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 :

DIF Chahinaz

Titre :

Sur l'estimation paramétrique

Membres du Comité d'Examen :

Pr.	CHERFAOUI Mouloud	UMKB	Président
Dr.	ABDELLI Jihane	UMKB	Encadreur
Dr.	SOLTANE Luiza	UMKB	Examinatrice

Juin 2024

Dédicace

Après un parcours académique de plusieurs années, traversant de nombreuses difficultés et fatigues, me voici aujourd'hui au seuil de ma graduation, voyant les résultats de mes efforts et levant ma casquette avec fierté...

À cet homme formidable qui a fait ressortir le meilleur en moi et m'a toujours encouragé à atteindre mes ambitions, celui qui m'a appris la vie de la plus belle manière et a tout donné sans jamais rien retenir, mon père, que dieu te garde à nos côtés...

À la femme qui a fait de moi une fille ambitieuse aimant les défis, mon premier modèle à travers lequel j'ai appris la force et la confiance en soi, celle dont satisfaction crée la réussite pour moi, ma mère, que dieu te prolonge la vie en santé et bonheur...

À l'étreinte chaleureuse qui englobe tout l'univers, à celle qui rend la vie plus belle par sa présence, celle qui m'a donné accès à ses sources de connaissance et d'expérience de vie, à mon soutien et deuxième mère "ASSASSI Fatima Zahra"...

À la montagne qui me protège des tempêtes de la vie, au soutien inébranlable, mes frères (Chamsse eddine,Anwar), À ceux qui ont partagé avec moi les épreuves des études et les nuits blanches, ceux qui ont été les meilleurs alliés sur mon chemin, mes sœurs (hayette,hassiba,hanine)...

Merci à mes amies "Selma, Ibtihal, Chahrazed, Rayan, Amine, Noureddine" qui ont été à mes côtés dans les moments les plus difficiles et les plus heureux, à ma seconde famille, l'association "Biskra Reads" Culturelle, et à tous ceux qui m'ont soutenu et encouragé dans la réalisation de cette réussite.

Et enfin, je dédie cette mémoire à moi-même qui a travaillé dur et avec dévouement pour atteindre cet objectif, espérant que ce soit un pas vers un avenir radieux et un succès continu.

REMERCIEMENTS

Tout d'abord, je tiens à exprimer ma profonde gratitude envers Allah Tout-Puissant, qui m'a accordé la force et la capacité de mener à bien ce travail avec succès. Je voudrais également exprimer ma gratitude envers ma superviseure **Dr. Abdelli Jihane**, pour ses conseils et son soutien.

Je remercie également les membres du jury, le **Pr. CHERFAOUI Mouloud** et **Dr. SOLTANE Luiza**, pour leurs efforts et leur précieuse contribution en fournissant leurs conseils et leurs commentaires constructifs qui ont contribué à améliorer cette recherche.

Enfin, je voudrais exprimer ma gratitude envers mes parents ainsi que toute ma famille, ainsi qu'à tous ceux qui m'ont soutenu et aidé tout au long de ce voyage, y compris mes amis et camarades d'études. Un grand merci à vous tous.

Table des matières

Table des figures

Liste des tableaux

Introduction

Les concepts et les méthodes statistiques ne sont pas seulement utiles mais nécessaires pour comprendre le monde qui nous entoure, fournissant des moyens pour avoir de nouvelles idées sur les différents phénomènes dans tous les domaines de la vie sociale et économique. L'estimation paramétrique est l'une des activités principales en statistique, consistant à estimer certaines caractéristiques statistiques de la distribution (comme la moyenne, la variance, ...etc) sur la base de données mesurées expérimentalement, utilisées pour estimer les paramètres de la population. L'objectif est de comparer les distributions des estimateurs et de déterminer la valeur du paramètre inconnu, en utilisant les informations extraites de l'échantillon permettant d'obtenir des résultats concernant la population par estimation, la valeur inconnue pour la population à estimer de l'échantillon étant appelée le paramètre. Le paramètre de la population est estimé à partir d'une statistique calculée sur la base de l'échantillon. L'estimation paramétrique étant divisée en deux parties :

- l'estimation ponctuelle du paramètre qui consiste à évaluer la valeur du paramètre de la population à l'aide d'une valeur unique de l'échantillon.

- l'estimation par intervalle de confiance qui implique de déterminer un intervalle de confiance incluant la vraie valeur du paramètre avec une probabilité prédéterminée. Il existe de nombreuses méthodes pour estimer le paramètre telles que la méthode des moindres carrés, la méthode des moments, la méthode du maximum de vraisemblance...etc. Dans le cadre de cette recherche, nous fournirons un aperçu historique de trois méthodes d'estimation (méthode du maximum de vraisemblance, méthode des moments et méthode des intervalles de confiance).

La méthode du maximum de vraisemblance ayant été développée par le statisticien Ronald Fisher en 1922, cette approche est l'une des premières à se baser sur la probabilité pour les estimations. Nous concentrerons également sur les propriétés de cette méthode (telles que la convergence, la densité...etc) par rapport à plusieurs modèles statistiques.

La méthode des moments, d'abord discutée par Pearson, a été généralisée par Hansen au XIXe siècle.

La méthode d'intervalle de confiance a été introduite par Jerzy Neyman.

Cette recherche sera divisée en deux chapitres.

Dans le premier chapitre, nous présentons les concepts généraux des probabilités et de la statistique en donnant une définition des lois de probabilité, la moyenne, la variance...etc, en mentionnant toutes les caractéristiques de la loi de probabilité.

Dans le deuxième chapitre, nous commençons par rappeler la définition de l'estimation et de l'estimateur, leurs caractéristiques, puis nous discuterons de trois méthodes d'estimation auto (la méthode du maximum de vraisemblance, la méthode des moments, et la méthode de l'intervalle de confiance).

Finalement, les résultats numériques et graphiques sont obtenus en utilisant le logiciel de statistiques R, où les codes de programmation sont compilés et les raccourcis ainsi que les expressions sont utilisés dans deux annexes.

Chapitre 1

Sur les lois du distribution paramétrique

Dans ce chapitre, on va présenter des notions de base sur la probabilité en général, en donnant après quelques lois de probabilité qui seront importantes pour la suite. Dans ce dernier nous allons utiliser les références[1, 2, 4]

1.1 Notions de base

1.1.1 Loi de probabilité

Dans le domaine des statistiques, une loi de probabilité décrit le comportement aléatoire d'un phénomène qui dépend du hasard. Grâse à ces lois, on peut modéliser des résultats inconnus à l'avance, tels que des phénomènes physiques, des observation biologiques ou des résultats économique. Une loi de probabilité pour une variable aléatoire X permet de calculer la probabilité d'obtenir différents résultats possibles pour cette variable aléatoire (v.a), et nous allons deux types de lois "les lois discrètes et les lois continues".

Définition 1.1.1 *On dit que la loi de probabilité de la v.a X est discrète, où encore que*

X est une v.a discrète s'il existe un ensemble dénombrable E telque $P(X \in E) = 1$. Dans ce cas, la loi de probabilité de X est uniquement déterminée par la probabilité $P(X = x)$, $x \in E$.

Définition 1.1.2 *on dit que la loi de probabilité de la v.a X est continue, où encore que X est une v.a continue s'il existe une fonction $f : \mathbb{R} \to \mathbb{R}$ non négative et telle que, pour tous ensemble $A \in \mathbb{R}$, on peut écrire*

$$P(X \in A) = \int_A f(x)dx.$$

1.1.2 Espérance et Variance

Espérance :

L'espérance E d'une v.a X correspond à la moyenne des valeurs possibles de X pondérées par les probabilités associées à ces valeurs, c'est un paramètre de position qui correspond au moment d'ordre 1 de la v.a X. C'est l'équivalente de la moyenne arithmétique $\bar{X}$. En effet lorsque le nombre d'epeuves n est grand, X tend vers $E(x)$.

Définition 1.1.3 *si X est une v.a descrète définie sur un univers probabilité Ω. On appelle espérance de X, le réel définie par :*

$$E(x) = \sum_{\omega \in \Omega} X(\omega) P(\omega) = \sum_{x \in E} x P(X = x).$$

Remarque 1.1.1 *Si $X(\Omega)$ est infini, on n'est pas sûr que l'espérance existe.*

Définition 1.1.4 *Si X est une v.a absulment continue de densité f, on appelle espérance de X, le réel définie par :*

$$E(x) = \int_E x dP_X = \int_{-\infty}^{+\infty} x f(x) dx.$$

Si cette intégrale est convergente.

Variance :

Soit $X \in L^2(\Omega, \mathcal{F}, P)$, la variance de X est :

$$var(x) = E(x - E(x))^2$$

1.1.3 Fonction de répartition

La densité de probabilité $P(X)$ ou la fonction de répertition $F(x)$ définissent la loi de probabilité d'une v.a continue X. La fonction de distribution cumulée $F(x)$ exprime la probabilité que X n'exéde pas la valeur x :

$$F(x) = P(X \leq x).$$

De même, la probabilité que X soit entre a et b $(b \succ a)$ vaut :

$$F(a \prec X \prec b) = F(b) - F(a).$$

1.1.4 La Convergence d'une suite de v.a

Définition 1.1.5 *Soit $(X_n)_{n\geq 1}$ une suit de v.a réelles definies sur un espace de probabilité* $(\Omega, \mathcal{F}, P)$.

Convergence en probablilité :

Définition 1.1.6 *On dit qu'une suit de v.a $(X_n)_{n\geq 1}$ converge en probabilité vers une v.a X, si*

$$\forall \varepsilon \succ 0, \qquad \lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

On note : $X_n \xrightarrow{p} X.$

Convergence presque sûre :

Définition 1.1.7 *On dit qu'une suite de v.a $(X_n)_{n\geq 1}$ converge presque sûrement vers une v.a X, si*

$$P(\omega \in \Omega,\ \lim_{n\to\infty} X_n(\omega) = X(\omega) = 1.$$

On note $X_n \xrightarrow{p.s} X$.

1.2 Les lois du distribution paramètrique

1.2.1 La loi normale

On dit qu'une v.a X obéit à une loi normale de moyenne μ (ou $-\infty \prec \mu \prec +\infty$) et de variance $\delta^2 \succ 0$ lorsqu'elle présente la fonction de densité :

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} \exp\left(-(1/2)\left[(x-\mu)/\delta\right]^2\right) \ ,\quad -\infty \prec x \prec +\infty.$$

En abrégé, on écrit $X \sim N(\mu, \delta^2)$.

La moyenne et la variance de la loi normale :

On peut aisement détrmine la moyenne et la variance de la loi normale. En effet ,puisque

$$E(x) = \int_{-\infty}^{+\infty} \frac{x}{\delta\sqrt{2\pi}} \exp\left(-(1/2)\left[(x-\mu)/\delta\right]^2\right) dx.$$

En posant $z = (x-\mu)/\delta$ et en utilisant la méthode d'integration par parties, on obtient

$$\begin{aligned} E(x) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} (\mu + \delta z) \exp\left(-z^2/2\right) dz \\ &= \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right) dz + \delta \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right) dz. \end{aligned}$$

Or, comme la première fonction à integrer n'est autre que la fonction de densité d'une variable normale de paramètres $\mu = 0$ et $\delta = 1$. La première integrale a une valeur de 1. La valeur de la deuxième integrale est par ailleurs nulle, puisque :

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} z \exp\left(-z^2/2\right) dz = -\frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right) |_{-\infty}^{+\infty} = 0.$$

Par conéquent,

$$E(x) = \mu(1) + \delta(0) = 0.$$

Pour déterminer la variance, il faut evaluer

$$var(x) = E[(x-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\delta\sqrt{2\pi}} \exp\left(-(1/2)\left[(x-\mu)/\delta\right]^2\right) dx.$$

Or, en posant $z = (x-\mu)/\delta$, on obtient :

$$\begin{aligned} var(x) &= \int_{-\infty}^{+\infty} \delta^2 z^2 \frac{1}{\sqrt{2\pi}} \exp\left(-z^2/2\right) dz = \delta^2 \int_{-\infty}^{+\infty} \frac{z^2}{\sqrt{2\pi}} \exp\left(-z^2/2\right) dz \\ &= \delta^2[0+1] = \delta^2. \end{aligned}$$

La fonction de répartition :

La fonction de répartition F est

$$F(X, \mu, \delta) = \frac{1}{2}[1 + \operatorname{erf}((x-\mu)/(\delta\sqrt{2}))].$$

Dans cette formule, X représente la valeur à lequelle nous voulons calculer la probabilitè, μ est la moyenne de la distribution normale, δ est l'ecart-type de la distribution normale, et erf est la fonction d'erreur.

1.2.2 La loi exponentielle

une v.a continue X est dite de "loi exponentielle de paramètre λ", si sa fonction de densité est de la forme :

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{si } x \succ 0 \\ 0 & \text{sinon} \end{cases}.$$

Où le paramètre λ est un nombre réel positif. En abrégé, on écrit $X \sim Exp(\lambda)$.

La moyenne et la variance de la loi exponentielle :

La moyenne de la loi exponentielle est donnée par :

$$E(x) = \int_0^{+\infty} x\lambda \exp(-\lambda x)dx = -x\exp(-\lambda x) \mid_0^{+\infty} + \int_0^{+\infty} \exp(-\lambda x)dx = \frac{1}{\lambda},$$

et sa variance, par :

$$\begin{aligned} var(x) &= \int_0^{+\infty} x^2\lambda \exp(-\lambda x)dx - \left(\frac{1}{\lambda}\right)^2 \\ &= [-x^2\lambda \exp(-\lambda x) \mid_0^{+\infty} +2\int_0^{+\infty} x\exp(-\lambda x)dx] - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}. \end{aligned}$$

La fonction de répartition :

La fonction de répartition F est

$$F(X, \lambda) = 1 - \exp(-\lambda x).$$

Dans cette formule X répresente la valeur à lequelle nous voulons calculer la probabilité, λ répresente le paramètre de taux de la distribution exponentielle.

1.2.3 la loi de Poisson

on dit qu'une v.a X est distribuée selon une "loi de poisson de paramètre $c \succ 0$", si sa fonction de masse donnée par :

$$P(x) = \begin{cases} c^x \exp(-c) & \text{si } x = 0, 1, 2, ... \\ 0 & \text{sinon} \end{cases} .$$

En abrégé, on écrit $X \sim Poisson(c)$.

La moyenne et la variance de la loi de Poisson :

La loi de poisson possède une propriété particulière, sa moyenne et sa variance sont toujours égales.

La moyenne de la loi poisson est donnée par :

$$E(x) = \sum_{x=0}^{\infty} \frac{xe^{-c}c^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-c}c^x}{(x-1)!} = ce^{-c}[1 + \frac{c}{1!} + \frac{c^2}{2!} + ...]$$
$$= ce^{-c} - e^c = c,$$

et sa variance, par :

$$E(x(x-1)) = \sum_{x=0}^{\infty} \frac{x(x-1)e^{-c}c^x}{x!} = c^2,$$

alors

$$var\,(x) = E\left(x^2\right) - [E\,(x)]^2 = c.$$

La fonction de répartition :

$$F\,(k, \lambda) = P\,(X \leq k) = \sum_{x=0}^{k} \frac{\lambda^n \exp(-\lambda)}{x!}.$$

Dans cette formule, k représente le nombre de la valeur à lequelle nous voulons calculer la probabilité, λ représente le taux moyen d'occurance des événements.

1.2.4 La loi Binomiale

La v.a X représente le nombre de succès obtenus en reálisant n epreuves de Bernoulli indépendantes, ayant chacune une probabilité de succés p, est dite de "loi Binomiale de paramètre n et p". Sa fonction de masse $P(x)$ est définie par :

$$P(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, ..., n, \\ 0 & \text{sinon.} \end{cases}$$

En abrégé, on écrit $X \sim Binomiale(n, p)$.

La moyenne et la variance de la loi Binomiale :

La moyenne est donnée par :

$$E(x) = \sum_{x=0}^{n} x \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x} = np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1} (1-p)^{n-x}.$$

Et en posant $y = x - 1$, on obtient :

$$E(x) = np \sum_{x=1}^{n} \frac{(n-1)!}{y!\,(n-x)!} p^y (1-p)^{n-1-y} = np.$$

Sa variance est :

$$\begin{aligned} var\,(x) &= E\left(x^2\right) - \left(E\,(x)\right)^2 = E\,(x\,(x-1)) + E\,(x) - \left(E\,(x)\right)^2 \\ &= n\,(n-1)\,p^2 + np - (np)^2 = np\,(1-p)\,. \end{aligned}$$

La fonction de repartition :

$$F\left(k;n,p\right)=P(X\leq k)=\sum_{x=0}^{kn}C_x^n p^x(1-p)^{n-x}.$$

Dans cette formule, k répresente le nombre de succès souhaité, n représente le nombre total d'essais, C_x^n est le coeffient binomiale qui représente le nombre de façons de choisir x succès parmi n essai.

1.2.5 La loi Gamma

Une v.a continue X est dite de "loi gamma de paramètre r et λ", si sa fonction de densité est de la forme :

$$f\left(x\right)=\begin{cases}\frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}\exp(-\lambda x) & \text{si } x\succ 0,\\ 0 & \text{sinon.}\end{cases}$$

Où $r\succ 0$ et $\lambda\succ 0$. En abrégé, on écrit $X\sim Gamma(r,\lambda)$.

La moyenne et la variance de la loi Gamma :

La moyenne est donnée par :

$$E\left(x\right)=\frac{r}{\lambda}.$$

Sa variance est :

$$var\left(x\right)=\frac{r}{\lambda^2}.$$

La fonction de répartition :

La fonction de répartition F est

$$F\left(x\right)=\begin{cases}0 & \text{si } \; x\leq 0,\\ 1-\int_x^{+\infty}\frac{1}{\Gamma(r)}\left(\lambda t\right)^{r-1}\exp\left(-\lambda t\right)dt & \text{si } x\succ 0.\end{cases}$$

1.2.6 La loi de Bernoulli

C'est la loi d'une variable X ne pouvant prendre que les deux valeurs 1 et 0 avec les probabilitè p et $(1-p)$; X est la fonction indicatrice d'un événement A de probabilite p.

La moyenne et la variance de la loi Bernoulli :

La moyenne donnée par :

$$E(x) = p$$

et sa variance par :

comme $X^2 = X$, $E(x^2) = p$, d'où :

$$var(x) = p(1-p).$$

La fonction de répartition :

La fonction de répartition F est :

$$F(X,p) = \begin{cases} 0 & \text{si } \ x \prec 0 \\ 1-p & \text{si } \ 0 \preceq x \prec 1 \\ 1 & \text{si } \ x \succeq 1 \end{cases}$$

1.2.7 La loi de Bêta

c'rst la loi d'une variable X ; $0 \leq X \leq 1$ dependant de deux paramétres n et p dont la densité est :

$$f(x) = \frac{1}{B(n,p)} x^{n-1}(1-x)^{p-1} \qquad n, p > 0 \text{ où } B(n,p) = \frac{\Gamma(n)\Gamma(p)}{\Gamma(n+p)}.$$

La moyenne et la variance de la loi Bêta :

La moyenne donnée par :

$$E(x) = \frac{n}{n+p}$$

et sa variance par :

$$var(x) = \frac{np}{(n+p+1)(n+p)^2}$$

1.2.8 La loi de khi-carré

Théorème 1.2.1 *soit $Z_{1,}Z_2, ..., Z_k$ des v.a indépendantes et normalement distribuées avec une moyenne $\mu = 0$ et une variance $\delta^2 = 1$. La v.a $U = Z_1^2 + Z_2^2 + ... + Z_k^2$ obéit alors a une loi du khi-carré k degrés de liberté notée $\varkappa_k^2$, et sa fonction de densité est :*

$$f(u) = \begin{cases} \frac{1}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} u^{\left(\frac{k}{2}-1\right)} \exp(-\frac{u}{2}) & si\ u \succ 0, \\ 0 & sinon. \end{cases}$$

La moyenne et la variance de la loi khi-carré :

La moyenne se traduit par :

$$\mu = k,$$

et sa variance par :

$$\delta^2 = 2k.$$

La fonction de répartition :

La fonction de répartition F est :

$$F(u) = P(X \leq u) = \int_0^u \frac{1}{2^{\frac{k}{2}}} u^{\frac{k}{2}-1} \exp(\frac{-u}{2}) dx.$$

1.2.9 La loi de Student

Théorème 1.2.2 *Soit Z une v.a suivant une loi $N(0,1)$ et V une v.a suivant une loi de $\varkappa_k^2$. Si ces deux variables sont indépendantes, alors la v.a :*

$$T = \frac{Z}{\sqrt{\frac{V}{k}}},$$

obéit à une loi t avec k degrés de liberté, notée t_k, et sa fonction de densité est :

$$f(t) = \frac{\Gamma[(k+1)/2]}{\sqrt{\pi k}\Gamma(k/2)} \times \frac{1}{[(t^2/k)+1]^{(k+1)/2}}, \qquad -\infty \prec t \prec +\infty.$$

1.2.10 La loi de de Fisher snedecor

Définition 1.2.1 *Soient Q_1 et Q_2 deux v.a indépendents telles que Q_1 suit $\varkappa^2(v_1)$ et Q_2 suit $\varkappa^2(v_2)$ alors la v.a*

$$F = \frac{Q_1/v_1}{Q_2/v_2}$$

suit une loi de Fisher snedcor à (v_1, v_2) degrés de liberte, notee $F(v_1, v_2)$.

Proposition 1.2.1 *La densité de la loi $F(v_1, v_2)$ est :*

$$f(x) = \begin{cases} \frac{\Gamma(\frac{v_1+v_2}{2})}{\Gamma(v_1/2)\Gamma(v_2/2)} (\frac{v_1}{v_2})^{v_1/2} \times \frac{x^{v_1/2-1}}{(1+\frac{v_1}{v_2}x)^{\frac{v_1+v_2}{2}}}, & si\ x \succ 0 \\ 0 & si\ non \end{cases}$$

son esperance n'existe que si $v_2 \geq 3$ et vaut $\frac{v_2}{v_2-2}$. Sa variance n'existe que si $v_2 \geq 5$ et vaut $\dfrac{2v_2^2(v_1+v_2-2)}{v_1(v_2-2)^2(v_2-4)}$.

Chapitre 2

Les méthodes d'estimation

L'estimation paramétrique fait référence à un processus statistique utilisé pour estimer les valeurs inconnues des variables paramétriques et déterminer les relations entre elles. Cela se fait généralement en utilisant deux types d'estimation, à savoir la méthode ponctuelle comprenant les méthodes des moments et les méthodes du maximum de vraisemblance, ainsi que la méthode de l'intervalle de confiance. C'est ce qui sera discuté ici.

Estimation :

L'estimation consiste à rechercher la valeur numérique d'un ou plusieurs paramètres inconnus d'une loi de probabilité a partir d'observation (valeurs prises par la v.a qui suit cette loi de probabilité).

Estimateur :

Un estimateur est une fonction statistique utilisée pour estimer un paramètre inconnu d'une population a partir d'un échantillon. Il s'agit d'une formule mathématique qui prend les donnes de l'échantillon en entrée et produit une estimation du parametre. L'estimation est utilise pour obtenir une valeur approximative du paramétre inconnu, en se basant sur les informations disponibles dans l'échantillon. Il existe différents types d'estimateurs, tel que les estimateurs ponctuelle et par intervalle de confiance, qui permettent d'estimer différent caractéristiques de la population, comme la moyenne, la variance, la proportion,...,etc.

2.1 Propriétés des estimateurs

Du retour à la référence [4]

Les estimateurs peuvent presentes differentes propriétés, telles que la biais, la convergence,...etc. De plus, tout estimateur peut etre exprime par l'equation :

$$E\left(T_n\right) = \theta + B\left(n, \theta\right)$$

tel que : T_n designe l'estimateur du paramétre θ ou $B\left(n, \theta\right)$ est le biais de T_n.

Estimateur sans biais :

On dit que T_n est un estimateur sans biais de θ si :

$$E(T_n) = \theta \text{ c à d } B\left(n, \theta\right) = 0.$$

Romarque : Un estimateur est asymptotiquement sans biais si $E(T_n) \rightarrow \theta$ lorsque $n \rightarrow \infty$.

Estimateur avec biais :

Un estimateur T_n est considéré biaisé pour le paramètre θ si :

$$E(T_n) \neq \theta \text{ c à d } B\left(n, \theta\right) = E(T_n - \theta) = E\left(T_n\right) - \theta \neq 0.$$

Convergence :

Un estimateur T_n est dit convergent si :

$$\forall \varepsilon > 0, P(|\hat{\theta} - \theta| > \varepsilon) \rightarrow 0 \; lorsque \; n \rightarrow \infty$$

Ceci équivaut à dire qu'en limite $\hat{\theta} \rightarrow \theta$ lorsque $n \rightarrow \infty$.

2.2 statistique exhaustive

Du retour à la référence [3]

Soit n échantillon d'une v.a X. La statistique T sera dite exhaustive pour θ si la conditionnelle de X sachant $T(x) = t$ n'est pas une fonction du paramétre θ : $P_\theta(x \backslash T(x) = t)$ ne depende pas de θ.

Thèoreme de factorisation :

Soit le modèle $(X. \; P_\theta$ et T une statistique $\left(X^n, B_n\right) \longrightarrow (Y, C))$. T est exhaustive pour θ si et seulment s'il existe deux fonctions mesurables $g : X \longrightarrow \mathbb{R}_+$

et $h : Y \longrightarrow \mathbb{R}_+$ telles que $f\left(x, \theta\right) = h\left(x\right) g\left(T\left(x\right), \theta\right)$ où $x = \left(x_1, x_2, ..., x_n\right)$.

statistique exhaustive minimale :

Une statistique T est dit exhaustive minimale pour θ si elle est exhaustive et si pour toute autre statistique que exhaustive S pour θ, il existe une application φ telle que $T = \varphi o S$.

2.3 Information de fisher

On appelle information de fisher au point θ la matrice :

$$I(\theta) = E\left(S(X,\theta)\, S(X,\theta)^t\right) = E\left[\left(\frac{\partial L(X,\theta)}{\partial \theta}\right)^2\right].$$

Théorème 2.3.1 *Si le domaine de définition de X ne dépende pas du paramètre θ alors :*

$$I(\theta) = -E\left[\left(\frac{\partial S(X,\theta)}{\partial \theta}\right)\right] = -E\left[\left(\frac{\partial^2 L(X,\theta)}{\partial \theta^2}\right)^2\right].$$

Preuve. Voir [7], page 295-296. ■

2.4 Estimation ponctuelle

La méthode d'estimation ponctuelle consiste à trouver une valeur approximative d'un paramètre de la population en se basant sur les résultats d'un échantillon, representée par un estimateur qui depend des donneés de l'échantillon. Constriarement a l'estimation par intervalle de confiance que fournit également une indication de la précision de l'estimation. L'estimation ponctuelle ne fournit qu'une seule valeur.

2.4.1 Methode de maximum de vraisemblance

La methode de maximum de vraisemblance(MMV) est la meilleurs methode d'estimation, cette méthode consiste à prendre comme estimation de θ la valeur qui maximise la vraisemblance $L(x_1, ..., x_n; \theta)$ étant donné un échantillon de valeurs $(x_1, ..., x_n)$.

Définition 2.4.1 *[4]On appelle fonction de vraisemblance de θ pour une réalisation $(x_1, ..., x_n)$*

d'un échantillon, la fonction de θ :

$$L(x_1, ..., x_n; \theta) = \begin{cases} \prod_{i=1}^{n} f(x_i; \theta) & \text{si continue} \\ \prod_{i=1}^{n} P(x_i; \theta) & \text{si discrètes} \end{cases}$$

Définition 2.4.2 *On appelle estimateur du maximum de vraisemblance (EMV) du paramètre θ la statistique $\hat{\theta}_n$ rendant maximale, selon θ, la fonction de vraisemblance du modele $L(x_1, ..., x_n; \theta)$, soit :*

$$\hat{\theta}_n = \arg\max_{\theta \in \ominus} L(X, \theta)$$

Exemple 2.4.1 ***(Cas continues)***

Soit X une v.a suit la loi normale(μ, δ^2), telle que sa densite est :

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} \exp(-\frac{1}{2\delta^2}(x - \mu)).$$

Pour appliquer MMV on va trouver la fonction de vraisemblance de l'echantillon

$X = (x_1, ..., x_n)$:

$$\begin{aligned} L(x_1, ..., x_n; \mu, \delta^2) &= \prod_{i=1}^{n} f(x_i) \\ &= \prod_{i=1}^{n} \frac{1}{\delta\sqrt{2\pi}} \exp(-\frac{1}{2\delta^2}(x_i - \mu)^2) \\ &= \frac{1}{\delta\sqrt{2\pi}} \exp(-\frac{1}{2\delta^2}(x_1 - \mu)^2) ... \frac{1}{\delta\sqrt{2\pi}} \exp(-\frac{1}{2\delta^2}(x_n - \mu)^2) \\ &= (\frac{1}{\delta\sqrt{2\pi}})^n \exp\left(-\frac{1}{2\delta^2}\left[(x_1 - \mu)^2 ... (x_n - \mu)^2\right]\right) \\ &= (\frac{1}{\delta\sqrt{2\pi}})^n \exp(-\frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i - \mu)^2). \end{aligned}$$

Maintenant on va trouver la log-vraisemblance :

$$\begin{aligned}
\log(L(x_1,...,x_n;\mu,\delta^2)) &= \log\left[(\frac{1}{\delta\sqrt{2\pi}})^n \exp(-\frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2))\right] \\
&= \log\left[\left(\frac{1}{\delta\sqrt{2\pi}}\right)^n\right] + \log\left[\exp\left(-\frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right] \\
&= n\log\left(\frac{1}{2\sqrt{2\pi}}\right) - \frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&= n\log\left[\left(\left(\delta^2\right)^{1/2}(2\pi)^{1/2}\right)^{-1}\right] - \frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&= -\frac{n}{2}\log\left(\delta^2\right) - \frac{n}{2}\log(2\pi) - \frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2 .
\end{aligned}$$

Alors :

$$\begin{aligned}
&\frac{\partial}{\partial\mu}\log L(x_1,...,x_n;\mu) = 0 \\
&\qquad\Rightarrow \frac{\partial}{\partial\mu}(-\frac{n}{2}\log\left(\delta^2\right) - \frac{n}{2}\log(2\pi) - \frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2) = 0 \\
&\qquad\Rightarrow \frac{\partial}{\partial\mu}(-\frac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2) = -\frac{1}{2\delta^2}\sum_{i=1}^{n}2(-1)(x_i-\mu) = 0 \\
&\qquad\Rightarrow \frac{1}{\delta^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \\
&\qquad\Rightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \\
&\qquad\Rightarrow \sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\mu = 0 \\
&\qquad\Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n}x_i}{n} = \bar{X}
\end{aligned}$$

et

$$\begin{aligned}
\frac{\partial}{\partial \delta^2} \log L(x_1, ..., x_n; \delta^2) = 0 \\
&\Rightarrow \frac{\partial}{\partial \delta^2}(-\frac{n}{2}\log\left(\delta^2\right) - \frac{n}{2}\log\left(2\pi\right) - \frac{1}{2\delta^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2) = 0 \\
&\Rightarrow \frac{\partial}{\partial \mu}(-\frac{n}{2}\log\left(\delta^2\right) - \frac{1}{2\delta^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2) = 0 \\
&\Rightarrow -\frac{n}{2}\frac{1}{\delta^2} - \frac{2}{(2\delta^2)^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2) = 0 \\
&\Rightarrow -\frac{1}{2\delta^2}(n - \frac{1}{\delta^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2) = 0 \\
&\Rightarrow n - \frac{1}{\delta^2}\sum_{i=1}^{n}\left(x_i - \mu\right)^2 = 0 \\
&\Rightarrow \hat{\delta}^2 = \frac{\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}{n}
\end{aligned}$$

Maintenant on va appliquer quelques propriétés pour évaluer la précision des résultats

La biaisi :

$$\begin{aligned}
E(\hat{\mu}) = E\left(\frac{\sum x_i}{n}\right) = E\left(\frac{x_1 + x_2 + ... + x_n}{n}\right) = \frac{E(x_1) + E(x_2) + ... + E(x_n)}{n} \\
&\Rightarrow \frac{E(x) + E(x) + ... + E(x)}{n} = \frac{nE(x)}{n} = E(x) = \mu \\
&\Rightarrow E(\hat{\mu}) = \mu
\end{aligned}$$

D'où $\hat{\mu}$ *est un estimateur sans biais.*

La convergence :

$$\begin{aligned}
var(\hat{\mu}) &= var\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2}var(x_1 + x_2 + ... + x_n) \\
&= \frac{1}{n^2}\left(var(x_1) + ... + var(x_n)\right) = \frac{1}{n^2}(var(x) + ... + var(x)) \\
&= \frac{1}{n^2}nvar(x) = \frac{\delta^2}{n}
\end{aligned}$$

$$\lim_{n\to\infty} var(\hat{\mu}) = \lim_{n\to\infty} \frac{\delta^2}{n} = 0.$$

D'ou on conclure que $\hat{\mu}$ est convergent.

L'effecacite :

$$\begin{aligned}
BCR(\mu) &= \frac{1}{I(\mu)} \\
\frac{\partial}{\partial\mu} L(x_1, ..., x_n; \theta) &= \frac{1}{\delta^2} \sum_{i=1}^{n} (x_i - \mu) \\
\frac{\partial^2}{\partial\mu^2} L(x_1, ..., x_n; \theta) &= \frac{\partial}{\partial\mu} \left(\frac{1}{\delta^2} \sum_{i=1}^{n} (x_i - \mu) \right) = \frac{1}{\delta^2} \sum_{i=1}^{n} (-1) = \frac{-n}{\delta^2} \\
E\left(\frac{-n}{\delta^2}\right) &= \frac{-n}{\delta^2} \\
I(\mu) &= -E\left(\frac{-n}{\delta^2}\right) = \frac{n}{\delta^2} \\
BCR(\mu) &= \frac{1}{I(\mu)} = \frac{\delta^2}{n} = var(\hat{\mu})
\end{aligned}$$

D'où on conclure que $\hat{\mu}$ est un estimateur efficace.

– Le même pour δ^2.

Exemple 2.4.2 ***(Cas continues)***

Soit X une v.a suit la loi normale exponentielle(λ), telle que :

$$\begin{aligned}
L(x_1, ..., x_n; \lambda) &= \prod_{i=1}^{n} f(x_i) \\
&\Leftrightarrow \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \exp(-\lambda x_1 - ... - \lambda x_n) \\
&\Leftrightarrow \lambda^n \exp(-\lambda(x_1 - ... - x_n)) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i).
\end{aligned}$$

Maintenant on va trouver la log-vraisemblance :

$$\log(L(x_1, ..., x_n; \lambda)) = \log\left[\lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i)\right]$$
$$\Rightarrow \log(\lambda^n) + \log\left[\exp(-\lambda \sum_{i=1}^{n} x_i)\right] = n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

Alors :

$$\frac{\partial}{\partial\lambda} \log L(x_1, ..., x_n; \lambda) = \frac{\partial}{\partial\lambda}\left(n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i\right) = 0$$
$$\Leftrightarrow \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$
$$\Leftrightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

Exemple 2.4.3 ***(Cas discretes)***

Soit X une v.a suit la loi bernoulli (p), telle que :

$$L(x_1, ..., x_n; p) = \prod_{i=1}^{n} P(x_i)$$
$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

on pose $t = \sum_{i=1}^{n} x_i$

donc

$$L(x_1, ..., x_n; p) = p^t(1-p)^{n-t}$$

Maintenant on va trouver la log-vraisemblance :

$$\begin{aligned}
\log(L(x_1, ..., x_n; p)) &= \log\left(p^t(1-p)^{n-t}\right) \\
&= \log(p^t) + \log((1-p)^{n-t}) \\
&= t \log p^t + (n-t)\log(1-p).
\end{aligned}$$

Alors

$$\begin{aligned}
\frac{\partial}{\partial p} \log L(x_1, ..., x_n; p) &= \frac{\partial}{\partial p}(t \log p^t + (n-t)\log(1-p)) = 0 \\
&\Rightarrow \frac{t}{p} - \frac{n-t}{1-p} = 0 \\
&\Rightarrow \frac{t(1-p) - p(n-t)}{p(1-p)} = 0 \\
&\Rightarrow t(1-p) - p(n-t) = 0 \\
&\Rightarrow t - pn = 0 \\
&\Rightarrow \hat{p} = \frac{t}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{X}.
\end{aligned}$$

2.4.2 Méthodes des moments

La méthode d'estimation des moments (MM) consiste à égaliser les moments théoriques aux moments empiriques des données pour obtenir les estimations des paramètres d'une distribution, en résolvant les équations basées sur les moments théorique et empirique.

Soit X une v.a qui suit une loi de probabilité dépendant d'un paramètre $\theta = (\theta_1, ..., \theta_2)$ donc l'estimateur de θ par la mèthode des moments est :

$$\begin{cases} m_1 = E\left(x\right) = g_1(\theta_1, ..., \theta_k) \\ m_2 = E\left(x^2\right) = g_2(\theta_1, ..., \theta_k) \\ \vdots \\ m_k = E\left(x^k\right) = g_k(\theta_1, ..., \theta_k) \end{cases} \Leftrightarrow \begin{cases} \theta_1 = f_1(m_1, ..., m_k) \\ \theta_2 = f_2(m_1, ..., m_k) \\ \vdots \\ \theta_k = f_k(m_1, ..., m_k) \end{cases}$$

telle que les moments empirique est :

$$\begin{cases} m_1 = \frac{1}{n}\sum_{i=1}^{n} X \\ m_2 = \frac{1}{n}\sum_{i=1}^{n} X^2 \\ \vdots \\ m_k = \frac{1}{n}\sum_{i=1}^{n} X^k \end{cases}$$

Alors :

$$\begin{cases} \hat{\theta}_1 = f(\hat{m}_1, ..., \hat{m}_k) \\ \hat{\theta}_2 = f_2(\hat{m}_1, ..., \hat{m}_k) \\ \vdots \\ \hat{\theta}_k = f_k(\hat{m}_1, ..., \hat{m}_k) \end{cases}$$

Exemple 2.4.4 ***(Cas continues)***

Soit X une v.a suit la loi gamma de paramétre r et λ, telle que :

$$\begin{cases} m_1 = E(x) = \frac{r}{\lambda} \\ m_2 = E(x^2) = \frac{r - r^2}{\lambda^2} \end{cases} \Rightarrow \begin{cases} r = \lambda E(x) \\ \lambda = var\left(x\right) = \frac{\lambda E(x)}{\lambda^2} = \frac{E(x)}{\lambda} \end{cases}$$

$$\Rightarrow \begin{cases} \lambda = \frac{E(x)}{var(x)} = \frac{m_1}{m_2 - m_1} \\ r = \frac{E(x)^2}{var(x)} = \frac{m_1^2}{m_2 - m_1} \end{cases}$$

et donc on va remplacer le moment théorique par le moment empirique :

$$\begin{cases} \lambda = \frac{E(x)}{var(x)} = \frac{\hat{m}_1}{\hat{m}_2 - \hat{m}_1} \\ r = \frac{E(x)^2}{var(x)} = \frac{\hat{m}_1^2}{\hat{m}_2 - \hat{m}_1} \end{cases} \Rightarrow \begin{cases} \hat{\lambda} = \frac{\bar{X}}{\frac{1}{n}\sum X - \bar{X}^2} = \frac{\bar{X}}{S^2} \\ r = \frac{\bar{X}^2}{\frac{1}{n}\sum X - \bar{X}^2} = \frac{\bar{X}^2}{S^2} \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\lambda} = \frac{\bar{X}}{S^2} \\ \hat{r} = \frac{\bar{X}^2}{S^2} \end{cases}.$$

Exemple 2.4.5 ***(Cas discrétes)***

Soit X une v.a suit la loi de poisson de paramétre c, telle que :

$$m = E(x) = c$$

$$\Rightarrow c = m_1$$

et donc on va remplacer le moment théorique par le moment empirique :

$$c = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\Rightarrow \hat{c} = \bar{X}$$

Remarque 2.4.1 *Si la distribution de la v.a X est connue on utilise la méthode de maximum de vraisemblance ou bien la méthode de moment, et si la distribution n'est pas connue on utilise la méthode de moinders carrès.*

2.5 Estimation par intervalle de confiance

Dans de nombreux cas, l'estimation ponctuelle ne fournit pas suffisamment d'informations sur le paramètre, car elle donne une valeur numérique unique pour cette variable sans prendre en compte les erreurs dues aux fluctuations de l'échantillonnage. L'intervalle de confiance estimé par la formule $\theta_1 \leq \theta \leq \theta_2$ pourrait être plus utile où θ_1 et θ_2 sont des v.a.

Définition 2.5.1 *Soit $\alpha \in [0,1]$. On appelle intervalle de confiance du paramètre θ de niveau (de confiance) $1-\alpha$ la donnée de deux statistiques θ_1 et θ_2 vèrifiant :*

$$P(\theta_1 \leq \theta \leq \theta_2) = 1 - \alpha.$$

2.5.1 Intervalles de confiance pour les paramètres gaussiens

Soit $(X_1, ..., X_n)$ un échantillon d'une v.a qui suit la loi normale de paramètres μ et δ^2. On souhaite construire des IC pour ces dernier.

Remarque 2.5.1 – *La construction de l'intervalle de confiance dépend des valeurs α à α_1 et α_2, ainsi que de leur choix en fonction des problèmes à résoudre, qui se présentent en deux cas :*

1. *Intervalles de confiance bilatérale où $\alpha_1 = \alpha_2 = \frac{\alpha}{2} \neq 0$ et sont habituellement utilisés dans certaines lois symétriques.*
2. *Intervalles de confiance unilatérale où l'on prend $\alpha_1 = 0$ ou $\alpha_2 = 0$ soit $[\theta_1, +\infty[$ lorsque $\alpha_1 = \alpha$ et $\alpha_2 = 0$, et $]-\infty, \theta_2]$ lorsque $\alpha_2 = \alpha$ et $\alpha_1 = 0$.*

– *Les valeurs de risque de α habituelles sont de 0.1, 0.05 et 0.01, correspondant respectivement aux niveaux de confiance de 90%, 95% et 99%.*

– *Pour déterminer l'intervalle de confiance, il faut connaître l'estimateur des paramètres ainsi que la loi de distribution associée.*

Estimation de la moyenne

1. **Si la variance est connue :**

   Lorsque X suit une loi normale, la moyenne de $\bar{X}$ corresponde à μ et sa variance à $\frac{\delta^2}{n}$. En consequence, la statistique

$$Z = \sqrt{n}\frac{\bar{X} - \mu}{\delta}$$

suit une loi normale centée reduit, c'est-à-dire une loi $N(0,1)$.

Alors,

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

où $Z_{1-\frac{\alpha}{2}}$ est le quantille d'ordre $1-\frac{\alpha}{2}$ de la loi $N(0,1)$, c'est-à-dire :

$$P\left(N(0,1) \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}$$
$$\Leftrightarrow P\left(\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\delta}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{1-\frac{\alpha}{2}}\frac{\delta}{\sqrt{n}}\right)$$

Par consèquente, l'IC est :

$$IC_{1-\alpha}(\mu) = \left[\bar{X} - Z_{1-\frac{\alpha}{2}}\frac{\delta}{\sqrt{n}}, \bar{X} + Z_{1-\frac{\alpha}{2}}\frac{\delta}{\sqrt{n}}\right].$$

Exemple 2.5.1 *Supposons que la force de compression pour un type de béton soit une v.a suivant la loi normale $N(\mu, \delta^2)$, où $\delta = 47.28Mpa$, comme l'ont montré les tests réalisés sur 20 échantillons $\bar{X} = 2276.75Mpa$, l'IC au niveau de confiance $1-\alpha = 0.95$ est :*

$$P\left(2276.75 - 1.96\frac{47.28}{\sqrt{20}} \leq \mu \leq 2276.75 + 1.96\frac{47.28}{\sqrt{20}}\right)$$
$$= P(2256.028 \leq \mu \leq 2297.471)$$

Donc, l'IC est :

$$IC_{1-\alpha}(\mu) = [2256.028; 2297.471].$$

Remarque 2.5.2 *Si $n \succ 30$ et δ est inconnu en remplace cet dernier par $S' = \sqrt{\frac{n}{n-1}}S$.*

2. **Si la variance est inconnue**$(n \prec 30)$ **:**

   Dans ce cas on utilise la relation $T = \frac{\bar{X}-\mu}{S/\sqrt{n-1}} \sim \tau_{n-1}$.

Alors :

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

où $t_{1-\frac{\alpha}{2}}$ est le quantille d'ordre $1-\frac{\alpha}{2}$ de la loi student avec $(n-1)$ degré de libertè, c'est-à-dire :

$$P\left(\tau_{n-1} \leq t_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}$$
$$\Leftrightarrow P\left(\bar{X}-\frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}} \leq \mu \leq \bar{X}+\frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}}\right)$$

Par consèquente, l'IC est :

$$IC_{1-\alpha}(\mu) = \left[\bar{X}-\frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}}, \bar{X}+\frac{S}{\sqrt{n-1}}t_{1-\frac{\alpha}{2}}\right].$$

Exemple 2.5.2 *Lors d'une expérience menée sur* 15 *batteries pour déterminer leur durée de vie, nous avons obtenu les moments expérimentaux* $\bar{X} = 360h$ *et* $S = 28h$, *l'*IC *au niveau de confiance* $1-\alpha = 0.95$ *est :*

$$P\left(360-2.093\frac{28}{\sqrt{14}} \leq \mu \leq 360+2.093\frac{28}{\sqrt{14}}\right)$$
$$= P(344.337 \leq \mu \leq 375.663)$$

Donc, l'IC *est :*

$$IC_{1-\alpha}(\mu) = [344.337; 375.663].$$

Estimation de la variance

1. **Si la moyenne est connue :**

   Si X_k suit une loi $N(\mu, \delta^2)$, alors :

$$P\left(u_1 \leq \frac{nk}{\delta^2} \leq u_2\right) = 1-\alpha$$

où $u_1$est le quantille d'ordre $\frac{\alpha}{2}$ et u_2 est le quantille d'ordre $1-\frac{\alpha}{2}$de la loi $\varkappa_n^2$, c'est-à-dire :

$$P\left(\varkappa_n^2 \leq u_1\right) = P\left(\varkappa_n^2 \geq u_2\right) = \frac{\alpha}{2}$$
$$\Leftrightarrow P\left(\frac{nk}{u_2} \leq \delta^2 \leq \frac{nk}{u_1}\right) = 1-\alpha.$$

Par consèquente, l'IC est :

$$IC_{1-\alpha}\left(\delta^2\right) = \left[\frac{nk}{u_2}, \frac{nk}{u_1}\right].$$

Exemple 2.5.3 *Soit X une v.a suit la loi $N(24, \delta^2)$, cherchons un IC au niveau de confiance $1-\alpha = 0.90$, pour un echantillon de taille $n = 25$ et $k = 13$.*

On a $u_1 = \varkappa^2_{\frac{\alpha}{2}} = \varkappa^2_{0,05} = 36.415$ et $u_2 = \varkappa^2_{1-\frac{\alpha}{2}} = \varkappa^2_{0,95} = 14.611$, alors

$$P\left(\frac{25\times 13}{37.652} \leq \delta^2 \leq \frac{25\times 13}{14.611}\right)$$
$$= P\left(8.631 \leq \delta^2 \leq 22,243\right).$$

Donc, l'IC est :

$$IC_{1-\alpha}(\delta^2) = [8.631; 22,243]$$

2. **Si la moyenne est inconnue :**

   Dans ce cas, on utilise $S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ l'estimateur de δ^2 par la variance empirique et on sait que $\frac{nS^2}{\delta^2} \sim \varkappa^2_{n-1}$.

   Alors,
$$P\left(T_1 \leq \frac{nS^2}{\delta^2} \leq T_2\right) = 1-\alpha$$

   où T_1est le quantille d'ordre $\frac{\alpha}{2}$ et T_2 est le quantille d'ordre $1-\frac{\alpha}{2}$de la loi $\varkappa^2_{n-1}$, c'est-à-dire :

$$P\left(\varkappa_{n-1}^2 \leq T_1\right) = P\left(\varkappa_{n-1}^2 \geq T_2\right) = \frac{\alpha}{2}$$
$$\Leftrightarrow P\left(\frac{nS^2}{T_2} \leq \delta^2 \leq \frac{nS^2}{T_1}\right) = 1-\alpha.$$

Par consèquente, l'IC est :

$$IC_{1-\alpha}\left(\delta^2\right) = \left[\frac{nS^2}{T_2}, \frac{nS^2}{T_1}\right].$$

Exemple 2.5.4 *Soit X une v.a suit la loi $N(\mu, \delta^2)$, cherchons un IC au niveau de confiance $1-\alpha = 0.95$, pour un echantillon de taille $n = 26$ et $S^2 = 14$.*

On a $T_1 = \varkappa_{\frac{\alpha}{2}}^2 = \varkappa_{0,025}^2 = 40.646$ et $T_2 = \varkappa_{1-\frac{\alpha}{2}}^2 = \varkappa_{0,975}^2 = 13.119$, alors

$$P\left(\frac{26 \times 14}{40.646} \leq \delta^2 \leq \frac{26 \times 14}{13.119}\right)$$
$$= P\left(8.955 \leq \delta^2 \leq 27.746\right).$$

Donc, l'IC est :

$$IC_{1-\alpha}(\delta^2) = [8.955; 27.746].$$

Remarque 2.5.3 *Si $n > 30$ alors nous pouvons utiliser les deux approximations suivantes :*

Approximation de fisher

$$\sqrt{2}\varkappa_p^2 - \sqrt{2p-1} \sim N(0,1) \quad si\ p \succ 30$$

Approximation de Wilson Hilferty

$$\varkappa_p^2 = p\left(u\sqrt{\frac{2}{9p}} + 1 - \frac{2}{9p}\right)^3$$

valable même pour les valeurs faible de p.

2.5.2 Intervalles de confiance à de deux échantillons des paramètres gaussiens

considérons les échantillons $(X_1, ..., X_2)$ et $(Y_1, ..., Y_2)$ suivant la loi $N(\mu_1, \delta_1^2)$ et $N(\mu_2, \delta_2^2)$ respctivement, ces deux échantillons sont independants, notre objectif est de comparer les moyennes μ_1 et μ_2, ainsi que les variances δ_1^2 et δ_2^2, en utilisant ces échantillons. Pour ce faire nous allons construire des IC pour $\mu_1 - \mu_2$ et pour δ_1^2 et δ_2^2.

Intervalles de confiance de la difference de deux moyennes

On pose

$$\bar{X} \sim N\left(\mu_1, \frac{\delta_1^2}{n_1}\right) \qquad \text{et} \qquad \bar{Y} \sim N\left(\mu_2, \frac{\delta_2^2}{n_2}\right)$$

Alors :

$$B = \bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}\right)$$

Remarque 2.5.4 *L'estimateur B est un estimateur sans biais.*

1. **si l'ecart-type est connus (δ_1^2 et δ_2^2 connus)**

   un IC de $\mu_1 - \mu_2$ du niveau de confiance $1 - \alpha$ est :

$$P\left(-Z_{1-\frac{\alpha}{2}} \leq \frac{B - (\mu_1 - \mu_2)}{\sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(B - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq B + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\delta_1^2}{n_1} + \frac{\delta_2^2}{n_2}}\right) = 1 - \alpha.$$

   Où $Z_{1-\frac{\alpha}{2}}$ est le quantille d'ordre $1 - \frac{\alpha}{2}$ de la loi $N(0, 1)$.

Alors, l'IC est :

$$IC_{1-\alpha}(\mu_1-\mu_2)=\left[B-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\delta_1^2}{n_1}+\frac{\delta_2^2}{n_2}},B+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\delta_1^2}{n_1}+\frac{\delta_2^2}{n_2}}\right].$$

2. **si l'ecart-type est inconnus (δ_1^2 et δ_2^2 inconnus)**

Il y a deux cas :

. *Ecarts-types égaux et inconnus ($\delta_1^2=\delta_1^2=\delta^2$)*

On a :

$$\frac{B-(\mu_1-\mu_2)}{\sqrt{\frac{\delta_1^2}{n_1}+\frac{\delta_2^2}{n_2}}}=\frac{B-(\mu_1-\mu_2)}{\delta\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$$

et

$$S_x^2=\frac{1}{n_1-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 \qquad S_y^2=\frac{1}{n_2-1}\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2$$

$$\text{et}\quad S_{xy}^2=\frac{1}{n_1+n_2-2}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)=\frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}.$$

Puis que δ est inconnu et si $n\geq 30$ en remplace δ^2 par S_{xy}^2, et on a :

$$\frac{(n_1-1)S_x^2}{\delta^2}\sim\varkappa_{n_1-1}^2 \qquad \text{et} \qquad \frac{(n_2-1)S_y^2}{\delta^2}\sim\varkappa_{n_2-1}^2$$

Alors,

$$n_1+n_2-2\frac{S_{xy}^2}{\delta^2}=\frac{(n_1-1)S_x^2}{\delta^2}+\frac{(n_2-1)S_y^2}{\delta^2}=\frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}\sim\varkappa_{n_1+n_2-2}^2.$$

Donc,

$$\frac{B-(\mu_1-\mu_2)}{S_{xy}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim\tau_{n_1+n_2-2}.$$

Alors, l'IC de $\mu_1 - \mu_2$ du niveau de confiance $1-\alpha$ est :

$$P\left(-t_{1-\frac{\alpha}{2}} \leq \frac{B-(\mu_1-\mu_2)}{S_{xy}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \leq t_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Où $t_{1-\frac{\alpha}{2}}$ est le quantille d'ordre $1-\frac{\alpha}{2}$ de la loi $\tau_{n_1+n_2-2}$.

$$\Leftrightarrow P\left(B - t_{1-\frac{\alpha}{2}}S_{xy}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \leq \mu_1-\mu_2 \leq B + t_{1-\frac{\alpha}{2}}S_{xy}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right) = 1-\alpha.$$

Alors, l'IC est :

$$IC_{1-\alpha}(\mu_1-\mu_2) = \left[B - t_{1-\frac{\alpha}{2}}S_{xy}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, B + t_{1-\frac{\alpha}{2}}S_{xy}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right].$$

. *Ecarts-types quelconques ($\delta_1^2 \neq \delta_1^2$)*

Si les échantillons ont des taille importantes et égales à $n_1 = n_2 = n \succ 30$, on a :

$$P\left(-Z_{1-\frac{\alpha}{2}} \leq \frac{\sqrt{n}\,(B-(\mu_1-\mu_2))}{\sqrt{S_x^2+S_y^2}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

$$\Leftrightarrow P\left(B - Z_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sqrt{S_x^2+S_y^2} \leq \mu_1-\mu_2 \leq B - Z_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sqrt{S_x^2+S_y^2}\right) = 1-\alpha$$

Alors, l'IC est :

$$IC_{1-\alpha}(\mu_1-\mu_2) = \left[B - Z_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sqrt{S_x^2+S_y^2}, B - Z_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{n}}\sqrt{S_x^2+S_y^2}\right]$$

IC **du rapport de deux variance**

On a

$$\bar{X} \sim N\left(\mu_1, \frac{\delta_1^2}{n_1}\right) \qquad \bar{Y} \sim N\left(\mu_2, \frac{\delta_2^2}{n_2}\right)$$

$$\text{et} \quad \frac{(n_1-1)\,S_x^2}{\delta^2} \sim \varkappa^2_{n_1-1} \qquad \frac{(n_2-1)\,S_y^2}{\delta^2} \sim \varkappa^2_{n_2-1}$$

Alors, d'apres l'application des résultats de fisher-snedecor nous trouvons :

$$\frac{S_x^2/\delta_1^2}{S_y^2/\delta_2^2} = \frac{\delta_1^2/\delta_2^2}{S_x^2/S_y^2} \sim F\left(n_1-1, n_2-1\right).$$

D'où :

$$P\left(F_{\frac{\alpha}{2}}\left(n_1-1, n_2-1\right) \leq \frac{S_x^2/\delta_1^2}{S_y^2/\delta_2^2} \leq F_{1-\frac{\alpha}{2}}\left(n_1-1, n_2-1\right)\right) = 1-\alpha.$$

Nous avons que $F_{\frac{\alpha}{2}}\left(n_1-1, n_2-1\right) = \frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)}$

donc :

$$P\left(\frac{S_x^2}{S_y^2}\frac{1}{F_{1-\frac{\alpha}{2}}\left(n_1-1, n_2-1\right)} \leq \frac{\delta_1^2}{\delta_2^2} \leq \frac{S_x^2}{S_y^2}\frac{1}{F_{\frac{\alpha}{2}}\left(n_1-1, n_2-1\right)}\right) = 1-\alpha.$$

Alors, l'IC est :

$$IC_{1-\alpha}\left(\frac{\delta_1^2}{\delta_2^2}\right) = \left[\frac{S_x^2}{S_y^2}\frac{1}{F_{1-\frac{\alpha}{2}}\left(n_1-1, n_2-1\right)}, \frac{S_x^2}{S_y^2}\frac{1}{F_{\frac{\alpha}{2}}\left(n_1-1, n_2-1\right)}\right].$$

2.5.3 IC de paramètre p d'une loi binomiale(proportion)

Considérons X comme une v.a suivant une loi $B(n,p)$ et n comme la taille de la population. La statistique $T = \frac{X}{n}$ est notée comme la meilleur estimateur sans biais de p. De plus, la distribution de l'échantillon p sera approximativement normale avec une moyenne p et une variance $\frac{p(1-p)}{n}$, pour autant que la valeur de p ne soit pas trop proche de 0 ou 1, et que la valeur de n soit suffisamment grand (d'aprés le théorème central limite). En conséquence

$$\frac{T-p}{\sqrt{\frac{p(1-p)}{n}}} \overset{L}{\sim} N(0,1)$$

alors,

$$P\left(-Z_{1-\frac{\alpha}{2}} \leq \frac{T-p}{\sqrt{\frac{p(1-p)}{n}}} \leq Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Leftrightarrow P\left(T - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} \leq p \leq T + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha.$$

Cela ne produit pas un IC car les limites dépendent de la valeur inconne p neànmois, nous obtenons le même résultat de convergence en remplacant p par son estimateur convergente T [pour plus de détails voir [7] page 310] . Ainsi, nous obtenons

$$P\left(T - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{T(1-T)}{n}} \leq p \leq T + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{T(1-T)}{n}}\right) = 1-\alpha.$$

Alors, l'IC est :

$$IC_{1-\alpha}(p) = \left[T - Z_{1-\frac{\alpha}{2}}\sqrt{\frac{T(1-T)}{n}}, T + Z_{1-\frac{\alpha}{2}}\sqrt{\frac{T(1-T)}{n}}\right].$$

Remarque 2.5.5 *En cas de petit valeur de n, les tables de la loi binomial ou un graphique seront utilisées.*

Exemple 2.5.5 *Dans un échantillon pris au hasard de 200 réservoir de capacités différenter on constant $X = 75$ d'entre eux ont une capacité supérieure à $1000L$. Pour calculer l'IC pour le pourcentage des réservoires ayant une capacité supérieure à $1000L$ au niveau de confiance $\alpha = 5\%$ nous commençons par calculer :*

* *L'estimateur sans biais pour la proportion p :*

$$p = \frac{X}{n} = \frac{75}{200} = 0,375.$$

alors, l'IC pour p est

$$P\left(0.375 - 1.96\sqrt{\frac{0.375\,(1-0.375)}{200}} \leq p \leq 0.375 + 1.96\sqrt{\frac{0.375\,(1-0.375)}{200}}\right) = 0.95$$

Donc,

$$IC_{0,95}\,(p) = [0.308, 0.442]\,.$$

2.6 Application sous R

Au cours de ce section, nous appliquerons les résultats obtenus précédemment sur les données de simulation afin de confirmer leur validité, en utilisant la distribution normale et la loi exponentielle pour ce faire.

Paramètres d'une population normale

Estimation de la moyenne

Si la variance connue :

On va présenter une illustration numérique et graphique de l'exemple qu'on déja donne2.5.1, les resultats numérique sont résumes dans le tableau 2.1.

$1-\alpha$	$b\_$ inf	$b\_$ sup
0.90	2259.36	2294.14
0, 95	2256.029	2297.471
0, 99	2249.518	2303.982

TAB. 2.1 – Intervalles de confiance, de differents niveaux, pour la moyenne si la variance connue

Et Les représentations graphiques sont présentées dans la figure2.1

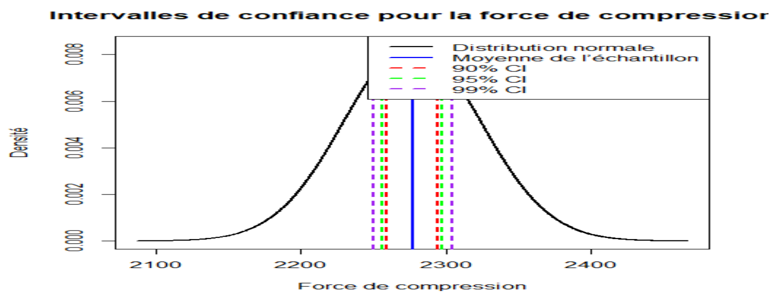

FIG. 2.1 – Intervalle de confiance pour la force de compression

Si la variance inconnue :

On va présenter une illustration numérique et graphique de l'exemple qu'on déja donne2.5.2, les resultats numérique sont résumes dans le tableau 2.2.

$1-\alpha$	$b\_$ inf	$b\_$ sup
0.90	346.8196	373.1804
0, 95	343.9499	376.0501
0, 99	337.7233	382.2767

TAB. 2.2 – Intervalles de confiance, de differents niveaux, pour la moyenne si la variance inconnue

Et Les représentations graphiques sont présentées dans la figure2.2

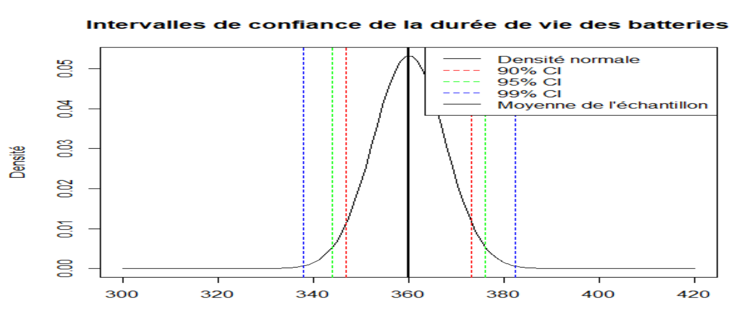

FIG. 2.2 – Intervalle de confiance de la durée de vie des batteries

Estimation de la variance

Si la moyenne connue :

On va présenter une illustration numérique et graphique de l'exemple qu'on déja donne2.5.3, les resultats numérique sont résumes dans le tableau 2.3.

$1-\alpha$	b_ inf	b_ sup
0.90	8.924	23.468
0, 95	8.256	26.207
0, 99	7.133	32.874

TAB. 2.3 – Intervalles de confiance, de differents niveaux, pour la variance si la moyenne connue

Et Les représentations graphiques sont présentées dans la figure2.3

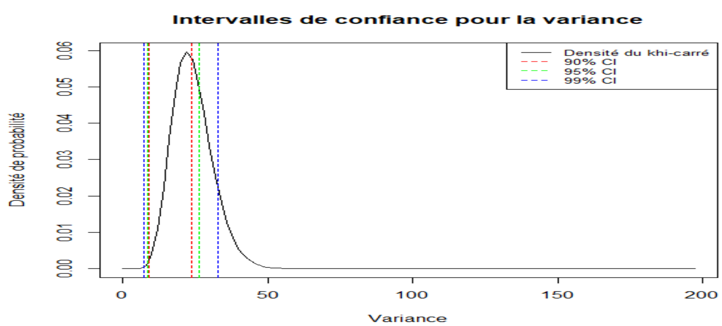

FIG. 2.3 – Intervalle de confiance de la variance si la moyenne connue

Si la moyenne inconnue :

On va présenter une illustration numérique et graphique de l'exemple qu'on déja donne2.5.4, les resultats numérique sont résumes dans le tableau 2.4.

$1-\alpha$	b_ inf	b_ sup
0.90	9.667	24.912
0, 95	8.955	27.744
0, 99	7.756	34.601

TAB. 2.4 – Intervalles de confiance, de differents niveaux, pour la variance si la moyenne inconnue

Et Les représentations graphiques sont présentées dans la figure2.4

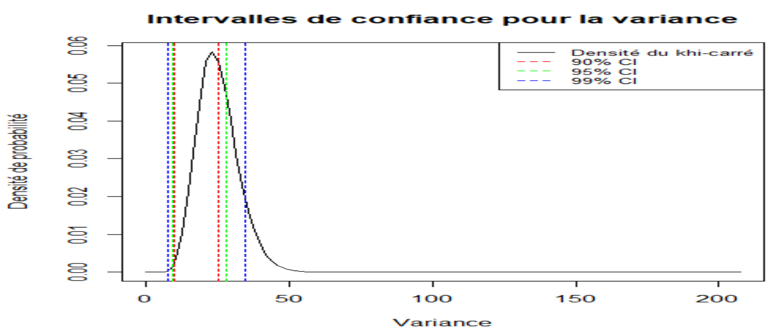

FIG. 2.4 – Intervalle de confiance pour la variance si la moyenne inconnue

Discussion :

Ces quatre exemples illustrent différentes applications de l'estimation des informations statistiques en utilisant les intervalles de confiance, avec des méthodes et des niveaux de confiance variés. On peut en conclure que plus le niveau de confiance est élevé, plus l'intervalle de confiance est large, ce qui signifie une plus grande confiance que la vraie valeur du paramètre se trouve dans l'intervalle, mais la précision est moindre. Ces exemples mettent en évidence la souplesse de l'utilisation des intervalles de confiance dans l'estimation, pouvant être employés pour estimer différentes informations statistiques. De plus, plus la taille de l'échantillon est grande, plus les intervalles de confiance sont petit.

Paramètres d'une population quelconque (exponentielle)

Dans un échantillon de taille $n = 70$ issu d'une population exponentielle, avec un paramètres $\lambda = 1$, nous utilisons un ensemble de niveaux de confiance et résumons les résultats dans le tableau2.5

$1 - \alpha$	$b\_\inf$	$b\_\sup$
0.90	0.830	1.231
0,95	0.801	1.282
0,99	0.749	1.390

TAB. 2.5 – Intervalles de confiance, de differents niveaux, pour la distribution exponentielle

Et Les représentations graphiques sont présentées dans la figure2.5

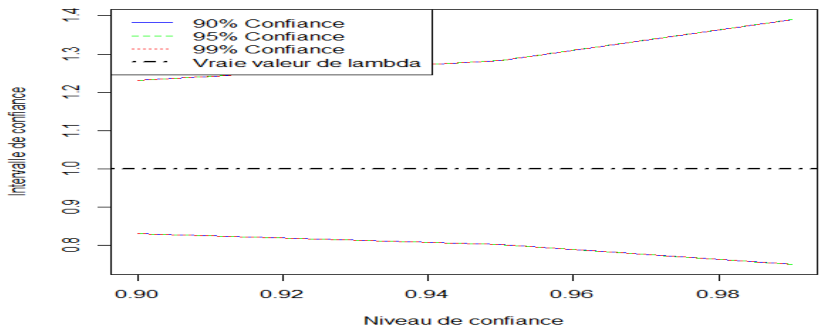

FIG. 2.5 – Intervalle de confiance pour le paramètre lambda d'une distribution exponentiell

Discussion :

Il semble que les intervalles de confiance s'élargissent avec l'augmentation du niveau de confiance. La plupart des intervalles de confiance, à l'exception de celui de 90%, contiennent la vraie valeur du paramètre, ce qui est conforme à l'interprétation des intervalles de confiance où il y a une probabilité de 10% que l'intervalle de 90% n'inclue pas la vraie valeur. Plus la taille de l'échantillon est grande, plus l'intervalle de confiance est petit.

Conclusion

En conclusion, cette recherche explore le domaine de l'estimation paramétrique sous ses deux formes : l'estimation ponctuelle et l'estimation par intervalle de confiance. L'estimation statistique suppose que les données suivent une distribution statistique connue, caractérisée par un ensemble des paramètres, et l'objectif est d'estimer ces paramètres à partir des données observées. L'estimation ponctuelle fournit une seule valeur représentant la meilleure estimation du paramètre recherché, mais il faut se rappeler que l'estimateur est une variable aléatoire avec une certaine variance. En revanche, l'estimation par intervalle de confiance ne fournit pas une seule valeur, mais plutôt une gamme de valeurs possibles pour le paramètre recherché. Cette approche permet de prendre en compte l'incertitude des données et d'obtenir une estimation plus fiable. Les deux méthodes d'estimation sont efficaces dans des cas spécifiques.

Bibliographie

[1] Hines, W. W., Montgomery, D. C., & Borror, D. M. G. C. M. (2003). Probabilités et statistique pour l'ingénieurs. John Wiley & Sons.

[2] Delmas, JF. (2010). Introduction au calcul des probabilités et à la statistique. France.

[3] Lejeune, M. (2010). Statistique la théorie et ses application. Springer-Verlag France. Paris.

[4] Pierre, D. (2017). Cours de statistique inférentilles. Licence 2-S4.SI-Mass.

[5] Ribereau, P. (2016). Cours de statistique inférentilles.

[6] Ruch, JJ. (2012-2013). statistique : Estimation : Préparation à l'Agrégation Bordeaux1.

[7] Saporta, G. (2006). Probabilités, analyse des données et statistique. Technip, Paris.

[8] Tassi, P. (1989). Méthodes statistique. Economica, Paris.

[9] Veysseyre, R. (2006). Aide mémoire, statistique et probabilités pour l'ingénieur. Dunod, Paris.

[10] Ycart, B. (2002). Estimation paramétrique tests statistique. Centre de publication universitaire. Tunis.

Annexe A1 : Logiciel R

2.7 Qu'est-ce-que le langage R ?

• Le langage R est un langage de programmation et un environnement mathématique utilisés pour le traitement de données. Il permet de faire des analyses statistiques aussi bien simples que complexes comme des modèles linéaires ou non-linéaires, des tests d'hypothèse, de la modélisation de séries chronologiques, de la classification, etc. Il dispose également de nombreuses fonctions graphiques très utiles et de qualité professionnelle.

• R a été créé par Ross Ihaka et Robert Gentleman en 1993 à l'Université d'Auckland, Nouvelle Zélande, et est maintenant développé par la R Developement Core Team.

L'origine du nom du langage provient, d'une part, des initiales des prénoms des deux auteurs (Ross Ihaka et Robert Gentleman) et, d'autre part, d'un jeu de mots sur le nom du langage S auquel il est apparenté.

Annexe A2 : Codes R

Les résultats numériques et les représentations graphiques obtenus à partir des différents programmes mis en œuvre sont présentés dans la section suivante.

IC de la moyenne si la variance connue :

```
# Fonction pour calculer l'intervalle de confiance de la moyenne
intervalle_confiance_moyenne <- function(moyenne_echantillon, variance, taille_echantillon,
niveau_confiance = 0.95)
{# Calculer l'écart type
ecart_type <- sqrt(variance)
# Calculer la valeur z pour le niveau de confiance donné
valeur_z <- qnorm((1 + niveau_confiance) / 2)
# Calculer la marge d'erreur
marge_erreur <- valeur_z * ecart_type / sqrt(taille_echantillon)
# Calculer l'intervalle de confiance
borne_inferieure <- moyenne_echantillon - marge_erreur
borne_superieure <- moyenne_echantillon + marge_erreur
# Afficher les résultats
cat("Intervalle de confiance à", niveau_confiance * 100, "% pour la moyenne :\n")
cat("Borne inférieure :", borne_inferieure, "\n")
```

```
cat("Borne supérieure :", borne_superieure, "\n")
# Retourner l'intervalle de confiance
return(c(borne_inferieure, borne_superieure))}
# Exemple d'utilisation
variance <- 2235.398
taille_echantillon <- 20
moyenne_echantillon <- 2276.75
niveaux_confiance <- c(0.90, 0.95, 0.99)
for (niveau_confiance in niveaux_confiance) { intervalle_confiance_moyenne
(moyenne_echantillon, variance, taille_echantillon, niveau_confiance)}
```

IC de la moyenne si la variance inconnue :

```
# Fonction pour calculer l'intervalle de confiance de la moyenne
intervalle_confiance_moyenne <- function(X, S, n, alpha = 0.05)
{ # Calculer la valeur critique t
t_crit <- qt(1 - alpha / 2, df = n - 1)
# Calculer les limites de l'intervalle de confiance
lower <- X - (S / sqrt(n - 1)) * t_crit
upper <- X + (S / sqrt(n - 1)) * t_crit
# Afficher l'intervalle de confiance
cat("Intervalle de confiance à", 100 * (1 - alpha), "% pour la moyenne : [", lower, ",", upper,
"]\n")
# Retourner l'intervalle de confiance
return(c(lower, upper))}
# Exemple d'utilisation
```

```
X <- 360 # Moyenne de l'échantillon
S <- 28 # Écart type de l'échantillon
n <- 15 # Taille de l'échantillon
alphas <- c(0.1, 0.05, 0.01) # Niveaux de confiance (90%, 95%, 99%)
for (alpha in alphas) {
intervalle_confiance_moyenne(X, S, n, alpha)}
```

IC de la variance si la moyenne connue :

```
# Fonction pour calculer l'intervalle de confiance de la variance
intervalle_confiance_variance <- function(n, k, conf_levels = c(0.90, 0.95, 0.99))
{ # Calculer les valeurs critiques du chi-carré
chi2_crits <- matrix(nrow = length(conf_levels), ncol = 2)
for (i in 1 :length(conf_levels))
{conf_level <- conf_levels[i]
alpha <- 1 - conf_level
chi2_crits[i, 1] <- qchisq(alpha / 2, df = n - 1) # Chi-carré inférieur
chi2_crits[i, 2] <- qchisq(1 - alpha / 2, df = n - 1) # Chi-carré supérieur}
# Créer un vecteur pour stocker les intervalles de confiance
intervalles_confiance <- matrix(nrow = length(conf_levels), ncol = 2)
# Calculer l'intervalle de confiance pour chaque niveau de confiance
for (i in 1 :length(conf_levels))
{# Calcul de l'intervalle de confiance pour la variance
intervalles_confiance[i, ] <- c((n * k) / chi2_crits[i, 2], (n * k) / chi2_crits[i, 1])
# Affichage des résultats
cat("Intervalle de confiance à", conf_levels[i] * 100, "% :", intervalles_confiance[i, ], "\n")}
```

```
# Retourner l'intervalle de confiance
return(intervalles_confiance)}
# Exemple d'utilisation
n <- 25 # Taille de l'échantillon
k <- 13 # Somme des observations de l'échantillon
conf_levels <- c(0.90, 0.95, 0.99) # Niveaux de confiance
intervalles_confiance <- intervalle_confiance_variance(n, k, conf_levels)
```

IC de la variance si la moyenne inconnue :

```
# Fonction pour calculer l'intervalle de confiance de la variance
intervalle_confiance_variance <- function(n, S2, conf_levels = c(0.90, 0.95, 0.99))
{ # Créer une matrice pour stocker les intervalles de confiance
intervalles_confiance <- matrix(nrow = length(conf_levels), ncol = 2)
# Calculer les valeurs critiques du chi-carré
chi2_crits <- matrix(nrow = length(conf_levels), ncol = 2)
for (i in 1 :length(conf_levels))
{conf_level <- conf_levels[i]
alpha <- 1 - conf_level
chi2_crits[i, 1] <- qchisq(alpha / 2, df = n - 1) # Chi-carré inférieur
chi2_crits[i, 2] <- qchisq(1 - alpha / 2, df = n - 1) # Chi-carré supérieur}
# Calculer l'intervalle de confiance pour chaque niveau de confiance
for (i in 1 :length(conf_levels))
{# Calcul de l'intervalle de confiance pour la variance
intervalles_confiance[i, ] <- c((n * S2) / chi2_crits[i, 2], (n * S2) / chi2_crits[i, 1])
# Affichage des résultats
```

cat("Intervalle de confiance à", conf_levels[i] * 100, "% :", intervalles_confiance[i,], "\n")}

Retourner l'intervalle de confiance

return(intervalles_confiance)}

Exemple d'utilisation

n <- 26 # Taille de l'échantillon

S2 <- 14 # Variance de l'échantillon

conf_levels <- c(0.90, 0.95, 0.99) # Niveaux de confiance

intervalles_confiance <- intervalle_confiance_variance(n, S2, conf_levels)

IC pour le paramètre lambda d'une distribution exponentiell :

Fonction pour calculer les intervalles de confiance pour

le paramètre lambda d'une distribution exponentielle

intervalle_confiance_lambda <- function(n, lambda,

niveaux_de_confiance = c(0.90, 0.95, 0.99))

{ # Calculer les intervalles de confiance

intervalles_de_confiance <- lapply(niveaux_de_confiance, function(niveau)

{alpha <- 1 - niveau

Obtenir les quantiles de la distribution khi-deux

quantile_inf <- qchisq(alpha/2, 2*n)

quantile_sup <- qchisq(1-alpha/2, 2*n)

Calculer les bornes de l'intervalle de confiance

borne_inf <- (2*n)/quantile_sup

borne_sup <- (2*n)/quantile_inf

c(borne_inf, borne_sup)})

Afficher les intervalles de confiance

```
for (i in 1 :length(niveaux_de_confiance))

{cat("Niveau de confiance :", niveaux_de_confiance[i], "\n")

cat("Intervalle de confiance :", intervalles_de_confiance[[i]], "\n\n")}

# Retourner l'intervalle de confiance

return(intervalles_de_confiance)}

# Exemple d'utilisation

n <- 70 # Taille de l'échantillon

lambda <- 1 # Paramètre de l'exponentielle

niveaux_de_confiance <- c(0.90, 0.95, 0.99)

intervalles_de_confiance <- intervalle_confiance_lambda(n, lambda,

niveaux_de_confiance)
```

Annexe B : Abréviations et Notations

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

v.a	:	variable aléatoire
c à d	:	C'est à dire
$E(x)$	:	Esperence mathámatique
$var(x)$	:	variance mathámatique
$\sim$	:	Suit la loi
iid	:	Indépendantes identiquement distribuée
$\xrightarrow{p}$	:	Convergence en probabilité
$\xrightarrow{p.s}$	:	Convergence presque sûre
$\xrightarrow{L}$	:	Convergence en loi
$B(n,\theta)$	:	La biais de l'estimateur T_n
$\bar{X}$	:	Moyenne empirique
$I(\theta)$	:	Information de fisher
MMV	:	methode de maximum de vraisemblance
$L(x,\theta)$	:	Fonction de vraisemblance
MM	:	Methode du moment
IC	:	Intervalle de confiance

Résumé :

L'estimation paramétrique est un processus statistique visant à estimer les valeurs d'un paramètre ou de plusieurs paramètres dans une distribution de probabilité donnée. Dans ce mémoire, nous avons introduit les principes et les techniques de l'estimation paramétrique, puis nous avons appliqué l'une de ces techniques en utilisant le langage R. L'analyse des résultats obtenus a contribué à renforcer la compréhension des forces et des faiblesses de cette méthode et à évaluer sa pertinence dans différents contextes.

Mots clés : Estimation paramétrique. Maximum de vraisemblance. Méthode du moment. Loi Normale. Intervalle de confiance.

Abstract :

Parametric estimation is a statistical process aimed at estimating the values of a parameter or multiple parameters within a given probability distribution. In this paper, we introduced the principles and techniques of parametric estimation, and then applied one of these techniques using the R language. Analyzing the results obtained contributed to enhancing the understanding of the strengths and weaknesses of this method and evaluating its relevance in different contexts.

Key words : Parametric estimation. Maximum likelihood. Method of moments. Normal distribution. Confidence interval.

ملخص :

التقدير المعلمي هو عملية في الاحصاء تهدف الى تقدير قيم معلمة أو معلمات في توزيع احتمال معين. في هذه المذكرة، قدمنا مقدمة حول مبدا التقدير المعلمي وتقنياته ثم تم تطبيق احدى هذه التقنيات باستخدام لغة **R**

قام تحليل النتائج المحصلة بتعزيز فهم القوة والضعف لهذه الطريقة وتقدير ملاءمتها في سياقات مختلفة .

الكلمات المفتاحية : تقدير المعلمة. الاحتمال الأقصى. طريقة اللحظات. التوزيع العادي. فاصل الثقة.