)
{ n=length(x)
# sd=sqrt((1/n)*sum((x-mean(x))^2) # sigma inconnue
# sd=sqrt((1/(n-1))*sum((x-mean(x))^2) # sigma corrigé
moy=mean(x)
```

```
int <- 1 - (alpha/2)
binf <- moy - qnorm(int) * sd/sqrt(n)
bsup <- moy + qnorm(int) * sqrt(sd/n)
bornes <- c(moy, binf, bsup)
return(bornes) }
N=300
M=rep(0,(N-30)); Minf=M; Msup=M
for (n in 30 :N){
x=rnorm(n,,4) # sd = 2
M[n]=ICmoyenne(x ,2, 0.05)[1]
Minf[n]=ICmoyenne(x ,2, 0.05)[2]
Msup[n]=ICmoyenne(x , 2, 0.05)[3] }
n=1 :N##### M[1 :30]=0; Minf[1 :30]=0; Msup[1 :30]=0
plot(n,M,type='l',lwd=2,ylim=c(-2,3),xlim=c(30,N))
lines(Minf,col=2,lty=2,lwd=2)
lines(Msup,col=4,lty=4,lwd=2)
abline(h=0,lwd=2)
legend(30,3,c("moy", "binf", "bsup"),
col=c(1,2,4), lty=c(1,2,4), lwd=c(2,2,2), bty = "n", cex=1)
```

**Remarque 2.2.4** *Il en est de même pour les autres codes R utilisés, avec de légères modifications selon chaque cas.*

**Exemple 2.2.2 (IC  $\mu$  en fonction de  $\alpha$  pour  $\sigma^2$  connue)** Reprenons l'exemple [2.2.1](#) précédent, dont on a choisi un niveau de confiance  $\gamma = 95\%$ . Nous étudierons ici, l'IC de la moyenne  $\mu$  en fonction de  $\gamma = 1 - \alpha$ , tel que  $\alpha \in (0; 1)$ . On prélève des échantillons de taille  $n = 100$  d'une population normale  $\mathcal{N}(0; 1)$ , et on fait varier  $\gamma$  entre 0 et 1. Il est clair que par construction l'augmentation du niveau de confiance  $\gamma$  implique une augmentation des bornes

de l'IC. Par exemple,

$$IC_1(\mu) = \bar{X} \pm \frac{1}{\sqrt{n}}z_1 = \bar{X} \pm \infty, \quad z_1 = \infty$$

Autrement dit,  $\mu \in [-\infty, \infty] = R$  à 100% et inversement,

$$IC_0(\mu) = \bar{X} \pm \frac{1}{\sqrt{n}}z_{1/2} = \bar{X} \pm 0, \quad z_{1/2} = 0.$$

Autrement dit,  $\mu = \bar{X}$  à 0%. La figure 2.2, donne une illustration de l'IC $_{\gamma}(\mu)$  en fonction du niveau de confiance  $\gamma \in (0, 1)$ .

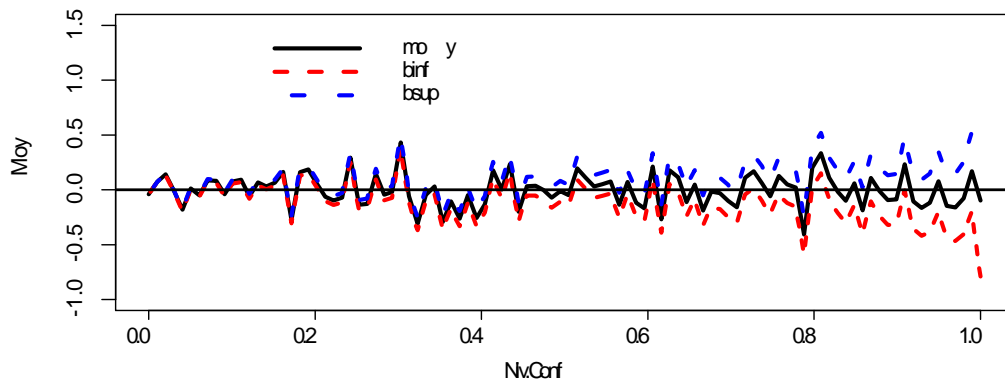
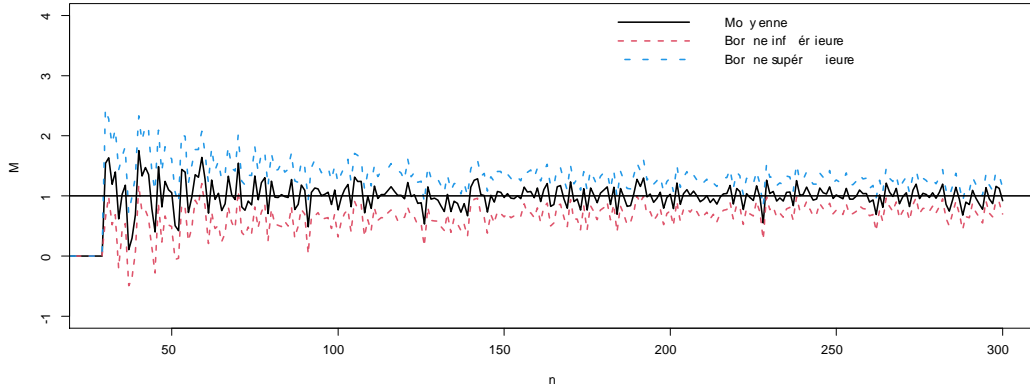


FIG. 2.2 – IC  $\mu$  en fonction du niveau de confiance  $\gamma$ .

**Exemple 2.2.3 (IC pour la moyenne dans le cas  $\sigma^2$  inconnue)** Reprenons l'exemple 2.2.1 précédent. Nous étudierons ici, l'IC de la moyenne  $\mu$  dans le cas où la variance est inconnue (elle sera donc estimée et remplacé par la variance empirique  $S^2$ ). Pour une population normale  $\mathcal{N}(\mu; \sigma)$  D'après (2.1) :

$$IC_{0.95}(\mu) = \bar{X} \pm \frac{S}{\sqrt{n}}t_{1-\frac{\alpha}{2}}^{(n-1)} = \bar{X} \pm \frac{S}{\sqrt{n}}t_{0.975}^{(n-1)}.$$

L'IC pour la moyenne ( $\mu = 1$ ) si  $\sigma^2$  inconnue est illustré par la figure 2.3, qui donne l'IC $_{0.95}(\mu)$  en fonction de  $n$  (de 30 à 300) et pour  $\gamma = 95\%$ .


 FIG. 2.3 – IC de niveau  $\gamma = 0.95$  pour la moyenne dans le cas  $\sigma^2$  inconnue.

## 2.2.2 IC pour la variance

### • IC pour la variance dans le cas $\mu$ inconnue :

Nous supposons que la moyenne  $\mu$  est également inconnue. Reprenons le résultat du (1.1.1)

et d'après (1.1) on a :

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

d'où :

$$P\left(\chi_{0.025}^{2(n-1)} < \frac{(n-1)S_n^2}{\sigma^2} < \chi_{0.975}^{2(n-1)}\right) = 0.95$$

où  $\chi_{\alpha}^{2(n-1)}$  est le quantile d'ordre  $\alpha$  de la loi  $\chi^2(n-1)$  à  $n-1$  d.d.l, (pour  $\alpha = 0.05$  on trouve le quantile agouche dans IC  $1 - \frac{\alpha}{2} = 0.975$  et on trouve le quantile adroite dans IC  $\frac{\alpha}{2} = 0.025$  et  $1 - \alpha = 0.95$ ). Ces quantiles se trouvent dans les tables statistiques ordinaires. On peut directement isoler  $\sigma^2$  pour obtenir :

$$P\left(\frac{(n-1)S_n^2}{\chi_{0.975}^{2(n-1)}} < \sigma^2 < \frac{(n-1)S_n^2}{\chi_{0.025}^{2(n-1)}}\right) = 0.95$$

et

$$IC_{0.95}(\sigma^2) = \left[ \frac{(n-1)S_n^2}{\chi_{0.975}^{2(n-1)}}, \frac{(n-1)S_n^2}{\chi_{0.025}^{2(n-1)}} \right] \quad (2.3)$$

cet IC est peu robuste vis-à-vis de l'hypothèse gaussienne, contrairement à celui sur  $\mu$ . On

ne peut donc l'utiliser dans des situations où la loi mère diffère d'une loi normale. Ceci est vrai même pour une grande taille d'échantillon car on montre que la loi asymptotique de  $S_n^2$  (plus précisément de  $\sqrt{n}(S_n^2 - \sigma^2)$ ) dépend de la loi mère.

**Remarque 2.2.5** De (2.3), on peut déduire un IC pour l'écart type  $\sigma$  de celui sur la variance :

$$IC_{0.95}(\sigma) = \left[ \sqrt{(n-1)S}/\sqrt{\chi_{0.975}^{2(n-1)}}, \sqrt{(n-1)S}/\sqrt{\chi_{0.025}^{2(n-1)}} \right].$$

Au passage on peut comparer la variabilité de l'écart type empirique  $S$  et de la moyenne empirique  $\bar{X}$  pour une loi de gauss tout du moins. En première approximation en appliquant la formule pour une fonction d'une v.a,  $V(S) \simeq \frac{\sigma^2}{2(n-1)}$  alors que  $V(\bar{X}) = \frac{\sigma^2}{n}$ . La fluctuation de  $S$  est plus faible, ce qui se retrouve au niveau des précisions des IC.

• **IC pour la variance dans le cas  $\mu$  connue :**

Dans le cas où  $\mu$  est connue, la statistique  $S_n'^2$  exprimée en termes de valeur exacte de  $\mu$ , et défini par :

$$S_n'^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

tel que  $S_n'^2$  constitue un estimateur de  $\sigma^2$  plus précis que  $S_n^2$ , car la moyenne  $\mu$  de  $X$  est inconnue alors la variance empirique représente l'estimateur naturel de  $\sigma^2$ , on a

$$\frac{nS_n'^2}{\sigma^2} \rightsquigarrow \chi^2(n)$$

d'où

$$P \left( \chi_{0.025}^{2(n)} < \frac{nS_n'^2}{\sigma^2} < \chi_{0.975}^{2(n)} \right) = 0.95$$

ou  $\chi_{\alpha}^{2(n)}$  est le quantile d'ordre  $\alpha$  de la loi  $\chi^2(n)$  de  $n$  d.d.l. On peut directement isoler  $\sigma^2$  pour obtenir :

$$P \left( \frac{nS_n'^2}{\chi_{0.975}^{2(n)}} < \sigma^2 < \frac{nS_n'^2}{\chi_{0.025}^{2(n)}} \right) = 0.95.$$

Finalement nous trouvons l'intervalle de confiance dans le cas  $\mu$  est connue :

$$IC_{0.95}(\sigma^2) = \left[ nS_n'^2 / \chi_{0.975}^{2(n)} \quad , \quad nS_n'^2 / \chi_{0.025}^{2(n)} \right]. \quad (2.4)$$

**Exemple 2.2.4 (IC pour la variance dans le cas  $\mu$  connu)** Dans cet exemple on prélève des échantillons de taille  $n$  variée de 30 à 500 d'une population mère  $\mathcal{N}(\mu, \sigma^2 = 4)$ , pour un niveau de confiance 95%. En utilisant (2.4), l'IC pour la variance si  $\mu$  connue ( $\mu = 0$ , dans cet exemple), est illustré par la figure 2.4, qui donne l' $IC_{0.95}(\sigma^2)$  en fonction de  $n$ .

\*) La variance empirique (la courbe en noir) converge vers la vraie valeur  $\sigma^2 = 4$  (la ligne horizontale),

\*) Les bornes inf et sup sont respectivement représentés par les lignes en rouge et en bleu.

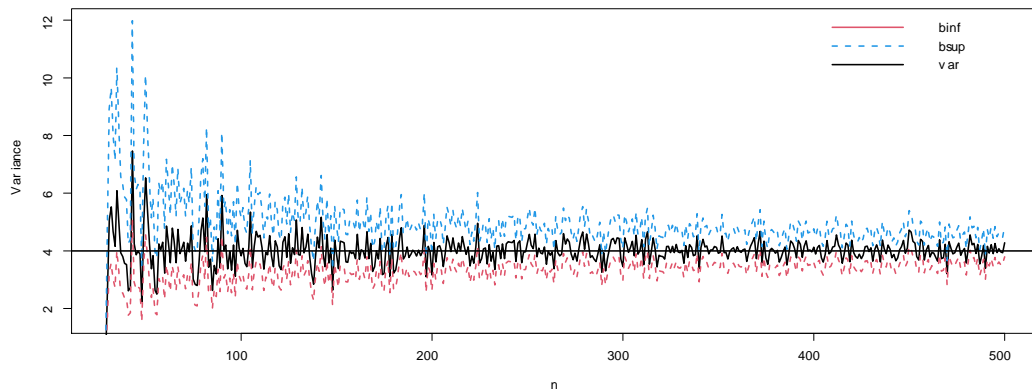
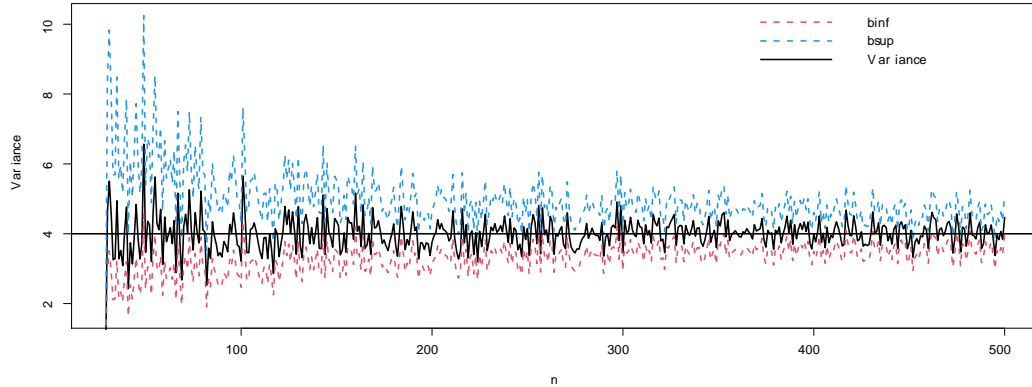


FIG. 2.4 – IC de niveau  $\gamma = 0.95$  pour la variance dans le cas  $\mu$  connue.

**Exemple 2.2.5 (IC pour la variance dans le cas  $\mu$  inconnue)** Reprenons l'exemple 2.2.4 précédent, supposons que la population mère est  $\mathcal{N}(\mu, \sigma^2 = 4)$ , avec une moyenne  $\mu$  inconnue, donc remplacer par la moyenne empirique. En utilisant (2.3), l'IC pour la variance si  $\mu$  inconnue, est illustré par la figure 2.5 suivante, qui donne l' $IC_{0.95}(\sigma^2)$  en fonction de  $n$ .


 FIG. 2.5 – IC de niveau  $\gamma = 0.95$  pour la variance dans le cas  $\mu$  inconnue.

### 2.2.3 IC pour la différence des moyennes

Nous considérons ici deux lois mères (en pratique, souvent, deux populations) et souhaitons construire un *IC* sur la différence de leurs moyennes. Comme exemples, l'écart entre la taille moyenne des filles et des garçons de même âge, l'écart de revenu moyen des actifs entre deux régions, ... Pour cela on dispose de deux échantillons indépendants pris de chaque population (le fait de prendre une sœur et un frère pour l'exemple de la taille ne respecterait pas cette hypothèse d'indépendance des deux échantillons).

La procédure classique que nous allons développer suppose que les deux lois ont même variance  $\sigma^2$ . Soit un échantillon de taille  $n_1$  issu de la loi  $\mathcal{N}(\mu_1, \sigma^2)$  et un échantillon indépendant du premier de taille  $n_2$  issu de la loi  $\mathcal{N}(\mu_2, \sigma^2)$ . Soit  $\bar{X}_1$  et  $S_1^2$  la moyenne et la variance empirique du premier échantillon et de même  $\bar{X}_2$  et  $S_2^2$  pour le deuxième échantillon. On a

$$\bar{X}_1 \rightsquigarrow \mathcal{N}\left(\mu_1, \frac{\sigma^2}{n_1}\right) \quad \text{et} \quad \bar{X}_2 \rightsquigarrow \mathcal{N}\left(\mu_2, \frac{\sigma^2}{n_2}\right)$$

donc,

$$\bar{X}_1 - \bar{X}_2 \rightsquigarrow \mathcal{N}\left(\mu_1 - \mu_2, \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right).$$

Alors,

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow \mathcal{N}(0; 1).$$

Le problème qui se pose est celui de l'estimation de  $\sigma$  que l'on effectue, en fait via  $\sigma^2$ . Sachant que

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \rightsquigarrow \chi^2(n_1 - 1) \quad \text{et} \quad \frac{(n_2 - 1)S_2^2}{\sigma^2} \rightsquigarrow \chi^2(n_2 - 1),$$

l'indépendance des deux échantillons entraîne que

$$\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2} \rightsquigarrow \chi^2(n_1 + n_2 - 2).$$

En faisant le rapport de la va  $\bar{X}_1 - \bar{X}_2$  centrée réduite à la racine carrée de la va ci-dessus divisée par ses d.d.l, et en posant :

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2},$$

on obtient

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \rightsquigarrow t(n_1 + n_2 - 2).$$

La fonction ci-dessus est une fonction pivot qui aboutit immédiatement à

$$IC_{0,95}(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{0,975}^{(n_1+n_2-2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (2.5)$$

où

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

est la variance empirique pondérée en fonction des tailles d'échantillon respectives. Qu'en est-il de la condition très restrictive d'égalité des variance? En fait, on a pu montrer que celle-ci n'est pas si cruciale si les tailles d'échantillons  $n_1$  et  $n_2$  diffèrent peu. Dans ce cas un facteur 2 pour le rapport des variances reste acceptable. En revanche si  $n_1$  et  $n_2$  diffèrent substantiellement la formule ci-dessus s'applique mal quand les variances ne sont pas proches.



Alors, on peut effectuer les mêmes développements que précédemment en introduisant les variances respectives des deux lois  $\sigma_1^2$  et  $\sigma_2^2$  pour obtenir :

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightsquigarrow N(0; 1)$$

et si les tailles d'échantillons sont élevées, disons au-delà de 100, conserver une approximation raisonnable en substituant à  $\sigma_1^2$  et  $\sigma_2^2$  leurs estimations  $S_1^2$  et  $S_2^2$ , d'où :

$$IC_{0,95}(\mu_1 - \mu_2) \simeq (\bar{X}_1 - \bar{X}_2) \pm 1,96 \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

On remarquera que si  $n_1 = n_2$  cette formule est identique à celle du cas où  $\sigma_1^2 = \sigma_2^2$  (mais à parts les quantiles d'ordre 0.975 qui seront cependant pratiquement identiques pour les grands échantillons). D'autres formules d'approximation plus précises ont été développées, mais elles donnent des résultats numériques proches de ceux obtenus avec l'hypothèse d'égalité des variances ce qui encourage peu leur utilisation par les participants.

**Remarque 2.2.6** *Indiquons qu'il existe un usage assez répandu consistant à effectuer au préalable un test de l'hypothèse d'égalité des variances. Même si l'on peut admettre que cela a l'avantage de constituer un garde-fou, cette procédure ne fournit pas une garantie suffisante quant à l'applicabilité de la formule classique en cas d'acceptation de l'hypothèse par le test. Quant à l'usage de celle-ci en dehors des conditions de «normalité» des deux lois, il est acceptable pour les mêmes raisons que celles exposées dans le cas d'une seule loi. En résumé, le point critique est une différence trop sensible des dispersions des deux lois.*

**Exemple 2.2.6 (IC pour la différence entre deux moyennes)** *On prélève deux échantillons de taille  $n_1$  et  $n_2$  variées de 30 à 400 de deux populations de la loi normale  $\mathcal{N}(\mu_1, \sigma^2)$  et  $\mathcal{N}(\mu_2, \sigma^2)$ , pour un niveau de confiance 95%. On trouve par le résultat (2.5) précédent, l'IC de  $\mu_1 - \mu_2$  :*

$$IC_{0,95}(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{1-\frac{\alpha}{2}}^{(n_1+n_2-2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

*Fixons  $\mu_1 = 0$  et  $\mu_2 = 1$ , et  $S_1 = S_2 = 2$  l'écart type de l'échantillon  $X_1$  et  $X_2$  respectivement.*

L'IC pour la différence entre les deux moyennes  $\mu_1 - \mu_2$  en fonction de  $n$  est illustré par la figure 2.6.

La courbe en noir est l'estimateur empirique de la différence entre les moyennes, qui converge vers la vraie valeur ( $\mu_1 - \mu_2 = -1$ ) représenté par la ligne horizontale.

Les bornes inf et sup sont respectivement les lignes en rouge et en bleu.

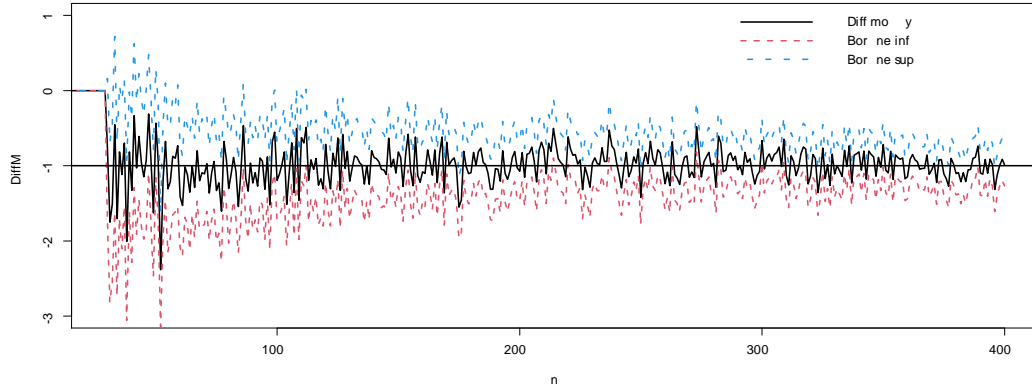


FIG. 2.6 – IC de niveau  $\gamma = 0.95$  pour la différence entre deux moyennes.

## 2.2.4 Intervalle de confiance sur le rapport des variances

De même que la section précédente, on considère ici, le rapport  $\frac{\sigma_1^2}{\sigma_2^2}$  des vraiances pour les lois mères  $\mathcal{N}(\mu_1, \sigma_1^2)$  et  $\mathcal{N}(\mu_2, \sigma_2^2)$ . Comme

$$\frac{(n_1 - 1)S_1^2}{\sigma_1^2} \rightsquigarrow \chi^2(n_1 - 1) \quad \text{et} \quad \frac{(n_2 - 1)S_2^2}{\sigma_2^2} \rightsquigarrow \chi^2(n_2 - 1)$$

On a,

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow F(n_1 - 1, n_2 - 1).$$

Soit finalement après pivotement et compte tenu du fait que  $F_\alpha^{(v_1, v_2)} = 1/F_{1-\alpha}^{(v_2, v_1)}$  :

$$IC_{0,95} \left( \frac{\sigma_1^2}{\sigma_2^2} \right) = \left[ \frac{S_1^2}{S_2^2} F_{0.025}^{(n_2-1, n_1-1)}, \frac{S_1^2}{S_2^2} F_{0.975}^{(n_2-1, n_1-1)} \right] \quad (2.6)$$

**Remarque 2.2.7** Comme pour la procédure relative à une variance, cette formule n'est pas robuste lorsque les lois s'écartent de lois gaussiennes. Son usage est donc très limité.

**Exemple 2.2.7 (IC sur le rapport des variances)** On prélève deux échantillons de taille  $n$  variée de 30 à 400 issus de population mères de la loi normale  $\mathcal{N}(\mu_1, \sigma_1^2)$  et  $\mathcal{N}(\mu_2, \sigma_2^2)$ , pour un niveau de confiance 95%. On trouve par le résultat (2.6) précédent

$$IC_{0,95} \left( \frac{\sigma_1^2}{\sigma_2^2} \right) = \left[ \frac{S_1^2}{S_2^2} F_{\frac{\alpha}{2}}^{(n_2-1, n_1-1)}, \quad \frac{S_1^2}{S_2^2} F_{1-\frac{\alpha}{2}}^{(n_2-1, n_1-1)} \right]$$

Fixons  $\sigma_1 = 4$  et  $\sigma_2 = 2$  les deux écart type, et  $\mu_1 = \mu_2 = 0$ .

L'IC sur le rapport des variances est illustrés par la figure 2.7, dont la courbe en noir qui représente le rapport empirique

$$S_1^2/S_2^2 \simeq \sigma_1^2/\sigma_2^2 = 16/4 = 4,$$

qui converge vers la vraie valeur = 4 (la ligne horizontale).

Les bornes inf et sup sont respectivement les lignes en rouge et en bleu.

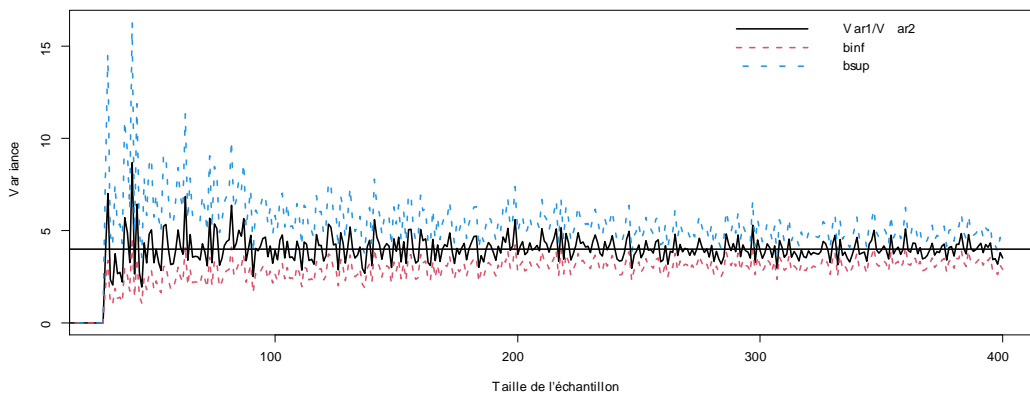


FIG. 2.7 – IC de niveau  $\gamma = 0.95$  pour le rapport entre deux variances.

## Conclusion

**D**ans ce travail, on s'est particulièrement intéressé à l'intervalle de confiance pour la moyenne et la variance dans le cas d'un échantillon gaussien. Nous avons étudié en détail le problème de construction des intervalles de confiance des paramètres (moyenne et variance) d'une loi mère gaussienne dans le cas où la moyenne est connue (respectivement inconnue) et de même pour la variance, ainsi que la différence et le rapport entre eux.

En conclusion, l'intervalle de confiance est une plage de valeurs, dérivée des données d'un échantillon, qui est susceptible de contenir un paramètre de la population avec une certaine probabilité ou niveau de confiance (par exemple, 95%). Il exprime l'incertitude autour de l'estimation de ce paramètre. Il dépend de trois quantités : la taille de l'échantillon, le niveau de confiance et de la loi limite utilisée.

Les résultats théoriques obtenus sont illustrés par des exemples numériques et graphiques de simulation à l'aide du logiciel de traitement statistique R. Nous envisageons dans nos études futures, l'application de ces différents résultats sur des données réelles.

# Bibliographie

- [1] Cavaillé, A . (2021). Cours de probabilités. IMT Mines Alès- Département de Mathématiques. [https://lms.fun-mooc.fr/asset-v1:MinesTelecom+04042+session01+type@asset+block/Polycopie\\_8.pdf](https://lms.fun-mooc.fr/asset-v1:MinesTelecom+04042+session01+type@asset+block/Polycopie_8.pdf)
- [2] Dusart, P. (2015). Cours de statistiques inférentielles. Université de Limoges.
- [3] Jacques, J. (2022). Statistiques inférentielles. <https://eric.univ-lyon2.fr/~jjacques/>.
- [4] Kim, N.S . (2016). La loi normale. <https://fr.scribd.com/document/702766944/2016-TC-Schalbar-Kim-Nora-2>.
- [5] Lejeune, M. (2010). *Statistique La Théorie et ses Applications*. Springer.
- [6] Le Coutre, J.P. (2016). *Statistique et Probabilités*. Dunod.
- [7] Meraghni B. (2019), Estimation par intervalle de confiance. Mémoire master en mathématiques, Université de Biskra.
- [8] Philippe, T.(1989). *Méthodes statistiques*. Economica.
- [9] Saporta, G. (2006). *Probabilités, Analyse des données et Statistique*. Technip.

# Annexe : Abréviations et Notations

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

$\mathbb{E}(X), \mu$	: Espérance de $X$ .
$\mathbb{V}(X), \sigma^2$	: Variance de $X$ .
$Cov(X, Y)$	: Covariance mathématique du couple $(X, Y)$ .
$va$	: Variable aléatoire.
$d.d.l$	: Degré de liberté.
$i.i.d$	: Indépendant identiquement distribué.
$IC$	: Intervalle de confiance.
$\mathcal{N}(\mu, \sigma^2)$	: Loi normale.
$\mathcal{N}(0; 1)$	: Loi normale centrée réduite.
$\chi^2(n)$	: Loi du khi-deux.
$t(n)$	: Loi de student.
$F(n_1, n_2)$	: Loi de fisher.
$f(x)$	: Fonction de densité.
$\mathbb{F}(x)$	: Fonction de répartition.
$\rightsquigarrow$	: Suit la loi .
$\sigma$	: L'écart-type.

$\varphi_X(t)$	: Fonction caractéristique.
$g(t)$	: Fonction génératrice des moments.
$\phi(x)$	: Fonction de répartition de la loi normale standard.
$\bar{X}$	: Moyenne empirique.
$S_n^2$	: Variance empirique.
$S^{*2}$	: Variance empirique corrigée.
$S$	: L'écart type empirique.
$z_\alpha$	: Valeur critique de la loi normale standard.
$t_\alpha^{(n-1)}$	: Valeur critique de la loi de student.
$\chi_\alpha^{2(n-1)}$	: Valeur critique de la loi du khi-deux.
$z_\alpha$	: Valeur critique de la loi normale standard.
$t_\alpha^{(n-1)}$	: Valeur critique de la loi de student.
$\chi_\alpha^{2(n-1)}$	: Valeur critique de la loi du khi-deux.
$F_\alpha^{(v_1, v_2)}$	: Valeur critique de la loi de fisher.
$\gamma = 1 - \alpha$	: Le niveau de confiance.
$\alpha$	: Le seuil.
$\xrightarrow{\mathcal{P}}$	: Convergence en probabilité.
$\xrightarrow{\mathcal{L}}$	: Convergence en loi.
$\xrightarrow{p.s.}$	: Convergence presque sûre.
$\xrightarrow{Mq}$	: Convergence en moyenne quadratique.
$\mathcal{U}[0; 1]$	: Loi uniforme continue de paramètres 0 et 1.

# ملخص

نقدم في هذه المذكرة لمحة عامة حول مفهومين أساسيين في الإحصاء، غالبًا ما يستخدمان معًا: مجال الثقة وقانون غاوس. في الواقع، سوف نستخدم خصائص التوزيع الطبيعي، وتقدير هذه المعلمات وسلوكها المقارب لإنشاء مجال الثقة للمتوسط، التباين، الفرق بين متوسطين والنسبة بين تباينين في مجتمع غاوسي. تم تقديم دراسة محاكاة باستخدام برنامج المعالجة الإحصائية R لتوضيح مختلف النتائج النظرية.

## Résumé

*Nous présentons dans ce mémoire de master un aperçu sur deux concepts fondamentaux en statistique, souvent utilisés ensemble: l'intervalles de confiance et la loi gaussienne. En effet, nous utiliserons les propriétés de la distribution normale, l'estimation de ces paramètres et leurs comportement asymptotiques pour établir des intervalles de confiance pour la moyenne, la variance, la différence deux moyennes et le rapport de deux variances d'une population gaussienne. Une étude de simulation à l'aide du logiciel de traitement statistique R est donnée pour illustrer les résultats théoriques.*

## Abstract

*In this mater dissertation we present an overview about two fundamental concepts in statistics, often used together: confidence intervals and the Gaussian law. In fact, we will use the properties of the normal distribution, the estimation of these parameters and their asymptotic behavior to establish confidence intervals for the mean, the variance, the difference of two means and the ratio of two variances of a Gaussian population. A simulation study using the R statistical software is given to illustrate the theoretical results.*