RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Abdelhamid Ibn Badis de Mostaganem

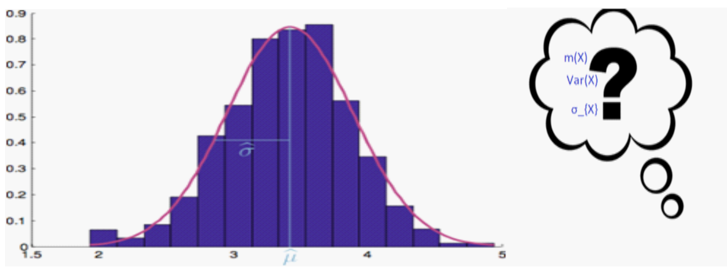
Faculté des Sciences Exactes et de l'Informatique

Département de Mathématique et de L'informatique

UNIVERSITE Abdelhamid Ibn Badis MOSTAGANEM

Polycopié pédagogique

FSEI FACULTÉ DES SCIENCES EXACTES ET DE L'INFORMATIQUE



Statistique paramétrique

Estimation des paramètres

Master 1- Analyse Fonctionnelle - S1

Djemaia BENSIKADDOUR

2021/2022

Table des matières

INTRODUCTION

La statistique paramétrique est une branche de statistique qui permet d'analyser les données (les observations) relatives à un phénomène aléatoire et les modélisées par une distribution (loi de probabilité parente), ayant un nombre fini de paramètres $\left\{\theta \in \mathbb{R}^k,\ k \in \mathbb{N}^*\right\}$.

L'estimation paramétrique est considérée comme l'outil essentiel pour décrire le phénomène étudié, faire des prévisions préalablement sur ses circonstances (ou ses résultats), et prendre des décisions à son sujet au vu des informations (les observations) contenues dans un échantillon de la population. Elle consiste à évaluer les paramètres inconnus θ à partir des observations $(X_1, X_2, ..., X_n)$ avec le plus faible seuil d'erreur. À ce propos, on distingue trois méthodes d'estimation paramétrique :

1/ les estimations ponctuelles,

2/ les estimations par des intervalles de confiance,

3/ les tests d'hypothèses.

L'objectif de ce cours semestriel **"Statistique I"** déstiné aux étudiants de **Master 1 Mathématiques**, option : **Analyse Fonctionnelle** est de présenter les principes des trois méthodes d'estimation paramétrique citées ci-dessus et leurs utilisations. Il nécessite la connaissance des lois de probabilités discrètes et continues (cours de probabilités L2) et quelques notions d'analyse mathématique (dérivation et intégration).

Ce polycopié est composé de quatre chapitres. Dans le premier, on rappelle les différents types d'échantillons puis on présente les concepts principaux de l'estimation paramétrique : les estimateurs et leurs propriétés. Le deuxième chapitre est consacré à exposer le principe de l'estimation ponctuelle des paramètres par la méthode des moments ainsi que la méthode du maximum de la vraisemblance. Dans le troisième chapitre, on aborde l'estimation paramétrique (la moyenne et la vraiance) par les intervalles de confiance en cas des distributions normales. Finalement, le quatrième chapitre consiste à expliquer comment tester les valeurs théoriques des paramètres proposées dans les hypothèses.

Chaque chapitre contient une section d'exercices corrigés et d'autres proposés sous forme d'un devoir, qui doit être fait sérieusement par les étudiants afin d'enrichir leur compréhension et développer leurs compétances.

Chapitre 1

Préliminaires

Souvent les analyses en sciences expérimentales (pharmacologie, biologie, agronomie...) portent sur l'étude des caractères d'une population donnée pour laquelle on n'a pas l'accès à toute l'information.

Exemple 1.0.1 *Pour étudier la crise du lait en Algérie, on choisi la population : les vaches laitières. On peut étudier plusieurs caractères : l'âge des vaches, le volume du lait, la durée de production,...*

On a donc besoin d'un processus d'échantillonnage pour obtenir une approximation des informations qu'on cherche.

1.1 Echantillonnage

Si la population est de taille N, on peut tirer de cette population plusieurs échantillons de taille n, $(n < N)$ et les résultats qu'on produit varieront d'un échantillon à un autre.

Alors, à partir d'un unique échantillon, on pourra déterminer une valeur estimée du paramètre de la population qu'on veut étudier. Ainsi, l'étude se base sur des estimations et non pas des certitudes.

1.1.1 Les différents types d'échantillons

1/ **Echantillon aléatoire (ou probabiliste) :** chaque élément (unité statistique) de la population a une chance d'être séléctionné (ou observé). Ce type d'échantillonnage garantit une représentation sans biais de l'ensemble de la population.

2/ **Echantillon non-aléatoire (ou non-probabiliste) :** est un échantillon sélectionné par une méthode non probabiliste, par exemple : les vaches d'une ferme.

3/ **Echantillon exhaustif :** est un échantillon pour lequel le tirage est réalisé sans remise (les unités statistiques sont choisies sans remise).

4/ **Echantillon non-exhaustif :** est un échantillon où le tirage est réalisé avec remise (les unités statistiques sont choisies avec remise).

5/ **Echantillon représentatif :** est un échantillon qui reproduit les caractéristiques d'une population. En fait, on peut généraliser les résultats obtenus avec cet échantillon à la population.

Remarque 1.1.1 *Si on veut comparer deux populations à l'aide de deux échantillons, alors on distingue deux types différents d'échantillons :*

i/ *Echantillons indépendants :* *si les deux échantillons sont composés de séries d'unités statistiques différentes (les deux échantillons ne sont pas identiques).*

ii/ *Echantillons appariés :* *si les deux échantillons sont composés de la même série d'unités statistiques (les deux échantillons sont identiques).*

Exemple 1.1.1 *On donne un exemple de chaque type d'échantillon :*

1/ *Comparer le nombre de filles en M1MCO avec le nombre de filles en M1AF* $\rightarrow$ *deux échantillons différents : deux échantillons indépendants .*

2/ *Comparer la valeur énergitique moyenne des produits laitiers avant et aprés refroidissement* $\rightarrow$ *le même échantillon : échantillon apparié.*

1.2 Echantillon aléatoire et statistique de l'échantillon

Définition 1.2.1 *Un échantillon alétoire d'effectif* n *est défini par un ensemble de* n *variables alétoires indépendantes* $(X_1, X_{2,...}, X_n)$ *suivant la même loi appelée* ***loi parente.*** *Une réalisation* $(x_1, x_{2,...}, x_n)$ *d'un échantillon aléatoire correspond à* n *répétitions indépendantes d'une même expérience.*

Remarque 1.2.1 *Un échantillon aléatoire d'effectif* n *est appelé* ***n-échantillon.***

Définition 1.2.2 *Toute fonction "*$\mathcal{S}$*" de l'échantillon aléatoire* $(X_1, X_{2,...}, X_n)$ *d'effectif* n *est appelée* ***une statistique de l'échantillon :*** $\mathcal{S}\,(X_1, X_{2,...}, X_n)$.

Définition 1.2.3 ***"Estimateur et estimation"*** *le plus important des objectifs d'une étude statistique, est l'obtention des estimations fiables des caractéristiques d'une population à partir d'un échantillon représentatif prélevé de cette population. Ceci est appelé problème de décision concernant des paramètres tels que :*

$\surd$ *L'espérence mathématique :* $E(X) = m = \bar{X}$.

$\surd$ *La variance* $Var(X)$ *ou l'écart type* σ_X.

$\surd$ *La proportion* p *en cas d'un caractère discret* (*dénombrable).*

Estimer un des paramètres cités ci-dessus, c'est déterminer une valeur approchée de ce paramètre, à partir des résultats obtenus sur un échantillon aléatoire et représentaif extrait de la population.

1.3 Propriétés d'un estimateur

On considère l'échantillon $(X_1, X_2, ..., X_n)$ de n-variables alétoires suivant toutes la même loi de probabilité du paramètre θ inconnu.

1/ *Un estimateur* T_n *du paramètre* θ *est dit "****sans biais****" si* $E(T_n) = \theta$. *De façon générale, un estimateur* T_n *de* $g(\theta)$ *est sans biais si* $E(T_n) = g(\theta)$. *Dans le cas contraire* $(E(T_n) \neq g(\theta))$, T_n *est un estimateur* ***biaisé.***

2/ T_n *est un estimateur asymptotiquement sans biais du paramètre* θ *si*

$$\lim_{n \to +\infty} E(T_n) = \theta.$$

3/ T_n *est un estimateur convergent (ou consistant) de* $g(\theta)$ *s'il converge en probabilité vers* $g(\theta)$, *cela peut être traduit par les deux conditions suivantes :*

$$\lim_{n \to +\infty} E(T_n) = g(\theta) \;\; et \;\; \lim_{n \to +\infty} Var(T_n) = 0.$$

4/ *T_n est un estimateur correct de $g(\theta)$, s'il est sans biais et convergent.*

Définition 1.3.1 *Soit T_n un estimateur du paramètre θ.*

1/ *L'écart entre θ et T_n : $|T_n - \theta|$ est appelé le risque de l'estimateur.*

2/ *Le biais de T_n est $E(T_n) - \theta$.*

3/ *Le risque quadratique ou l'erreur quadratique moyenne "EQM" est :*

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

On conclut donc que

$$EQM(T_n) = variance\ de\ l'estimateur\ +\ carré\ de\ son\ biais.$$

1/ *Si l'estimateur T_n est sans biais, alors*

$$EQM(T_n) = Var(T_n).$$

2/ *L'estimateur T_n converge en moyenne quadratique vers θ si et seulement si son EQM converge vers 0 quand n tend vers $+\infty$,*

$$T_n \overset{MQ}{\to} \theta \Leftrightarrow \lim_{n\to+\infty} \left[E(T_n - \theta)^2\right] = 0.$$

3/ *Si l'estimateur T_n est sans biais, alors*

$$T_n \overset{MQ}{\to} \theta \Leftrightarrow \lim_{n\to+\infty} Var(T_n) = 0.$$

4/ *Le meilleur estimateur de θ est un estimateur sans biais et de variance minimale.*

5/ *Un estimateur sans biais à variance élevée peut fournir des estimations très éloignées de la vraie valeur du paramètre.*

Chapitre 2

Estimation ponctuelle

On vient de voir que l'estimation paramétrique constitue un aspect essentiel de la statistique paramétique, elle permet de déterminer les paramètres de la loi de distribution des données d'une expérience aléatoire, et par conséquent décrire les caractères de la population. On distingue trois formes d'estimations : les estimations ponctuelles, les estimations par intervalle de confiance et les tests d'hypothèses. On commence par aborder le premier moyen de l'estimation : l'estimation ponctuelle.

On suppose que les données $(x_1, x_{2,...}, x_n)$ sont n réalisations de variables aléatoires $(X_1, X_{2,...}, X_n)$ indépendantes suivant la même loi de probabilité.

Les techniques de la statistique descriptive permettent d'adopter une famille de lois de probabilité bien précise pour la loi des X_i, mais les valeurs des paramètres de cette dernière sont inconnues. **Comment estimer ces paramètres ?**

Dans tout ce qui suit, θ est le paramètre inconnu à estimer.

Définition 2.0.2 *L'estimation ponctuelle du paramètre θ est la valeur la plus probable que prendra ce paramètre. Autrement dit, une approximation de θ est la plus proche valeur possible de sa vraie valeur inconnue.*

Notation : *Toute variable aléatoire X_i où* $(1 \leq i \leq n)$ est caractérisée par *sa fonction de répartition $F(x, \theta)$.*

Si pour tout i, X_i suit une loi de probabilité continue, $f\ (x, \theta)$ est la fonction densité (qui est la dérivée de la fonction de répartition). Dans le cas contraire (X_i suit une loi de probabilité discrète), $P\left(X_i = x_i,\ \theta\right)$ sont les probabilités élémentaires.

Définition 2.0.3 *Un estimateur d'un paramètre θ est une statistique $\mathcal{S}_n$ à valeurs dans l'ensemble des valeurs possibles de θ noté Θ. Une estimation de θ est une réalisation $\mathcal{S}_n = \mathcal{S}(x_1, x_{2,}..., x_n)$ de l'estimateur $\mathcal{S}_n = \mathcal{S}(X_1, X_{2,}..., X_n)$.*

Remarque 2.0.1 *Un estimateur est une variable aléatoire, et une estimation est sa valeur déterministe.*

2.1 Estimation ponctuelle par la méthode des moments (EPMM)

L'idée de base de cette méthode est d'estimer les paramètres de la loi de la variable X par les moments centrés ou non centrés de la série statistique.

Par exemple : on peut estimer l'espérance mathématique $E(X)$ d'une variable aléatoire X par une moyenne empirique m (moment non centré d'ordre 1) et sa variance $Var(X)$ par une variance empirique (moment centré d'ordre 2).

Soit X le vecteur formé par un n-échantillon $X_1, X_{2,}..., X_n$. On suppose que les variables X_i, $(1 \leq i \leq n)$ sont à valeurs dans un ensmble $\mathcal{X}$.

Soit $f = (f_1, f_{2,}..., f_n)$ une application de $\mathcal{X}$ dans $\mathbb{R}^k$ telle que l'application φ définie par :

$$\begin{array}{cccc} \varphi: & \Theta & \rightarrow & \mathbb{R}^k \\ & \theta & \mapsto & \varphi(\theta) = E_\theta(f(X_i)) \end{array}$$

est inversible. En fait,

$$\theta = \varphi^{-1}(E_\theta(f(X_i))).$$

Définition 2.1.1 *L'estimateur $\hat{\theta}_n$ du paramètre θ par la méthode des moments est défini comme la solution (quand elle existe et est unique) de l'équation*

$$\varphi(\theta) = \frac{1}{n}\sum_{i=1}^{n} f_i(X_i)$$

où $f_i(x) = x^i$, donc φ correspond au $i^{\text{ème}}$ moment (le moment d'ordre i) de la variable X_i. Si la loi des X_i a deux paramètres θ_1 et θ_2 tels que $\varphi(\theta_1, \theta_2) = (E(X), Var(X))$, alors

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

qui est le moment non centré d'ordre 1 *de* X*, et*

$$\begin{aligned}\hat{\theta}_2 &= S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\bar{X}_n\right)^2 \\ &= \bar{X_n^2} - \left(\bar{X}_n\right)^2 \\ &= Var\left(X_n\right).\end{aligned}$$

On estime donc la variance des X_i *par la variance empirique de l'échantillon notée par* S_n^2 *qui est le moment centré d'ordre* 2*. Par conséquent, les estimateurs* θ_1 *et* θ_2 *par la méthode des moments sont :*

$$\left(\hat{\theta}_{1_n}, \hat{\theta}_{2_n}\right) = \varphi^{-1}\left(\bar{X}_n, S_n^2\right).$$

Exemple 2.1.1 *Soit* $X = (X_1, X_{2,...}, X_n)$ *n-variables alétoires.*

1/ *Loi de Bernoulli :* *On suppose que toutes les variables alétoires* $X_1, X_{2,...}, X_n$ *suivent la loi de Bernoulli qui prend les valeurs* 0 *et* 1 *et de paramètre* p *qui est la proportion de* 1 *dans l'échantillon, la proportion du* 0 *est égale à* $q = 1 - p$*. Puisque l'espérance mathématique* $E(X) = p$*, alors, l'estimateur du paramètre "*p*" par la méthode des moments est* $\hat{p} = \bar{X}_n$*.*

2/ *Loi poissonnienne :* *Toutes les variables aléatoires* $X_1, X_{2,...}, X_n$ *suivent la loi de Poisson* $\mathcal{P}(\lambda)$ *de paramètre* $\lambda > 0$ et *d'espérance mathématique* $E(X) = \lambda$*. Alors, l'estimateur du paramètre* λ *par la méthode des moments est :*

$$\hat{\lambda}_n = \bar{X}_n.$$

3/ *Loi exponnentielle :* *Toutes les variables aléatoires* $X_1, X_{2,...}, X_n$ *suivent la loi exponentielle* $\exp(\lambda)$ *de paramètre* λ *(un réel positif non nul) et d'espérance mathématique* $E(X) = \frac{1}{\lambda}$*. Alors, l'estimateur du paramètre* λ *par la méthode des moments est :*

$$\hat{\lambda}_n = \frac{1}{\bar{X}_n}.$$

4/ *<u>Loi normale :</u> Toutes les variables aléatoires $X_1, X_{2,...}, X_n$ suivent la loi normale $\mathcal{N}(m, \sigma^2)$ de paramètres m et σ où l'espérance mathématique $E(X) = m$ et la variance $Var\,(X) = \sigma^2$. Alors, les estimateurs des paramètres m* et *σ par la méthode des moments sont :*

$$\hat{m}_n = \bar{X}_n \text{ et } \hat{\sigma}_n^2 = S_n^2.$$

5/ *<u>Loi Gamma :</u> Toutes les variables aléatoires $X_1, X_{2,...}, X_n$ suivent la loi Gamma $\mathcal{N}(\alpha, \beta)$ de paramètres α et β (deux réels positifs et β est non-nul) où l'espérance mathématique $E(X) = \frac{\alpha}{\beta}$ et la variance $Var\,(X) = \frac{\alpha}{\beta^2}$. Alors, les estimateurs des paramètres α et β par la méthode des moments sont :*

$$\hat{\beta}_n = \frac{\bar{X}_n}{S_n^2} \text{ et } \hat{\alpha}_n = \frac{(\bar{X}_n)^2}{S_n^2}.$$

Remarque 2.1.1 *La notation $\hat{\theta}_n$ désigne à la fois l'estimateur du paramètre θ qui est une variable alétoire et sa valeur estimée, c'est-à-dire : la variable θ et sa réalisation aléatoire sur l'expérience correspondante au phénomène étudié.*

2.2 Estimation ponctuelle par la méthode du maximum de vraisemblance (EPMMV)

La méthode du maximum de vraisemblance est une méthode efficace due à Fisher, elle permet de deviner un estimateur d'une loi parente.

On commence tout d'abord par définir la fonction de vraisemblance. Soient les obeservations $x_1, x_2, ..., x_n$ des variables aléatoires toutes discrètes ou continues $(X_1, X_2, ..., X_n)$.

Définition 2.2.1 ***"Fonction de vraisemblance"*** *On appelle fonction de vraisemblance pour l'échantillon $(X_1, X_2, ..., X_n)$, la fonction du paramètre θ définie par :*

$$\begin{cases} L(\theta, x_1, ..., x_n) &= P(X = (x_1, ..., x_n), \theta), & \text{si } X_i \text{ est discrète} \\ L(\theta, x_1, ..., x_n) &= f_{X_1, X_2, ..., X_n}(x_1, ..., x_n, \theta), & \text{si } X_i \text{ est continue} \end{cases}.$$

On suppose dans tout ce qui suit que les X_i , $(1 \leq i \leq n)$ sont toutes des variables alétoires indépendantes suivant la même loi de probabilité. On aura donc

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

Remarque 2.2.1 *La fonction de vraisemblance est considérée comme une fonction de θ qui dépend des observations $x_1, ..., x_n$. Ainsi, elle est décrite par rapport au paramètre θ qui peut être unidimensionnel ou multidimentionnel.*

Définition 2.2.2 *Le maximum de vraisemblance du paramètre θ est la valeur qui maximise la fonction de vraisemblance, c'est-à-dire :*

$$\hat{\theta} = \arg\max_{\theta} \left\{ L(\theta, x_1, ..., x_n) \right\}.$$

Remarque 2.2.2 *Pour déterminer le maximum de la fonction de vraisemblance, on maximise le logarithme de $L(\theta, x_1, ..., x_n)$. Par conséquent, $\hat{\theta}$ est la valeur unique si elle existe qui vérifie*

$$\partial_\theta \left(\ln L(\theta, x_1, ..., x_n) \right) = 0.$$

Exemple 2.2.1 *En cas discret, on considère la variable aléatoire X suivant la loi binomiale $\mathcal{B}\,(15, p)$, où p est inconnu. On cherche à estimer le paramètre p sachant que la valeur observée est $x = 5$. La fonction de vraisemblance dans ce cas est :*

$$\begin{aligned} L(p, 5) &= P(x = 15, p) \\ &= C_{15}^5 p^5 (1-p)^{10}. \end{aligned}$$

On suppose que p prend les valeurs 0.1, $0.2, ...,$ 0.9 puis on calcule $L(p, 15)$, les résultats obtenus sont :

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$L(p,15)$	0.01	0.1	0.21	0.19	0.09	0.02	0.003	10^{-4}	2.10^{-7}

Quand $p = 0.1$, la variable aléatoire $X \rightsquigarrow \mathcal{B}\,(15, 0.1)$ et $L(p, 5) = 0.01 = 1\%$. Alors, il y a une chance sur 100 d'observer $x = 5$. La valeur vraisemblance du paramètre p est celle pour la quelle la probabilité d'observer la valeur 5 est maximale. Du tableau, on a : si $p = 0.3 \rightarrow L(p, 5) = 0.21 = 21\%$. Pour déterminer le maximum de $L(p, 5)$, on maximise le logarithme de $L(p, 5)$. Comme

$$\ln (L(p, 5)) = \ln C_{15}^5 + 5 \ln p + 10 \ln (1-p)$$

alors

$$\begin{aligned} \partial_p (L(p, 5)) &= \frac{5}{p} - \frac{10}{1-p} \\ &= \frac{5 - 15p}{p(1-p)}. \end{aligned}$$

Ainsi,

$$\partial_p\left(L(p,5)\right)=0\Rightarrow p=\frac{1}{3}.$$

Par conséquent, la valeur la plus vraisemblance du paramètre p est $\frac{1}{3}$ et la vraisemblance maximale est

$$\begin{aligned} L(\frac{1}{3},5) &= C_{15}^{5}\left(\frac{1}{3}\right)^{5}\left(\frac{2}{3}\right)^{10} \\ &= 21.4\%. \end{aligned}$$

2.3 Exercices

Exercice 1 : Une étude sur la corrélation entre la prise du poids et le développement du diabète a été faite sur un échantillon de 40 personnes malades (diabétiques) ayant tous la même taille. La mesure de leurs poids (en kg) est récapitulé dans le tableau ci-dessous :

65.1	64.5	75.0	81.2	77.8	90.5	66.3	81.4	93.7	72.3
98.7	74.6	83.3	96.4	99.0	98.3	78.9	82.9	96.1	72.2
68.4	74.5	99.8	67.7	70.5	83.6	84.7	76.7	76.2	77.6
63.9	59.1	78.3	81.7	88.1	89.0	62.7	75.7	67.5	90.1

1/ Donner une valeur estimée pour le poids moyen de cet échantillon.

2/ Trouver la valeur estimée de la variance du poids.

Exercice 2 : On considère le modèle statistique de la loi Gamma : $G(\alpha,\beta)$, où α et β sont deux réels positifs et β est non nul. La densité de la variable aléatoire X suivant la loi $G(\alpha,\beta)$ est

$$f_{\alpha,\beta}(x)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\exp(-\beta x),\ x\geq 0$$

avec

$$\Gamma(\alpha)=\int_{0}^{+\infty}\exp(-y)\,y^{\alpha-1}dy.$$

1/ Calculer $E_{\alpha,\beta}(X)$ et $Var_{\alpha,\beta}(X)$.

2/ Par la méthode des moments, donner un estimateur du paramètre bidimentionnel (α,β) du modèle, basé sur l'observation d'un échantillon $X_1,...,X_n$.

Exercice 3 : Soit $X = (X_1, ..., X_n)$, un n-échantillon d'une loi normale de moyenne m et de variance $1/\theta$. On suppose que m est connue et on cherche à estimer θ. Calculer l'estimateur du maximum de vraisemblance (EMV) de θ.

Exercice 4 : On considère $n-$échantillon des variables aléatoires identiquement indépendantes $X_1, ..., X_n$, suivant la loi normale $\mathcal{N}\left(0, \sqrt{\theta}\right)$ où $\theta > 0$ est le paramètre inconnu à estimer.

1/ Déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ.

2/ Calculer le biais $b\,(\theta)$ et le risque quadratique $R(\hat{\theta}_n, \theta)$.

Exercice 5 : Soit $X_1, ..., X_9$ un échantillon aléatoire tiré d'une population de moyenne m et de variance σ^2. On considère $\hat{m}_1$ et $\hat{m}_2$ deux estimateurs de la moyenne m définis par :

$$\hat{m}_1 = \frac{1}{9}\sum_{i=1}^{9} X_i \text{ et } \hat{m}_2 = \frac{2X_1 - X_2 + X_3 + X_6 - X_9}{2}$$

1/ Les estimateurs $\hat{m}_1$ et $\hat{m}_2$ sont-ils sans biais ?

2/ Quel est le meilleur estimateur du paramètre m ?

2.4 Corrigés

Exercice 1 : Le caractère étudié est le poids en kg de 40 personnes malades (diabétiques), il est de type quantitatif continu. Pour cela, on peut réécrire ses valeurs sous forme d'une série classée (9 classes). Les résultats sont récapitulés dans le tableau ci-dessous :

Les classes	n_i	x_i	$n_i x_i$	$n_i x_i^2$
$[55, 65[$	4	60	240	14400
$[65, 75[$	10	70	700	49000
$[75, 85[$	15	80	1200	96000
$[85, 95[$	5	90	450	40500
$[95, 105[$	6	100	600	60000
Total	40	$\backslash$	3190	259900

1/ La valeur estimée du poids moyen de cet échantillon est la moyenne empirique :

$$\begin{aligned} \bar{X} &= \frac{1}{n}\sum_{i=1}^{5} n_i x_i \\ &= \frac{3190}{40} \\ &= 79.75\ Kg \end{aligned}$$

2/ La valeur estimée de la variance du poids de cet échantillon est la variance empirique :

$$\begin{aligned} Var(X) &= \bar{X}^2 - \left(\bar{X}\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{5} n_i x_i^2 - (79.75)^2 \\ &= \frac{259900}{40} - 6360.0625 \\ &= 137.4375\ Kg^2. \end{aligned}$$

<u>Exercice 2 :</u> La densité de la variable aléatoire X suivant la loi $G(\alpha, \beta)$ est :

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)\,,\ \ x \geq 0$$

avec

$$\Gamma(\alpha) = \int_0^{+\infty} \exp(-y)\, y^{\alpha-1} dy.$$

1/ Calculer $E_{\alpha,\beta}(X)$ et $Var_{\alpha,\beta}(X)$:

i/

$$\begin{aligned} E_{\alpha,\beta}(X) &= \int_0^{+\infty} x f_{\alpha,\beta}(x)\, dx \\ &= \int_0^{+\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^\alpha \exp(-\beta x)\, dx. \end{aligned}$$

En intégrant par partie, on obtient

$$E_{\alpha,\beta}(X) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left[\underbrace{\left(-\frac{1}{\beta} x^\alpha \exp(-\beta x)\right)_0^{+\infty}}_{\searrow = 0} + \frac{\alpha}{\beta} \int_0^{+\infty} x^{\alpha-1} \exp(-\beta x)\, dx \right].$$

En faisant le changement de variable $y = \beta x$, on aura

$$\begin{aligned} E_{\alpha,\beta}(X) &= \frac{\alpha}{\beta} \int_0^{+\infty} \left(\frac{y}{\beta}\right)^{\alpha-1} \exp(-y)\, \frac{1}{\beta} dy \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\alpha}{\beta} \frac{1}{\beta^\alpha} \underbrace{\int_0^{+\infty} y^{\alpha-1} \exp(-y)\, dy}_{=\Gamma(\alpha)} \\ &= \frac{\alpha}{\beta}. \end{aligned}$$

ii/

$$\begin{aligned} Var_{\alpha,\beta}(X) &= \int_0^{+\infty} x^2 f_{\alpha,\beta}(x)\,dx - \left(E_{\alpha,\beta}(X)\right)^2 \\ &= \underbrace{\int_0^{+\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha+1}\exp(-\beta x)\,dx}_{(I)} - \frac{\alpha^2}{\beta^2}. \end{aligned}$$

On intègre (I) deux fois par partie, on trouve

$$\begin{aligned} Var_{\alpha,\beta}(X) &= \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} \\ &= \frac{\alpha}{\beta^2}. \end{aligned}$$

2/ $E_{\alpha,\beta}(X)$ et $Var_{\alpha,\beta}(X)$ calculées en question 1/ représentent le moment non centré d'ordre 1 et le moment centré d'ordre 2 respectivement de l'échantillon $X = (X_1, ..., X_n)$. Donc, on peut estimer α et β en fonction de $E(X)$ et $Var(X)$. On a

$$\begin{aligned} \alpha &= \beta E_{\alpha,\beta}(X) \\ &= \beta^2 Var_{\alpha,\beta}(X), \end{aligned}$$

d'où

$$\left\{ \begin{array}{ccc} \beta & = & \frac{E(X)}{Var(X)} \\ & \text{et} & \\ \alpha & = & \frac{(E(X))^2}{Var(X)}. \end{array} \right.$$

<u>**Exercice 3 :**</u> $X = (X_1, ..., X_n)$, un n-échantillon d'une loi normale de moyenne m et de variance $1/\theta$, $\theta \in \mathbb{R}^{*,+}$.

<u>EMV de θ</u> : $\forall i = \overline{1-n}$, $X_i \rightsquigarrow \mathcal{N}\left(m, \frac{1}{\sqrt{\theta}}\right)$, de fonction densité

$$f(X_i/\theta) = \frac{1}{\sqrt{\frac{2\pi}{\theta}}} \exp\left(-\frac{1}{2} \frac{(X_i - m)^2}{\frac{1}{\theta}} \right).$$

La vraisemblance $L(\theta; X_1, ..., X_n)$ est définie par

$$\begin{aligned} L(\theta; X_1, ..., X_n) &= \prod_{i=1}^n f(X_i/\theta) \\ &= \left(\frac{\sqrt{\theta}}{\sqrt{2\pi}} \right)^n \exp\left(-\frac{\theta}{2} \sum_{i=1}^n (X_i - m)^2 \right). \end{aligned}$$

On applique la fonction ln, on obtient

$$\ln L\left(\theta; X_1, ..., X_n\right) = n \ln\left(\frac{\sqrt{\theta}}{\sqrt{2\pi}}\right) - \frac{\theta}{2}\sum_{i=1}^{n}\left(X_i - m\right)^2.$$

On dérive par rapport à θ

$$\partial_\theta \ln L\left(\theta; X_1, ..., X_n\right) = \frac{n}{2\theta} - \frac{1}{2}\sum_{i=1}^{n}\left(X_i - m\right)^2,$$

puis on résout l'équation

$$\begin{aligned}\partial_\theta \ln L\left(\theta; X_1, ..., X_n\right) = 0 \quad &\Leftrightarrow \quad \tfrac{n}{2\theta} - \tfrac{1}{2}\textstyle\sum_{i=1}^{n}\left(X_i - m\right)^2 = 0\\ &\Leftrightarrow \quad \tfrac{n}{\theta} = \textstyle\sum_{i=1}^{n}\left(X_i - m\right)^2.\end{aligned}$$

D'où

$$\begin{aligned}\theta &= \frac{n}{\sum_{i=1}^{n}\left(X_i - m\right)^2}\\ &= \frac{1}{Var(X)}.\end{aligned}$$

Finalement

$$\begin{aligned}EMV\left(\theta\right) &= \frac{1}{Var(X)}\\ &= \hat{\theta}.\end{aligned}$$

<u>**Exercice 4 :**</u> On considère $n-$échantillon des variables aléatoires identiquement indépendantes $X_1, ..., X_n$, suivant la loi normale $\mathcal{N}\left(0, \sqrt{\theta}\right)$ où $\theta > 0$ est le paramètre inconnu à estimer.

1/ L'estimateur du maximum de vraisemblance $\hat{\theta}_n$ du parmètre θ :

$$\begin{aligned}L\left(\theta; X_1, ..., X_n\right) &= \prod_{i=1}^{n} f_\theta\left(X_i\right)\\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}}\exp\left(-\frac{X_i^2}{2\theta}\right)^2\\ &= \left(\frac{1}{2\pi\theta}\right)^{n/2}\exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} X_i^2\right).\end{aligned}$$

En appliquant la fonction log népérien, on aura

$$\ln L\left(\theta; X_1, ..., X_n\right) = -\frac{n}{2}\ln\left(2\pi\theta\right) - \frac{1}{2\theta}\sum_{i=1}^{n} X_i^2.$$

On cherche la valeur qui annule la dérivée de $\ln L\left(\theta; X_1, ..., X_n\right)$ par rapport à θ, on trouve

$$\begin{aligned} \tfrac{\partial}{\partial\theta}\ln L\left(\theta; X_1, ..., X_n\right) = 0 \quad &\Rightarrow \quad -\tfrac{n}{2\theta} + \tfrac{1}{2\theta^2}\sum_{i=1}^{n} X_i^2 = 0 \\ &\Rightarrow \quad \theta = \tfrac{1}{n}\sum_{i=1}^{n} X_i^2. \end{aligned}$$

Alors,

$$EMV = \hat{\theta}_n = Var(X).$$

2/ Calculer le biais $b\left(\theta\right)$ et le risque quadratique $R(\hat{\theta}_n, \theta)$:

i/ <u>Le biais $b\left(\theta\right)$:</u>

$$\begin{aligned} b\left(\theta\right) &= E\left(\hat{\theta}_n\right) - \theta \\ &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - \theta \\ &= \frac{1}{n}E\left(\sum_{i=1}^{n} X_i^2\right) - \theta. \end{aligned}$$

Comme toutes les variables aléatoires $(X_i)_{1\leq i\leq n}$ sont de même loi de probabilité, alors :

$$E\left(X_1^2\right) = E\left(X_2^2\right) = ... = E\left(X_n^2\right).$$

Par conséquent,

$$b\left(\theta\right) = E\left(X_1^2\right) - \theta.$$

En plus, on a

$$\begin{aligned} Var(X) &= E\left(X^2\right) - \left(\underbrace{E(X)}_{=0}\right)^2 \\ &= E\left(X^2\right). \end{aligned}$$

Finalement,

$$\begin{aligned} b\left(\theta\right) &= Var(X) - \theta \\ &= 0, \end{aligned}$$

et EMV de θ est sans biais.

ii/ <u>Le risque quadratique $R(\hat{\theta}_n, \theta)$:</u>

$$\begin{aligned}
R(\hat{\theta}_n, \theta) &= E(\hat{\theta}_n - \theta)^2 \\
&= E\left(\hat{\theta}_n - E\left(\hat{\theta}_n\right) + E\left(\hat{\theta}_n\right) - \theta\right)^2 \\
&= E\left[\left(\hat{\theta}_n - E\left(\hat{\theta}_n\right)\right)^2 + 2\left(\hat{\theta}_n - E\left(\hat{\theta}_n\right)\right)\left(E\left(\hat{\theta}_n\right) - \theta\right) + \left(E\left(\hat{\theta}_n\right) - \theta\right)^2\right] \\
&= E\left(\hat{\theta}_n - E\left(\hat{\theta}_n\right)\right)^2 \\
&= Var(\hat{\theta}_n) \\
&= 0.
\end{aligned}$$

<u>**Exercice 5 :**</u> $X_1, ..., X_9$ un échantillon aléatoire tiré d'une population de moyenne m et de variance σ^2. On considère $\hat{m}_1$ et $\hat{m}_2$ deux estimateurs de la moyenne m définis par :

$$\hat{m}_1 = \frac{1}{9}\sum_{i=1}^{9} X_i \text{ et } \hat{m}_2 = \frac{2X_1 - X_2 + X_3 + X_6 - X_9}{2}$$

1/ Pour calculer l'espérance mathématique des estimateurs $\hat{m}_1$ et $\hat{m}_2$, On utilise la linéarité de l'espérance mathématique. On a

$$\begin{aligned}
E\left(\hat{m}_1\right) &= E\left(\frac{1}{9}\sum_{i=1}^{9} X_i\right) \\
&= \frac{1}{9}\sum_{i=1}^{9} E\left(X_i\right) \\
&= \frac{1}{9}(9m) \\
&= m.
\end{aligned}$$

D'où $\hat{m}_1$ est un estimateur sans biais. De manière similaire, on montre que $\hat{m}_2$ est aussi un estimateur sans biais.

$$\begin{aligned}
E\left(\hat{m}_2\right) &= E\left(\frac{2X_1 - X_2 + X_3 + X_6 - X_9}{2}\right) \\
&= \frac{1}{2}\left[E\left(2X_1\right) - E\left(X_2\right) + E\left(X_3\right) + E\left(X_6\right) - E\left(X_9\right)\right] \\
&= \frac{1}{2}\left(2m - m + m + m - m\right) \\
&= m.
\end{aligned}$$

2/ Pour déterminer le meilleur estimateur, il suffit de comparer leurs variances. On a

$$\begin{aligned} Var(\hat{m}_1) &= Var\left(\frac{1}{9}\sum_{i=1}^{9} X_i\right) \\ &= \frac{1}{9^2}\sum_{i=1}^{9} Var(X_i) \\ &= \frac{1}{9^2}(9\sigma^2) \\ &= \frac{1}{9}\sigma^2. \end{aligned}$$

et

$$\begin{aligned} Var(\hat{m}_2) &= Var\left(\frac{2X_1 - X_2 + X_3 + X_6 - X_9}{2}\right) \\ &= \frac{1}{2^2}\left[Var(2X_1) + Var(X_2) + Var(X_3) + Var(X_6) Var(X_9)\right] \\ &= \frac{1}{2^2}\left(2^2\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2 + \sigma^2\right) \\ &= 2\sigma^2. \end{aligned}$$

Il est clair que

$$Var(\hat{m}_1) < Var(\hat{m}_2).$$

Alors, $\hat{m}_1$ est le meilleur estimateur de m.

2.5 Devoir N 1

Exercice 1 : On suppose que $\hat{\theta}_1$ et $\hat{\theta}_2$ sont deux estimateurs du paramètre θ. Sachant que

$$E\left(\hat{\theta}_1\right) = \theta,\ E\left(\hat{\theta}_2\right) = \frac{\theta}{4},\ Var\left(\hat{\theta}_1\right) = 8 \text{ et } Var\left(\hat{\theta}_2\right) = 2.$$

1/ Comparer les deux estimateurs du paramètre θ.

2/ Lequel préférez-vous ? Pourquoi ?

Exercice 2 : On considère le modèle gaussien (loi normale) $\mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ est sa moyenne et σ est son écart type.

1/ Trouver l'estimateur du maximum de vraisemblance du paramètre bidimensionnel $\theta = (\mu, \sigma^2)$ d'un échantillon $x = (x_1, ..., x_n)$ issu de ce modèle.

2/ L'estimateur du paramètre $\theta = (\mu, \sigma^2)$ est-il biaisé ?

<u>Exercice 3 :</u> Soit X une variable aléatoire uniforme sur l'intervalle $[0,\ 2\theta]$. On considère une suite $X_1, ..., X_n$ de n variables aléatoires indépendantes et de même loi que X. On pose

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{et} \quad Y = max\ (X_1, ..., X_n)\,.$$

1/ Déterminer l'espérance $E\,(X)$ et la variance $Var(X)$ de la variable X.

2/ Montrer que $\bar{X}$ est un estimateur convergent de θ.

3/ Déterminer la densité de la variable Y.

4/ Calculer l'espérence et la variance de Y.

5/ En déduire un autre estimateur Y^* sans biais de θ.

6/ Comparer les deux estimateurs de θ.

Chapitre 3

Estimation par intervalle de confiance

Dans le chapitre précédent, on a illustré comment un paramètre peut être estimé à partir des données d'un échantillon, mais souvent dans l'étude de quelques phénomènes, les estimations ponctuelles n'apportent pas d'information sur la précision des résultats, parcequ'elles ne prennent pas en considération les erreurs dues aux choix des échantillons. Cependant, il est important de comprendre à quel point l'estimation du paramètre est bonne.

Pour évaluer la confiance que l'on peut avoir en une valeur estimée d'un paramètre, il est nécessaire de trouver un intervalle contenant sa vraie valeur, avec un certain niveau de confiance $1-\alpha$ fixé au préalable (α un réel positif non nul est appelé le seuil de signification ou le taux de risque). C'est ce qu'on appelle l'estimation par intervalle de confiance.

En fait, un intervalle de confiance doit avoir de grandes chances de contenir la vraie valeur du paramètre, il est toujours associé à un risque d'erreur α.

3.1 Définitions et propriétés

Soit X une variable aléatoire dont la fonction de répartition $F(x, \theta)$ et la fonction densité $f(x, \theta)$ dépendent d'un paramètre θ. On pose D_θ l'ensemble des valeurs possibles du pramètre θ.

On sintéresse à définir des statistiques particulières "ensemble des variables aléatoires" appelées **des estimateurs** $\hat{\theta}$ qui prennent des valeurs numériques appelées **les estimations du paramètre** θ.

Exemple 3.1.1 *Soit $\bar{X}$ la statistique définie par :*

$$S\left(X_1, X_2, ..., X_n\right) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

qui représente la fonction moyenne arithmétique des n observations d'un échantillon, elle peut être considérée comme un estimateur a priori de l'espérance mathématique $E(X) = m$. En plus, on a

$$E\left(S\right) = E\left(X\right) \text{ et } Var\left(S\right) = \frac{Var(X)}{n}.$$

Si la loi de la variable aléatoire X est connue, on peut en déduire celle de $\bar{X}$.

Définition 3.1.1 ***Le biais*** *de l'estimateur $\hat{\theta}$ du paramètre θ est défini par :*

$$B_0\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta_0$$

où θ_0 est la vraie valeur du paramètre θ et $\hat{\theta}$ est la valeur vraisemblable de θ_0 .

Définition 3.1.2 *Soit $\hat{\theta}$ un estimateur du paramètre θ.* ***Le risque quadratique*** *est défini par :*

$$R\left(\hat{\theta}, \theta\right) = E\left(\left(\hat{\theta} - \theta\right)^2\right).$$

Remarque 3.1.1 *En utilisant le risque quadratique, on peut comparer deux estimateurs du même paramètre θ.*

Définition 3.1.3 *Soient $\hat{\theta}_1$ et $\hat{\theta}_2$ deux estimateurs du même paramètre θ. L'estimateur $\hat{\theta}_1$ est dit meilleur que l'estimateur $\hat{\theta}_2$ si*

$$R\left(\hat{\theta}_1, \theta\right) \leq R\left(\hat{\theta}_2, \theta\right), \text{ pour tout } \theta$$

et

$$\exists\theta,\ R\left(\hat{\theta}_1, \theta\right) < R\left(\hat{\theta}_2, \theta\right).$$

Remarque 3.1.2 *Comme pour tout estimateur $\hat{\theta}$ du paramètre θ, on a*

$$\begin{aligned} E\left(\left(\hat{\theta} - \theta\right)^2\right) &= E\left(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2\right) \\ &= E\left(\hat{\theta}^2\right) - 2E\left(\hat{\theta}\theta\right) + E\left(\theta^2\right) \\ &= E\left(\hat{\theta}^2\right) - 2\theta E\left(\hat{\theta}\right) + \theta^2 - \left(E\left(\hat{\theta}\right)\right)^2 + \left(E\left(\hat{\theta}\right)\right)^2 \\ &= Var\left(\hat{\theta}\right) + \left(E\left(\hat{\theta}\right) - \theta\right)^2 \\ &= Var\left(\hat{\theta}\right) + B_\theta^2\left(\hat{\theta}\right). \end{aligned}$$

Alors, on peut comparer deux estimateurs sans biais de θ en considérant uniquement la variance.

Définition 3.1.4 *Soient $\hat{\theta}_1$ et $\hat{\theta}_2$ deux estimateurs sans biais du même paramètre θ, i.e*

$$B_\theta\left(\hat{\theta}_1\right) = B_\theta\left(\hat{\theta}_2\right) = 0$$

L'estimateur $\hat{\theta}_1$ est plus efficace que l'estimateur $\hat{\theta}_2$ si

$$\forall\theta,\ Var\left(\hat{\theta}_1\right) \leq Var\left(\hat{\theta}_2\right),$$

et

$$\exists\theta,\ Var\left(\hat{\theta}_1\right) < Var\left(\hat{\theta}_2\right).$$

3.2 Estimation de la moynne m et de la variance σ^2

Proposition 3.2.1 *Soit $X = (X_1, X_2, ..., X_n)$ un n-échantillon aléatoire de la loi parente, la loi de X d'espérance*

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

et de variance

$$Var(X) = \sigma_X^2.$$

i/ *L'estimateur sans biais de m* est

$$\begin{aligned} E\left(\bar{X}\right) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) \\ &= \frac{1}{n}.n.m \\ &= m \\ &= \hat{m}_n. \end{aligned}$$

ii/ *L'estimateur de la variance*

$$\begin{aligned} Var\left(\bar{X}\right) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_k\right) \\ &= \frac{1}{n^2} n\sigma^2 \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

On pose

$$\begin{aligned} S_n^2 &= Var\left(\bar{X}\right) \\ &= \frac{\sigma^2}{n}, \end{aligned}$$

on a aussi,

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

Définition 3.2.1 *L'estimateur corrigé*

$$\begin{aligned} S_{n,\ c}^2 &= \frac{nS_n^2}{n-1} \\ &= \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \hat{m}_n\right)^2 \\ &= \frac{n}{n-1}\left[E\left(X^2\right) - \left(E\left(\hat{m}_n\right)\right)^2\right]. \end{aligned}$$

3.2.1 Estimateur de σ^2 en supposant que m est connue

Si la moyenne m est connue, on a

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

Alors,

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - m\right)^2$$

est un estimateur sans biais qui converge vers σ^2.

En sciences expérimentales, il est constaté que si la taille n d'échantillons $(X_1, X_2, ..., X_n)$ est très grande (c'est-à-dire les mesures de l'expérience sont répétées un grand nombre de fois), alors la loi suivie par la moyenne des variables aléatoires $X_1,\ X_2, ...,\ X_n$ tendra vers une loi normale. Pour cela, on va se focaliser sur l'estimation des paramètres de la loi normale.

3.2.2 Estimation des paramètres de la loi normale

On suppose que X suit la loi normale $\mathcal{N}\left(m, \sigma^2\right)$. La variable

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

qui suit la loi normale $\mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$ est un estimateur sans biais de m. En plus, on a

$$E\left(S_n^2\right) = \sigma^2,$$

et

$$Var\left(S_n^2\right) = \frac{2\left(n-1\right)}{n^2}\sigma^4.$$

3.3 Estimation par intervalle de confiance

Définition 3.3.1 *L'estimation par intervalle d'un paramètre inconnu θ consiste à trouver à partir d'un estimateur choisi $\hat{\theta}$, un intervalle $IC = [LI,\ LS]$ dans lequel il est vraisemblable que la valeur correspondante du paramètre s'y trouve. La probabilité liée à cet intervalle de confiance $P(IC)$ est exprimée en pourcentage noté $1-\alpha$ où α est le taux (pourcentage) de risque. La probabilité $1-\alpha$ est appelée le niveau de confiance ou le seuil de confiance*

$$\begin{aligned} P(IC) &= P\left(LI \leq \theta \leq LS\right) \\ &= 1-\alpha \end{aligned}$$

où $1-\alpha$ est la probabilité associée à l'intervalle d'encadrer la vraie valeur du paramètre et LI (resp. LS) est la borne inférieure (resp. supérieure) de l'intervalle de confiance.

Remarque 3.3.1 *Le niveau de confiance est toujours associé à l'intervalle et non au paramètre inconnu θ.*

Pour déterminer l'intervalle de confiance IC, on choisit de prendre des risques symétriques, c'est-à-dire :

$$\begin{cases} P(\theta \leq LI) & = & \frac{\alpha}{2}, \\ P(\theta \geq LS) & = & \frac{\alpha}{2}. \end{cases}$$

3.4 Estimation de la moyenne de la loi normale par intervalle de confiance

Il s'agit de déterminer, à partir de la moyenne $\bar{X}$ de l'échantillon qui suit la loi normale, un intervalle $IC = [LI,\ LS]$ contenant la vraie valeur du paramètre moyenne m.

3.4.1 Estimation de la moyenne de la loi normale où l'écart type σ est connu

On dispose d'un grand échantillon $(n \geq 30)$ ou d'un petit échantillon $(n < 30)$ dont la distribution est normale de moyenne m inconnue et d'écart type $\frac{\sigma}{\sqrt{n}}$ connu.

Soit $(X_1, X_2, ..., X_n)$ un échantillon aléatoire de la loi normale d'espérance (moyenne) m et de variance $Var(X) = \sigma_X^2$, vérifiant

$$\begin{cases} & E(\bar{X}_n) & = m = \hat{m}_n \\ \text{et} & & \\ & Var(\bar{X}_n) & = \frac{\sigma^2}{n} \end{cases}$$

d'où

$$S_n = \frac{\sigma}{\sqrt{n}}.$$

Ainsi, la variable $\bar{X} \rightsquigarrow \mathcal{N}\left(m, \frac{\sigma}{\sqrt{n}}\right)$. On fait le changement de variable

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \tag{3.4.1}$$

tel que la nouvelle variable Z suit la loi normale standard (centrée réduite) : $Z \rightsquigarrow \mathcal{N}(0, 1)$.

<u>Problème :</u> On cherche à déterminer les bornes LI et LS telles que

$$P(LI \leq m \leq LS) = 1 - \alpha.$$

Puisque on choisit des risques symétriques, on commence par déterminer dans la table de la loi normale centrée réduite la valeur $z_{\alpha/2}$ telle que

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha \tag{3.4.2}$$

ce qui implique

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

d'où

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Par conséquent

$$IC = \left[\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{3.4.3}$$

et $1-\alpha$ est le niveau de confiance de contenir la vraie valeur de m.

Définition 3.4.1 *Soit m est la moyenne d'un échantillon aléatoire de taille n d'une population normale de variance σ^2 connue. Un $(1-\alpha)$ IC de m est donné par :*

$$\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

où $z_{\alpha/2}$ est le $\alpha/2$ point de pourcentage supérieur de la distribution normal standard vérifiant (3.4.2).

Exemple 3.4.1 *On considère un échantillon de n variables aléatoires suivant la loi normale $\mathcal{N}(m,1)$. On cherche un IC pour m avec un niveau de confiance égal à 95% sachant que $n=100$ et $\bar{X}=2$. On a*

$$\begin{aligned} P\left(Z \leq z_{\alpha/2}\right) &= 1-\frac{\alpha}{2} \\ &= 1-\frac{0.05}{2} \\ &= 0.975. \end{aligned}$$

Cette probabilité correspond à $z=1.96$, alors

$$\begin{aligned} IC &= \left[2-1.96\frac{1}{\sqrt{100}},\ 2+1.96\frac{1}{\sqrt{100}}\right] \\ &= [1.804,\ 2.196]. \end{aligned}$$

Limites de confiance unilatérales

L'intervalle de confiance dans l'expression (3.4.3) donne à la fois une limite de confiance inférieure et une limite de confiance supérieure pour la moyenne m. Ainsi, il fournit un IC bilatéral. Il est également possible d'obtenir une limite de confiance unilatérale pour m en posant $LI=-\infty$ ou $LS=+\infty$ et en remplaçant $z_{\alpha/2}$ par z_α.

Définition 3.4.2 *Une limite supérieure de confiance au seuil* $(1-\alpha)$ *pour* m *est :*

$$m \leq \bar{X} + z_\alpha \frac{\sigma}{\sqrt{n}} = LS \tag{3.4.4}$$

et une limite inférieure de confiance au seuil $(1-\alpha)$ *pour* m *est :*

$$m \geq \bar{X} - z_\alpha \frac{\sigma}{\sqrt{n}} = LI. \tag{3.4.5}$$

Exemple 3.4.2 *La durée de vie en heures d'une ampoule de* 100 *watts est normalement distribuée avec* $\sigma = 25$ *heures. Un échantillon aléatoire de* 100 *ampoules ont une durée de vie moyenne de* $\bar{X} = 1140$ *heures.*

1/ *Trouver un intervalle de confiance bilatéral à* 95% *de la moyenne de vie de l'ampoule.*

2/ *Déterminer une borne de confiance inférieure à* 95% *de la moyenne de vie de l'ampoule.*

Il est clair que la variable aléatoire T *qui représente la durée de vie de* 100 *ampoules suit la loi normale de moyenne* m *inconnue et d'écart type* σ *connu. Le seuil de confiance* $1-\alpha = 0.95 \rightarrow$ le taux de risque $\alpha = 0.05$.

1/ *De l'expression (3.4.3), l'intervale au seuil* 95% *est :*

$$IC = \left[\bar{X} - z_{\alpha/2}\frac{25}{\sqrt{100}},\ \bar{X} + z_{\alpha/2}\frac{25}{\sqrt{100}}\right]$$

où $z_{\alpha/2}$ *vérifie (3.4.3). C'est-à-dire*

$$\begin{aligned} P\left(Z \leq z_{\alpha/2}\right) &= 1 - \frac{\alpha}{2} \\ &= 0.975. \end{aligned}$$

De la table de la loi normale standard, on trouve $z_{\alpha/2} = 1.96$. *D'où*

$$IC = [1135.1\ h,\ 1144.9\ h].$$

2/ *De l'expression (3.4.5), l'intervale au seuil* 95% *est :*

$$LI = 1140 - z_\alpha \frac{25}{\sqrt{100}}$$

avec

$$\begin{aligned} P\left(Z \geq z_\alpha\right) &= 1 - \alpha \\ &= 0.95. \end{aligned}$$

Comme

$$P(Z \leq -z_\alpha) = P(Z \geq z_\alpha),$$

alors, de la table de la loi normale standard, on trouve $z_\alpha = -1.645$*. D'où*

$$LI = 1140 + 1.645 \times 2.5$$

Par conséquent,

$$IC = [1144.1125\ h,\ +\infty[.$$

3.4.2 Estimation de la moyenne de la loi normale où l'écart type σ est inconnu

Si on connait pas la valeur de σ, on peut déterminer IC, en utilisant un résultat issu du théorème de Fisher.

Tout d'abord, on rappelle les définitions et les propriétés de quelques lois continues dérivées de la loi normale.

Loi Khi-deux

Définition 3.4.3 *Soient* $(X_1, .., X_n)$ *n variables aléatoires suivant toutes la loi normale standard* $\mathcal{N}(0,1)$*. La variable aléatoire*

$$X = \sum_{i=1}^{n} X_i^2$$

est dite une loi Khi-deux à n *degrés de liberté. On note la loi Khi-deux par* $\chi^2(n)$ *ou* χ^2_n*. Sa fonction densité est :*

$$f(x,n) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{n/2-1} \exp(-x/2),\ \forall x \geq 0$$

où Γ *est la fonction Gamma définie par :*

$$\Gamma(t) = \int_0^{+\infty} \exp(-y) y^{t-1} dy.$$

L'espérance mathématique *de la loi Khi-deux* X *est :*

$$E(X) = n.$$

La variance *de la loi Khi-deux X est :*

$$Var(Y) = 2n.$$

L'écart type *de la loi Khi-deux X est :*

$$\delta\left(X\right) = \sqrt{VarX} = \sqrt{2n}.$$

L'approximation normale de la loi Khi-deux

Si la taille de l'échantillon $n > 100$, alors la variable aléatoire X suivant la loi Khi-deux peut être approchée par une loi normale de moyenne $m = n$ et de variance $2n$.

Loi de Student

Définition 3.4.4 *Soient deux variables alétoires Z et U telles que Z suit la loi normale standard $Z \rightsquigarrow \mathcal{N}(0,1)$ et U suit la loi Khi-deux de n degrés de liberté $U \rightsquigarrow \chi_n^2$. La variable alétoire*

$$T = \frac{Z}{\sqrt{\frac{U}{n}}}$$

suit la loi de Student à n degrés de liberté notée par St_n ou T_n. Sa fonction densité est :

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+t^2\right)^{-\frac{n+1}{2}},\ \textit{pour tout } n > 0.$$

Γ ici est la fonction Gamma d'Euler définie par :

$$\Gamma(z) = 2\int_0^{+\infty} y^{2z-1}\exp(-y^2)dy.$$

*Son **espérance mathématique** est nulle pour tout $n > 1$ et sa variance est égale à $\frac{n}{n-2}$ pour tout $n > 2$.*

Théorème 3.4.1 ***"Fisher"*** *Soit $X = (X_1, X_2, ..., X_n)$ une suite de variables aléatoires indépendantes de même loi $\mathcal{N}\left(m, \sigma\right)$. Si*

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

et

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

alors

i/ $\bar{X}_n$ *et* $S_n(X)$ *sont indépendantes.*

ii/ $(n-1)\, S_n^2(X)/\sigma^2 \rightsquigarrow \chi^2_{n-1}$.

iii/ $\sqrt{n}\left(\bar{X}_n - m\right)/S_n(X) \rightsquigarrow St_{n-1}$.

Remarque 3.4.1 *On veut estimer la variance de la loi* $X = (X_1, X_2, ..., X_n)$ *par la variance empirique, on a*

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

et

$$\begin{aligned} E\left(S_n^2\right) &= E\left(X^2\right) - E\left(\left(\bar{X}_n\right)^2\right) \\ &= Var(X) - (E(X))^2 - Var\left(\bar{X}_n\right) - \left(E\left(\bar{X}_n\right)\right)^2 \\ &= Var(X) - (E(X))^2 - \frac{Var(X)}{n} - (E(X))^2 \\ &= \frac{n-1}{n} Var(X) \\ &\neq Var(X). \end{aligned}$$

Alors, la variance empirique S_n^2 *n'est pas un estimateur sans biais de* $Var(X)$.
Par contre, on voit que

$$\begin{aligned} E\left(\frac{n}{n-1} S_n^2\right) &= \frac{n}{n-1} E\left(S_n^2\right) \\ &= Var(X) \end{aligned}$$

donc, on prend

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

Cette quantité $S_n'^2$ *est appelée variance estimée de l'échantillon, elle représente un estimateur sans biais de* $Var(X)$.

Maintenant, on peut déterminer l'intervalle de confiance de la moyenne m si σ est inconnu. En remplaçant σ dans l'expression (3.4.1) par un estimateur, on obtient

$$\begin{aligned} T &= \sqrt{n}\frac{\bar{X}-m}{S'_n} \\ &= \sqrt{n-1}\frac{\bar{X}-m}{S_n}. \end{aligned}$$

La nouvelle variable T suit la loi de Student de $(n-1)$ degrés de liberté.
Alors, on peut écrire

$$\begin{aligned} P\left(\left|\bar{X}-m\right| \leq \varepsilon\right) &= P\left(|T| \leq \sqrt{n}\frac{\varepsilon}{S'_n}\right) \\ &= 1-P\left(|T| > \sqrt{n}\frac{\varepsilon}{S'_n}\right) \end{aligned}$$

où la variable T suit la loi $St\,(n-1)$. Or, la table de la loi de Student donne la valeur $t_{n-1,\ \alpha/2}$ telle que $P\left(|T| > t_{n-1,\ \alpha/2}\right)$.
Par conséquent,

$$t_{n-1,\ \alpha/2} = \sqrt{n}\frac{\varepsilon}{S'_n}$$

ce qui donne

$$\varepsilon = \frac{S'_n}{\sqrt{n}} t_{n-1,\ \alpha/2}.$$

Enfin, l'intervalle de confiance devient

$$\begin{aligned} IC &= \left[\bar{X}-\frac{S'_n}{\sqrt{n}}t_{n-1,\ \alpha/2},\ \bar{X}+\frac{S'_n}{\sqrt{n}}t_{n-1,\ \alpha/2}\right] \\ &= \left[\bar{X}-\frac{S_n}{\sqrt{n-1}}t_{n-1,\ \alpha/2},\ \bar{X}+\frac{S_n}{\sqrt{n-1}}t_{n-1,\ \alpha/2}\right]. \end{aligned}$$

Remarque 3.4.2 *Le quantile $t_{n-1,\ \alpha/2}$ d'ordre $1-\frac{\alpha}{2}$ est la valeur lu dans la colonne $P=\alpha$ et la ligne $n-1$, (voir la table de Student en annexe 2).*

Définition 3.4.5 *Si $\bar{X}$ et S_n sont la moyenne et l'écart type d'un échantillon de distribution normale où l'écart type σ est inconnu, l'intervale de $(1-\alpha)$ niveau de confiance pour la moyenne m est donné par :*

$$IC = \left[\bar{X}-\frac{S_n}{\sqrt{n-1}}t_{n-1,\ \alpha/2},\ \bar{X}+\frac{S_n}{\sqrt{n-1}}t_{n-1,\ \alpha/2}\right]. \tag{3.4.6}$$

Exemple 3.4.3 *On prend $n = 20$, $\bar{X} = 64.2$, $S'_n = 5.15$ et $\alpha = 5\%$. De la table de la loi de Sudent, on trouve $t_{19,\ 0.05} = 2.093$, d'où*

$$IC = [61.8,\ 66.7]\,.$$

Remarque 3.4.3 *Un intervalle de confiance trop large n'a aucun intérêt, en pratique puisque la vraie valeur du paramètre ne peut pas être déterminée de façon précise. Alors, il faut parfois accepter un risque d'erreur relativement fort pour obtenir un intervalle de confiance utilisable.*

Définition 3.4.6 *Soient $\bar{X}$ et S'_n sont la moyenne et l'écart type d'un échantillon de distribution normale où l'écart type σ est inconnu. Une limite supérieure de niveau de confiance $(1-\alpha)$ pour la moyenne m est :*

$$m \leq \bar{X} + \frac{S_n}{\sqrt{n-1}} t_{n-1,\ \alpha} = LS,$$

et une limite inférieure de niveau de confiance $(1-\alpha)$ pour la moyenne m est :

$$m \geq \bar{X} + \frac{S_n}{\sqrt{n-1}} t_{n-1,\ \alpha} = LI.$$

Remarque 3.4.4 *Le quantile $t_{n-1,\ \alpha}$ d'ordre $1-\alpha$ est la valeur lu dans la colonne $P = 2\alpha$ et la ligne $n-1$, (voir la table de Student en annexe 2).*

3.5 Intervalle de confiance pour la variance

On utilise le théorème de Fisher (3.4.1) pour chercher une fonction des observations $X_1, X_2, ..., X_n$ et de σ^2 dont la loi de probabilité ne dépend ni de m ni de σ^2. Soient $a,\ b \in \mathbb{R}^{*,+}$ avec $0 < a < b$. D'une part, on a

$$\begin{aligned} P\left(a < \frac{nS_n^2}{\sigma^2} < b\right) &= P\left(\frac{nS_n^2}{b} < \sigma^2 < \frac{nS_n^2}{a}\right) \\ &= 1-\alpha \end{aligned}$$

d'autre part

$$P\left(a < \frac{nS_n^2}{\sigma^2} < b\right) = F_{\chi^2_{n-1}}(b) - F_{\chi^2_{n-1}}(a)$$

Afin d'équilibrer les risques, on prend a et b tels que

$$F_{\chi^2_{n-1}}(b) = 1 - \frac{\alpha}{2} \text{ et } F_{\chi^2_{n-1}}(a) = \frac{\alpha}{2}$$

La table de χ^2 donne la valeur $z_{n,\alpha}$ qui vérifie

$$\begin{aligned} P\left(Z > z_{n,\alpha}\right) &= 1 - F_{\chi^2_n}\left(z_{n,\alpha}\right) \\ &= \alpha. \end{aligned}$$

Alors, pour $b = z_{n-1,\alpha/2}$ et $a = z_{n-1,1-\alpha/2}$ on a

$$P\left(\frac{nS_n^2}{b} < \sigma^2 < \frac{nS_n^2}{a}\right) = 1 - \alpha.$$

Enfin, l'intervalle de confiance s'écrit

$$\begin{aligned} IC &= \left[\frac{nS_n^2}{z_{n-1,\alpha/2}}, \frac{nS_n^2}{z_{n-1,1-\alpha/2}}\right] \\ &= \left[\frac{(n-1)\,S_n'^2}{z_{n-1,\alpha/2}}, \frac{(n-1)\,S_n'^2}{z_{n-1,1-\alpha/2}}\right]. \end{aligned}$$

Exemple 3.5.1 *On prend* $n = 20,\ S_n'^2 = 26.5,\ \alpha = 0.05$. *De la table de la loi* χ^2*, on trouve*

$$z_{19,\ 0.025} = 32.85 \ \textit{et}\ z_{19,\ 0.975} = 8.91.$$

D'où

$$IC = [15.3,\ 56.6]\,.$$

Cet intervalle est trop large par rapport à celui de la moyenne $m \rightarrow$ *l'estimation de variance est moins précise que celle de la moyenne.*

Théorème 3.5.1 ***"Théorème central limite"*** *Soit* X *une variable aléatoire suivant une loi de probabilité de moyenne* m *et d'écart type* σ. *Lorsque la taille de l'échantillon* n *est trés grande, la loi de* X *peut être approximée (approchée) par une loi normale de moyenne* m *et d'écart type* $\frac{\sigma}{\sqrt{n}}$. *Par conséquent, la variable*

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}}$$

a approximativement une distribution normale standard.

3.6 Intervalle de confiance pour la proportion d'une population

Souvent, il est nécessaire de construire des intervalles de confiance sur une proportion de population. Par exemple, on suppose qu'un échantillon aléatoire de n éléments a été tiré d'une population de taille très grande (éventuellement infini). Si $X \leq n$ observations dans cet échantillon appartiennent à une classe d'intérêt, alors $\hat{p} = \frac{X}{n}$ est un estimateur ponctuel de la proportion p de la population qui appartient à cette classe. Ici, n et p sont les paramètres d'une distribution binomiale.

On rappelle que si $np \geq 5$ et $n(1-p) \geq 5$, la distribution binomiale peut être approximée par une distribution normale de moyenne np et de variance $np(1-p)$. Ainsi, la nouvelle variable

$$\begin{aligned} Z &= \frac{X - np}{\sqrt{np(1-p)}} \\ &= \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \end{aligned}$$

suit la loi normale standard.

Définition 3.6.1 *L'intervalle de confiance pour la proportion p d'une population donné par :*

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{3.6.1}$$

vérifie

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

où $1-\alpha$ est le niveau de confiance et $z_{\alpha/2}$ est le point $\alpha/2$ de pourcentage supérieur de la distribution normale standard, c'est-à-dire

$$P\left(Z \leq z_{\alpha/2}\right) = 1 - \frac{\alpha}{2}.$$

Exemple 3.6.1 *Dans un échantillon aléatoire de 100 pneus, 10 ont une finition de surface qui est plus rugueuse que les spécifications ne le permettent. Par conséquent, une estimation* *ponctuelle de la proportion des pneus dans la population qui dépasse la spécification de rugosité est*

$$\begin{aligned}\hat{p} &= \frac{X}{n} \\ &= \frac{10}{100} \\ &= 0.1.\end{aligned}$$

Un intervalle de confiance bilatéral à 95% *niveau de confiance pour* p *est calculé à partir de l'expression (3.6.1) comme suit :*

$$\hat{p} - z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

ce qui donne

$$0.1 - 1.96\sqrt{\frac{0.1 \times 0.9}{100}} \leq p \leq 0.1 + 1.96\sqrt{\frac{0.1 \times 0.9}{100}}.$$

D'où,

$$IC = [0.0412,\ 0.1588].$$

Définition 3.6.2 *Une limite supérieure de niveau de confiance* $(1-\alpha)$ *pour la proportion* p *est :*

$$p \leq \bar{X} + z_{\alpha}\frac{\sigma}{\sqrt{n}} = LS$$

et une limite inférieure de niveau de confiance $(1-\alpha)$ *pour la proportion* p *est :*

$$p \geq \bar{X} - z_{\alpha}\frac{\sigma}{\sqrt{n}} = LI.$$

3.7 Exercices

Exercice 1 : Un échantillon de 5000 personnes sur une population, on sait que le taux moyen de personnes à soigner pour un problème de cholestérol élevé est de 5.7%. Donner un intervalle dans lequel on soit «sûr» à 95%, de trouver le nombre exact de personnes à soigner sur les 5000.

Exercice 2 : Dans un centre avicole, des études ont montré que la masse d'un œuf choisi au hasard peut être considérée comme la réalisation d'une variable aléatoire normale X, de moyenne m et de variance σ^2. On admet que les masses des œufs sont indépendantes les unes des autres. On pèse œufs d'un échantillon de taille $n = 36$, les mesures obtenues sont en gramme :

50.34	52.62	53.79	54.99	55.82	57.67
51.41	53.13	53.89	55.04	55.91	57.99
51.51	53.28	54.63	55.12	55.95	58.10
52.07	53.30	54.76	55.24	57.05	59.30
52.22	53.32	54.78	55.28	57.18	60.58
52.38	53.39	54.93	55.56	57.31	63.15

1/ Calculer la moyenne et l'´ecart-type de cette série statistique. Tracer son histogramme.

2/ Donner une estimation des paramètres m et σ.

3/ Donner un intervalle de confiance au niveau 95%, puis 98%, de la masse moyenne m d'un œuf.

<u>**Exercice 3 :**</u> Dans une usine, on fait emballer le café dans des boites dont le poids net de la boite pleine X est une variable aléatoire qui suit une loi normale $\mathcal{N}(m, \sigma)$. Un contrôle de fabrication sur un échantillon de taille $n = 5000$ montre que 1650 ont un poids inférieur à 240 g et 1760 ont un poids supérieur à 244.1 g.

1/ Trouver les paramètres m et σ.

2/ Déterminer un intervalle de confiance du poids X des boites avec une probabilité égale à 90%.

3/ Déterminer un intervalle de confiance du poids moyen $\bar{X}$ des boites avec une probabilité égale à 90%.

<u>**Exercice 4 :**</u> Un ingénieur contrôle qualité a mesuré les diamètres de 25 flexibles PVC. La moyenne de l'échantillon était $\overline{X} = 4.05\ cm$, et l'écart type de l'échantillon était $S'_n = 0.08$ cm.

1/ Déterminer l'intervalle de confiance à niveau 95% pour le diamètre moyen des flexibles.

2/ Trouver une borne de confiance inférieure à niveau 95% pour le diamètre moyen des flexibles.

3/ Interpréter l'intervalle obtenu en 2/.

Exercice 5 : On veut étudier la proportion p des étudiants de la faculté des sciences exactes et de l'informatique "FSEI" absents chaque mois. On suppose que N le nombre des étudiants absents dans un échantillon de taille $n = 100$.

1/ Quelle est la loi de la variable N ? Par quelle loi peut-on l'approcher et pourquoi ? En déduire une approximation de la loi de $X = N/n$.

2/ On observe une proportion x des étudiants qui s'absentent chaque mois. Donner la forme d'un intervalle de confiance pour la proportion p, de niveau de confiance $1 - \alpha$.

3/ Application numérique : $x = 0.1,\ 1 - \alpha = 90\%,\ 95\%,\ 98\%$.

3.8 Corrigés

Exercice 1 : On a $n = 5000,\ p = 0.057$ et $1 - \alpha = 0.95$.

La variable aléatoire X : le nombre de personnes à soigner pour un problème de cholestérol élevé suit la loi binomiale $\mathcal{B}(n,\ p)$.

Comme $n >> 30,\ np \geq 5$ et $np(1-p) \geq 5$, alors la loi de X peut être approximée par la loi normale $\mathcal{N}(m, \sigma)$ avec

$$m = np \text{ et } \sigma = \sqrt{np(1-p)},$$

d'où

$$\begin{cases} m = 285, \\ \sigma = 16.4. \end{cases}$$

En faisant le changement de variable

$$Z = \frac{X - m}{\sigma},$$

on passe à la loi normale standard $\mathcal{N}(0, 1)$.

Le seuil de signification $\alpha = 1 - 0.95 = 0.05$, pour déterminer la valeur de $z_{\alpha/2}$ qui vérifie

$$\begin{aligned} P(Z \leq z_{\alpha/2}) &= 1 - \tfrac{\alpha}{2} \\ &= 0.975 \end{aligned}$$

on utilise la table de la loi $\mathcal{N}(0, 1)$. On trouve $z_{\alpha/2} = 1.96$.

L'intervalle de confiance du nombre de personnes à soigner vérifie

$$\begin{aligned}
P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) &= P\left(-z_{\alpha/2} \leq \frac{X-m}{\sigma} \leq z_{\alpha/2}\right) \\
&= P\left(m - z_{\alpha/2}\sigma \leq X \leq m - z_{\alpha/2}\sigma\right) \\
&= P\left(285 - 1.96 \times 16.4 \leq X \leq 285 + 1.96 \times 16.4\right).
\end{aligned}$$

Alors,

$$\begin{aligned}
IC &= [252.856,\ 317.144] \\
&\simeq [253,\ 317].
\end{aligned}$$

<u>Exercice 2 :</u> La variable aléatoire X représente les masses des œufs, elle suit la loi normale de moyenne m et de variance σ^2 : $X \rightsquigarrow \mathcal{N}\ (m, \sigma^2)$.

1/ On commence tout d'abord par réécrire les données du tableau sous forme de 5 classes toutes de même amplitude

$$\begin{aligned}
a &= \frac{e}{5} \\
&= \frac{x_{\max} - x_{\min}}{5} \\
&= \frac{63.15 - 50.34}{5} \\
&= 2.562.
\end{aligned}$$

On prend $a = 2.6$.

Les classes	n_i	c_i	$n_i c_i$	$n_i c_i^2$
$[50.34,\ 52.94[$	7	51.14	357.98	18307.1
$[52.94,\ 55.54[$	16	53.73	859.68	46190.6
$[55.54,\ 58.14[$	10	56.34	563.40	31742
$[58.14,\ 60.74[$	2	58.94	117.88	6947.8
$[60.74,\ 63.34[$	1	61.54	61.54	3787.2
Total	$n = 36$	$\backslash$	1960.48	106974.7

La moyenne arithmétique de la série statistique classée représentée dans le tableau ci-dessus est :

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{5} n_i c_i$$

où les $(c_i)_{1\leq i\leq 5}$ sont les centres des classes.

$$\begin{aligned}\bar{X} &= \frac{1960.48\ g}{36}\\ &= 54.45\ g.\end{aligned}$$

L'écart type est

$$\sigma_X = \sqrt{Var\,(X)}$$

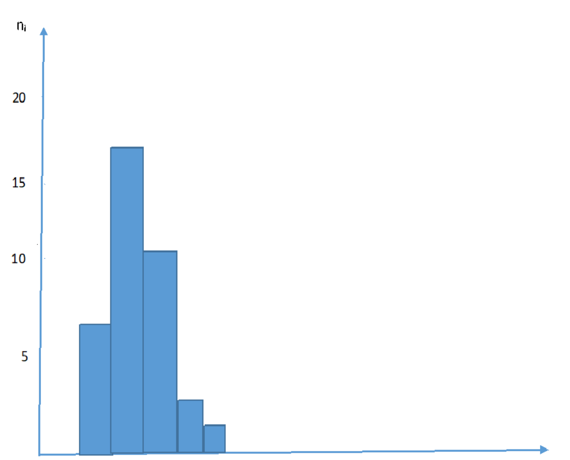
où

$$\begin{aligned}Var\,(X) &= \overline{X^2} - \left(\overline{X}\right)^2\\ &= \frac{1}{n}\sum_{i=1}^{5} n_i c_i^2 - \left(\overline{X}\right)^2\\ &= \frac{106974.7\ g^2}{36} - (54.45\ g)^2\\ &= 6.7\ g^2.\end{aligned}$$

Alors,

$$\begin{aligned}\sigma_X &= \sqrt{6.7\ g^2}\\ &= 2.58\ g.\end{aligned}$$

<u>**L'histogramme des effectifs :**</u> on représente sur l'axe des abscisses les classes de la série statistique et les effectifs partiels sur l'axe des ordonnées.



L'histogramme des effectifs

2/ La moyenne empirique $\bar{X}$ est une estimation de la moyenne m et la racine carrée de la variance empirique$\sqrt{Var\left(X\right)}$ est une estimation de l'écart type σ_X.

3/ L'intervalle de confiance de la moyenne m est :

$$IC = \left[\bar{X} - z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right]$$

i/ Si $1-\alpha = 95\%,\ z_{\alpha/2} = z_{0.025} = 1.96$ et

$$\begin{aligned} IC &= \left[54.45 - 1.96 \times \frac{2.58}{6},\ 54.45 + 1.96 \times \frac{2.58}{6}\right] \\ &= \left[53.6\ g,\ 55.3\ g\right]. \end{aligned}$$

ii/ Si $1-\alpha = 98\%,\ z_{\alpha/2} = z_{0.01} = 2.326$ et

$$\begin{aligned} IC &= \left[54.45 - 2.326 \times \frac{2.58}{6},\ 54.45 + 2.326 \times \frac{2.58}{6}\right] \\ &= \left[53.44\ g,\ 55.45\ g\right]. \end{aligned}$$

<u>Exercice 3 :</u> La variable aléatoire $X \rightsquigarrow \mathcal{N}(m, \sigma)$ et $n = 5000$.

1/ La moyenne m et l'écart type σ sont les solutions du système suivant :

$$\begin{cases} P\left(X < 240\right) &= \frac{1650}{5000} \\ P\left(X > 244.1\right) &= \frac{1760}{5000} \end{cases}$$

ce qui implique

$$\begin{cases} P\left(X < 240\right) &= 0.33 \\ 1 - P\left(X < 244.1\right) &= 0.352 \end{cases}$$

d'où

$$\begin{cases} P\left(X < 240\right) &= 0.33 \\ P\left(X < 244.1\right) &= 0.648. \end{cases}$$

Après le changement de variable $Z = \frac{X-m}{\sigma} \rightsquigarrow \mathcal{N}(0,1)$, on trouve de la table de la loi normale standard

$$\begin{cases} \frac{240-m}{\sigma} &= -0.44 \\ \frac{244.1-m}{\sigma} &= 0.38 \end{cases}$$

ceci implique

$$\begin{cases} 0.44\sigma + 240 &= m, \\ 0.82\sigma &= 4.1. \end{cases}$$

Finalement, on touve

$$\begin{cases} m &= 242.2\ g, \\ \sigma &= 5\ g. \end{cases}$$

2/ L'intervalle de confiance du poids X des boites avec un niveau de confiance 90% est :

$$IC = \left[m - z_{\alpha/2}\sigma,\ m + z_{\alpha/2}\sigma\right]$$

avec

$$P\left(-z_{\alpha/2} \leq \frac{X-m}{\sigma} \leq z_{\alpha/2}\right) = 0.9$$

d'où

$$\begin{aligned} P\left(Z \leq z_{\alpha/2}\right) &= 1 - \frac{\alpha}{2} \\ &= 0.95. \end{aligned}$$

De la table de la loi normale standard $\mathcal{N}(0,1)$, on trouve $z_{\alpha/2} = 1.645$. Donc,

$$\begin{aligned} IC &= [242.2 - 1.645 \times 5,\ 242.2 + 1.645 \times 5] \\ &= [233.975\ g,\ 250.425\ g]. \end{aligned}$$

3/ De façon similaire, on détermine l'intervalle de confiance de la moyenne $\bar{X}$ avec un niveau de confiance 90%. Comme $\bar{X} \rightsquigarrow \mathcal{N}\left(m, \frac{\sigma}{\sqrt{n}}\right)$, alors la nouvelle variable

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0,1).$$

Ainsi, l'intervalle de confiance de la moyenne $\bar{X}$ est

$$\begin{aligned} IC &= \left[m - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ m + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \\ &= \left[242.2 - 1.645 \times \frac{5}{\sqrt{5000}},\ 242.2 + 1.645 \times \frac{5}{\sqrt{5000}}\right] \\ &= [242.08\ g,\ 242.31\ g]. \end{aligned}$$

<u>**Exercice 4 :**</u> On sait que $\overline{X} = 4.05\ cm$, $S_n = 0.08\ cm$ et $n = 25$.

1/ L'intervalle de confiance inférieure à niveau 95% pour le diamètre moyen des flexibles :

$$IC = \left[\bar{X} - \frac{S_n}{\sqrt{25}} t_{n-1,\ \alpha/2},\ \bar{X} + \frac{S_n}{\sqrt{25}} t_{n-1,\ \alpha/2}\right].$$

Comme $\alpha = 0.05$, alors de la table de Student, on trouve

$$t_{n-1,\ \alpha/2} = 2.064.$$

Ainsi,

$$\begin{aligned} IC &= \left[4.05 - \frac{0.08}{\sqrt{25}} \times 2.064,\ 4.05 + \frac{0.08}{\sqrt{25}} \times 2.064\right] \\ &= [4.017\ cm,\ 4.083\ cm]. \end{aligned}$$

2/ La borne de confiance inférieure à niveau 95% pour le diamètre moyen des flexibles est :

$$m \geq \bar{X} + \frac{S'_n}{\sqrt{n}} t_{n-1,\ \alpha}.$$

De la table de Student, on trouve

$$t_{n-1,\ \alpha} = 1.711.$$

Donc,

$$\begin{aligned} m &\geq 4.05 - \frac{0.08}{\sqrt{25}} \times 1.711 \\ &\geq 4.023\ cm. \end{aligned}$$

3/ A partir de ce dernier intervalle, on peut supposer que le diamètre moyen des flexibles est supérieur à 4.023 cm.

Exercice 5 : Dans un échantillon de taille $n = 100$, on a N étudiant absent chaque mois.

1/ Déterminer la loi de N et sa loi approximée :

i/ La variable N suit la loi binomiale $\mathcal{B}(n, p)$ où $n = 100$ et p est la probabilité des étudiants absents.

ii/ Puisque $n = 100 >> 30$, on peut approximer la loi de variable N par la loi normale de moyenne $m = np$, et d'écart type $\sigma = \sqrt{np(1-p)}$: $N \rightsquigarrow \mathcal{N}\left(np, \sqrt{np(1-p)}\right)$.

iii/ De ii/, on peut en déduire une approximation de la loi de la variable $X = \frac{N}{n}$

$$N \rightsquigarrow \mathcal{N}\left(np, \sqrt{np(1-p)}\right) \Rightarrow \frac{N}{n} \rightsquigarrow \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

2/ La variable $X = \frac{N}{n}$ représente la proportion des étudiants absents chaque mois. D'après ce qui précède,

$$\frac{N}{n} \rightsquigarrow \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

alors, en faisant le changement de variable

$$Z = \frac{x-p}{\sqrt{\frac{p(1-p)}{n}}} \rightsquigarrow \mathcal{N}(0,1) .$$

On trouve dans la table de la loi normale standard la valeur $z_{\alpha/2}$ telle que

$$P\left(Z \leq z_{\alpha/2}\right) = 1 - \frac{\alpha}{2}$$

avec α est le seuil de risque (le niveau de confiance $(1-\alpha)$ est donné). On obtient donc l'intervalle de confiance en vérifiant :

$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$$

ce qui implique

$$P\left(-z_{\alpha/2} \leq \frac{x-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

D'où

$$IC = \left[x - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}},\ x - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right]. \tag{3.8.1}$$

3/ **<u>Application numérique :</u>**

i/ $x = 0.1$, $\alpha = 0.1$ et $z_{\alpha/2} = 1.645$. En remplaçant dans (3.8.1), on trouve

$$\begin{aligned} IC &= \left[0.1 - 1.645\sqrt{\frac{0.1\times 0.9}{100}},\ 0.1 + 1.645\sqrt{\frac{0.1\times 0.9}{100}}\right] \\ &= [0.050,\ 0.149] . \end{aligned}$$

ii/ $x = 0.1$, $\alpha = 0.05$ et $z_{\alpha/2} = 1.96$. En remplaçant dans (3.8.1), on trouve

$$\begin{aligned} IC &= \left[0.1 - 1.96\sqrt{\frac{0.1\times 0.9}{100}},\ 0.1 + 1.96\sqrt{\frac{0.1\times 0.9}{100}}\right] \\ &= [0.0412,\ 0.1588] . \end{aligned}$$

iii/ $x = 0.1$, $\alpha = 0.02$ et $z_{\alpha/2} = 2.326$. En remplaçant dans (3.8.1), on trouve

$$\begin{aligned} IC &= \left[0.1 - 2.326\sqrt{\frac{0.1\times 0.9}{100}},\ 0.1 + 2.326\sqrt{\frac{0.1\times 0.9}{100}}\right] \\ &= [0.03022,\ 0.16978] . \end{aligned}$$

3.9 Devoir N 2

Exercice 1 : Démontrer le théorème de Fisher (3.4.1).

Exercice 2 : La teneur en sucre du jus d'orange industriel est normalement distribué. Un échantillon aléatoire de $n = 10$ canettes donne un écart type d'échantillon de $S = 4,8\ mg$.

1/ Déterminer un intervalle à niveau 95% de confiance pour σ.

2/ Trouver une borne de confiance inférieure à niveau 95% pour σ.

Exercice 3 : Dans une usine, on produit en série des planches en bois. La surface d'une planche est une variable aléatoire X suivant la loi normale d'écart type $\sigma = 4$.

Après mise en place d'un nouveau processus de fabrication, afin de déterminer une estimation de la moyenne m de la variable X, on a prélevé un échantillon de 25 planches, et on a trouvé une moyenne empirique $\bar{X} = 45.25\ dm^2$.

1/ Trouver un intervalle de confiance pour la moyenne m au niveau 98% en supposant que l'écart type n'a pas changé au moment de la mise en place du nouveau processus.

2/ Pour un même niveau de confiance, on souhaite réduire la largeur de l'intervalle trouvé dans la question précédente en choisissant un échantillon de taille supérieure. En souhaitant obtenir une largeur d'intervalle de $0.75\ dm^2$. Quelle doit être la taille n du nouvel échantillon.

3/ Suite à la mise en place du nouveau processus, on considère maintenant que l'écart type de la variable X ne peut pas être supposé invariant. On relève l'écart type empirique de l'échantillon et on trouve $S_n = 4\ dm^2$. Trouver le nouvel intervalle de confiance de m avec le même niveau de confiance et le comparer à celui trouvé dans la première question.

4/ En considérant le même processus de la question 3, construire un intervalle de confiance de la variance.

5/ Commenter les résultats.

Chapitre 4

Tests d'hypothèses

4.1 Généralités

D'après ce qu'on a vu dans les deux chapitres précédents : estimation des paramètres (ponctuelle et par intervalle de confiance), il est bien clair que la statistique (variable d'échantillonnage) serve des estimateurs aux paramètres de la population qui n'ont prend pas les mêmes valeurs théoriques proposées dans l'hypothèse.

Le plus souvent, pour prendre une décision sur la base d'un ensemble d'observations, on fera un choix entre deux hypothèses contradictoires : H_0 (l'hypotèse nulle) et H_1 (l'hypothèse alternative) qui dépendent des paramètres théorique de la population.

Pour décider si l'hypothèse formulée H_0 est supportée ou non par les observations, il faut une méthode qui permettra de vérifier l'information hypothétique. C'est ce qu'on appelle les tests d'hypothèses.

4.1.1 Hypothèse statistique

En statistique mathématique, l'information hypothétique concerne la population à laquelle on s'intéresse. C'est une information qui peut être :

1/ Une distribution, par exemple : répartition des poids, de la température, de l'âge des joueurs,...

2/ Une valeur ponctuelle, par exemple : une proportion, un quantil, une moyenne, une variance,...

3/ Un intervalle de valeurs contenant la valeur d'une statistique.

4.1.2 Hypothèses de test

Tout d'abord, on formule les hypothèses suivantes :

1/ **L'hypothèse nulle** : Cette hypothèse notée H_0, représente la valeur a priori d'un paramètre de la population qu'on veut vérifier.

2/ **L'hypothèse alternative** : Elle représente toute hypothèse qui diffère de l'hypothèse nulle, on la note H_1.

Remarque 4.1.1 *H_1 est l'autre hypothèse que l'on souhaite démontrer en rejetant H_0.*

4.1.3 Test d'hypothèse

La construction d'un test d'hypothèse consiste à déterminer entre quelles valeurs peut varier la variable aléatoire, en supposant l'hypothèse vraie, sur la seule considération du hasard de l'échantillonnage.
Un test d'hypothèse peut aussi être utilisé pour vérifier le succés (ou l'échec) d'une action pour modifier la valeur d'une statistique de population.
On peut résumer les étapes du test d'hypothèse comme suit :

1/ Choisir un critère fonction des observations, pour lequel on rejette l'hypothèse nulle H_0 pour démontrer que la contre hypothèse H_1 "l'hypothèse alternative" est vraie.

2/ Calculer la zone de rejet de H_0 correspondant aux résultats qui sont tellement éloignés du résultat moyen attendu qu'ils n'ont presque aucune chance de se produire si H_0 est vraie.

3/ Comparer la distribution du paramètre sous H_0 avec les résultats du paramètre empirique réellement obtenus à partir d'un échantillon (les observations).

4/ Conclure que :

- **i/** H_0 est peu crédible $\Rightarrow$ on rejette H_0 si le paramètre empirique appartient à la zone de rejet, c'est-à-dire, il est en dehors de la région de confiance (la région crédible).

- **ii/** H_0 est crédible $\Rightarrow$ on accepte H_0 si le paramètre empirique appartient à la zone crédible.

Remarque 4.1.2 *La désision de rejeter H_0 signifie que l'hypothèse nulle est fausse. La décision de ne pas rejeter H_0 n'est pas équivalente à H_0 est vraie. On ne pourra prendre cette décision qu'en considérant un certain risque qu'elle soit erronée. Ce risque est donné par le seuil de signification du test.*

Risque d'erreur de prmière espèce

Définition 4.1.1 *Le risque noté α, s'appelle le seuil de signification du test, il est défini par :*

$$\alpha = P\left(rejeter\ H_0/\ H_0\ vraie\right).$$

Ainsi, α la probabilité de la région de rejet de l'hypothèse nulle est le risque d'erreur de première espèce.

Exemple 4.1.1 *Si $\alpha = 0.05$, on admet d'avance que la variable d'échantillonnage peut prendre dans 5% des cas, une valeur se situant dans la région de rejet de H_0, bien que H_0 soit vraie.*

La région complémentaire de la région de rejet ou ce qu'on appelle la région d'acceptation de H_0 a une probabilité égale à $1 - \alpha$.

Remarque 4.1.3 *La statistique qui convient pour le test est donc une variable aléatoire dont la valeur observée sera utinlisée pour décider* ***"du rejet"*** *ou* ***"du non rejet"*** *de H_0.*

La distribution d'échantillonnage de cette statistique sera déterminée en supposant que H_0 est vraie.

4.1.4 Formulation d'un test

La formulation d'un test passe par les étapes suivantes :

1/ On suppose que la valeur d'un paramètre θ (proportion, moment, quantile,...) d'une population est égale à θ_0.

2/ On distingue deux tests possibles du paramètre θ :

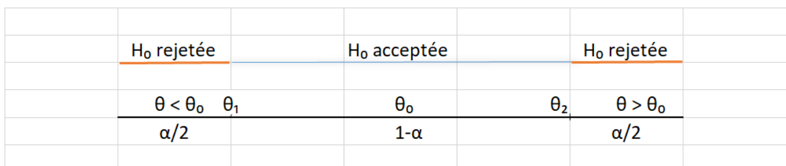
a/ Unilatéral :

$$\begin{cases} H_0 : & \theta = \theta_0, \\ H_1 : & \theta < \theta_0 \quad ou \quad \theta > \theta_0. \end{cases}$$

b/ Bilatéral :

$$\begin{cases} H_0 : & \theta = \theta_0, \\ H_1 : & \theta \neq \theta_0. \end{cases}$$

3/ Shématiser les régions de rejet et de non rejet de H_0



Exemple 4.1.2 *On s'intéresse à la proportion p de garçons dans un échantillon de 120 étudiants préparant le concours de doctorat en mathématiques.*

$$\begin{cases} H_0 : & \text{le pourcentage de garçons est } 20\% \Rightarrow p = 0.2. \\ \quad \text{contre} & \\ H_1 : & p \neq 0.2. \end{cases}$$

X : la variable aléatoire qui représente le nombre de garçons dans un échantillon de 120 étudiants (condidats) $\rightsquigarrow B(n,p)$ où $n = 120$ et $p = 0.2$. Comme $n = 120 >> 30$, on peut approximer la loi binomiale par la loi normale de paramètre

$$\begin{aligned} m &= \overline{X} \\ &= np \\ &= 120 \times 0.2 \\ &= 24, \end{aligned}$$

et

$$\begin{aligned} \sigma &= \sqrt{npq} \\ &= \sqrt{120 \times 0.2 \times 0.8} \\ &= 4.38. \end{aligned}$$

On peut donc estimer sous H_0 les paramètres m et σ. En considérant un intervalle de niveau de confiance 95% de la moyenne

$$IC_m = \left[\overline{X} - z_{1-\alpha/2}\hat{\sigma},\ \overline{X} + z_{1-\alpha/2}\hat{\sigma}\right]$$

où $\alpha = 5\%$. H_0 *est vraie si la moyenne* $m \in IC_m$, *et elle est rejetée si elle n'appartient pas à* IC_m. *L'intervalle de confiance de la proportion* p *est :*

$$IC_p = \left[p - z_{1-\alpha/2}\hat{\sigma},\ p + z_{1-\alpha/2}\hat{\sigma}\right].$$

Rejeter H_0 *veut dire qu'elle est fausse et* H_1 *est vraie avec un risque d'erreur* $\alpha = 5\%$.

Risque d'erreur de deuxième espèce

Définition 4.1.2 *Le risque d'erreur* β *est la probabilité que la valeur expérimentale ou calculée de la statistique n'appartienne pas à la région (zone) de rejet si* H_1 *est vraie. Dans ce cas,* H_0 *est acceptée et* H_1 *est fausse. Le risque* β *de deuxième espèce est celui d'accepter* H_0*, alors qu'elle est fausse ou rejeter* H_1 *alors qu'elle est vraie*

$$\begin{aligned}\beta &= P\left(accepter\ H_0\ /\ H_0\ est\ fausse\right)\\ &= P\left(accepter\ H_0\ /\ H_1\ est\ \ vraie\right)\\ &= P\left(rejeter\ H_1\ /\ H_1\ est\ vraie\right).\end{aligned}$$

Remarque 4.1.4 *Pour quantifier le risque* β, *il faut connaitre la loi de probabilité de la statistique* S *sous l'hypothèse* H_1.

Exemple 4.1.3 *On cherche à tester l'hypothèse qu'une pièce de monnaie n'est pas truquée. La probabilité* p *d'obtenir face est* 0.6 *pour une pièce truquée. On adopte la règle de décision suivante :*

$$H_0 :\quad \text{la pièce n'est pas truquée}\begin{cases}\text{acceptée} & \text{si} \qquad x \in [40,\ 60]\\ \text{rejetée} & \text{sinon}\end{cases}$$

où x *est le nombre de faces obtenus en lançant la pièce* 100 *fois. Quel est le risque d'erreur de deuxième espèce* β *?*

La puissance d'un test $(1-\beta)$

Définition 4.1.3 *L'aptitude d'un test à rejeter* H_0*, alors qu'elle est fausse constitue la puissance d'un test notée par* $1-\beta$,

$$\begin{aligned}1-\beta &= P(rejeter\ H_0\ /\ H_0\ est\ fausse)\\ &= P(accepter\ H_1\ /\ H_1\ est\ est\ vraie).\end{aligned}$$

Remarque 4.1.5 *La puissance d'un test augmente avec la taille de l'échantillon n étudié à valeur de α constant, et elle diminue lorsque α diminue.*

Exemple 4.1.4 *On reprend le même exemple (4.1.3), calculer la puissance du test lorsque la probabilité d'obtenir face est* 0.2 *puis* 0.5 *pour une pièce truquée. Que constatez vous ?*

Les règles de décision

Le tableau ci-dessous résume les deux types de risque ainsi que les régles de décision correspondantes

		H_0 est vraie	H_0 est fausse
Décision	H_0 est rejetée	Mauvaise décision : erreur de type 1. α: probabilité de rejeter H_0, alors qu'elle est vraie.	Bonne décision : puissance du test $1-\beta$: probabilité de rejeter H_0, alors qu'elle est fausse.
	H_0 est acceptée	Bonne décision : $1-\alpha$: probabilité de ne pas rejeter H_0, alors qu'elle est vraie.	Mauvaise décision : erreur de type 2. β: probabilité de ne pas rejeter H_0, alors qu'elle est fausse.

Choix d'un test statistique

Le choix d'un test statistique dépend de :

1/ La nature des données.

2/ Le type d'hypothèse que l'on désire contrôler.

3/ La nature des populations (égalité des variances ou non).

Tests de conformité

Les tests de conformité sont déstinés à vérifier si un échantillon peut être considéré comme extrait d'une population donnée, c'est-à dire l'échantillon est représentatif de la population vis-à-vis d'un paramètre : (moyenne, variance...). Ceci implique que la loi du paramètre est connue au niveau de la population.

4.2 Comparaison d'une moyenne observée et d'une moyenne théorique

On s'interesse à comparer la moyenne observée (moyenne calculée à partir d'un échantillon) à une moyenne théorique d'une population dans deux cas différents :

i/ La variance (variance théorique) de la population est connue.

ii/ La variance de la population est inconnue (elle est remplacée par son estimateur).

4.2.1 Principe du test

On considère une variable aléatoire X observée sur une population suivant une loi normale et un échantillon extrait de cette population.

Population connue		*Population inconnue*
$X \rightsquigarrow \mathcal{N}(m_0, \sigma_0)$		$X \rightsquigarrow \mathcal{N}(m, \sigma)$
$H_0 : m = m_0$		$H_1 : m \neq m_0$
	Echantillon $n,\ \overline{X},\ S_n^2$	

où $\overline{X}$ est un estimateur de m.

1/ Si $\overline{X}$ appartient à une population de référence connue d'espérance m_0, alors H_0 est vraie.

2/ Si $\overline{X}$ appartient à une population de référence inconnue d'espérance m, alors H_1 est vraie.

4.2.2 Variance de la population connue

Statistique du test

$\overline{X}$ la distribution d'échantillonnage de la moyenne dans la population inconnue suit la loi normale de moyenne m et d'écart type $\sqrt{\frac{\sigma^2}{n}}$

$$\overline{X} \rightsquigarrow \mathcal{N}\left(m, \sqrt{\frac{\sigma^2}{n}}\right).$$

La statistique étudiée est

$$\mathcal{S} = \overline{X} - m_0 \rightsquigarrow \mathcal{N}\left(0, \sqrt{\frac{\sigma^2}{n}}\right).$$

Sous l'hypothèse H_0, on a

$$E(\mathcal{S}) = 0 \text{ et } Var(\mathcal{S}) = \frac{\sigma^2}{n}$$

avec σ^2 est connu.

La nouvelle variable obtenue par passage à la loi normale centrée réduite est :

$$Z = \frac{\mathcal{S} - E(\mathcal{S})}{\sqrt{Var(\mathcal{S})}} = \frac{\overline{X} - m_0}{\sqrt{\frac{\sigma^2}{n}}} \rightsquigarrow \mathcal{N}(0,1).$$

L'hypothèse testée est

$$\boxed{H_0 : m = m_0 \quad \text{contre} \quad H_1 : m \neq m_0}$$

avec σ est connu.

La valeur z calculée

$$z = \frac{\left|\overline{X} - m_0\right|}{\sqrt{\frac{\sigma^2}{n}}}$$

est notée par ε_{obs}, elle est comparée avec la valeur ε_{seuil} lue sur latable de la loi $\mathcal{N}(0,1)$ pour un risque d'erreur α fixé (règle de décision 1).

Règle de décision

1/ Si $\varepsilon_{obs} > \varepsilon_{seuil}$, l'hypothèse H_0 est rejetée au risque d'erreur α.

2/ Si $\varepsilon_{obs} \leq \varepsilon_{seuil}$, l'hypothèse H_0 est acceptée au risque d'erreur α.

Exemple 4.2.1 *Soit X la variable alétoire représentant le taux de glycémie. X suit la loi normale de moyenne $m_0 = 1\ g/l$ et d'écart type $0.1\ g/l$. En étudiant un échantillon de 9 patients, on trouve le taux moyen de glycémie $\bar{X} = 1.12\ g/l$. Cet échantillon est-il représentatif ?*

$$\boxed{H_0 : m = m_0 = 1\ g/l \quad \textit{contre} \quad H_1 : m \neq m_0}$$

Un échantillon n, $\overline{X}$, S_n^2 où $\overline{X} \rightsquigarrow \left(m, \sqrt{\frac{\sigma^2}{n}}\right)$, la statistique

$$\mathcal{S} = \overline{X} - m_0 \rightsquigarrow \mathcal{N}\left(0, \sqrt{\frac{\sigma^2}{n}}\right).$$

Sous l'hypothèse H_0, on a

$$E(\mathcal{S}) = 0 \ \textit{et}\ Var(\mathcal{S}) = \frac{\sigma^2}{n}$$

avec σ^2 *est connu. La nouvelle variable obtenue par passage à la loi normale centrée réduite est :*

$$Z = \frac{\mathcal{S}-E(\mathcal{S})}{\sqrt{Var(\mathcal{S})}} = \frac{\overline{X} - m_0}{\sqrt{\frac{\sigma^2}{n}}} \rightsquigarrow \mathcal{N}(0,1).$$

En choisissant $\alpha = 0.05$, *on trouve la valeur* z *de la variable* Z

$$\begin{aligned} z &= \frac{1.12 - 1}{\sqrt{\frac{0.01}{9}}} \\ &= 3.6 > \varepsilon_{seuil} \end{aligned}$$

où $\varepsilon_{seuil} = 1.96$. *Alors, l'échantillon n'est pas représentatif de la population d'espérance* $m_0 = 1\ g/l$.

Test de Khi-deux Soient $X_1,\ X_2, ..., X_n$ un échantillon de n observations d'une distribution normale où la variance σ^2 est égale à une valeur spécifiée σ_0^2, (ou l'écart type est égal à σ_0). Pour tester l'hypothèse

$$H_0 : \sigma^2 = \sigma_0^2 \text{ contre } H_1 : \sigma^2 \neq \sigma_0^2$$

on va utiliser la variable

$$Z = \frac{(n-1)S^2}{\sigma_0^2}. \tag{4.2.1}$$

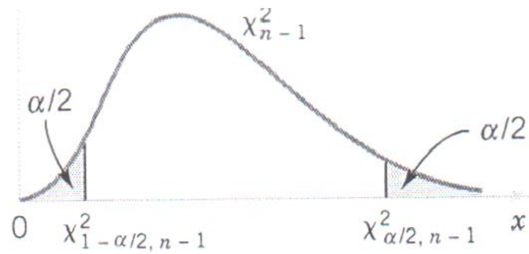
Si l'hypothèse nulle est vraie, la nouvelle variable Z donnée dans (4.2.1) suit la loi de Khi-deux χ^2_{n-1} à $(n-1)$ degrés de liberté.

Règle de décision

Pour un risque d'erreur α fixé, on calcule la valeur

$$\begin{aligned} \varepsilon_{obs} &= \frac{(n-1)S^2}{\sigma_0^2} \\ &= \chi^2_{n-1}. \end{aligned}$$

L'hypothèse nulle H_0 est rejetée si $\varepsilon_{obs} > \chi^2_{n-1,\ \alpha/2}$ ou $\varepsilon_{obs} < \chi^2_{n-1,\ 1-\alpha/2}$, voir la figure ci-dessous :



Test Khi-deux bilatral

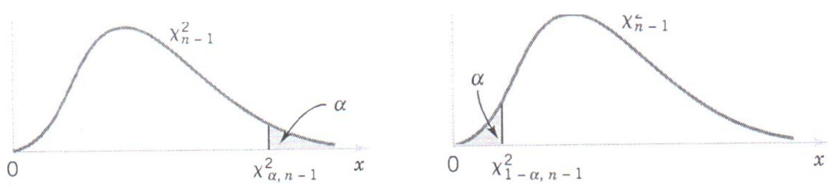
Le même test est utilisé pour une hypothèse alternative unilatérale

$$H_0 : \sigma^2 = \sigma_0^2 \ \text{ contre } H_1 : \sigma^2 > \sigma_0^2$$

où on rejette H_0 si $\varepsilon_{obs} > \chi^2_{n-1,\ \alpha}$. Ainsi, Pour l'autre hypothèse unilatérale

$$H_0 : \sigma^2 = \sigma_0^2 \ \text{ contre } H_1 : \sigma^2 < \sigma_0^2$$

on rejette H_0 si $\varepsilon_{obs} < \chi^2_{n-1,\ 1-\alpha}$, voir les figures suivantes :



4.2.3 Variance de la poulation inconnue

Comme la variance de la population est inconnue, alors elle est estimée par

$$\hat{\sigma}^2 = \frac{n}{n-1} S^2$$

où S est la variance empirique.

Statistique du test

La statistique étudiée est

$$\mathcal{S} = \overline{X} - m_0 \rightsquigarrow \mathcal{N}\left(0, \sqrt{\frac{\hat{\sigma}^2}{n}}\right).$$

avec

$$E(\mathcal{S}) = 0 \text{ et } Var(\mathcal{S}) = \frac{\hat{\sigma}^2}{n}.$$

La nouvelle variable obtenue par passage à la loi normale centrée réduite est :

$$\begin{aligned} Z &= \frac{\mathcal{S} - E(\mathcal{S})}{\sqrt{Var(\mathcal{S})}} \\ &= \frac{\overline{X} - m_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}} \rightsquigarrow St_{n-1} \end{aligned}$$

St_{n-1} est la loi de Student à $(n-1)$ degrés de liberté.

L'hypothèse testée est la suivante :

$$\boxed{H_0 : m = m_0 \quad \text{contre} \quad H_1 : m \neq m_0}$$

La valeur t de la variable Z suivant la loi St_{n-1} est calculée

$$t_{cal} = \frac{\left|\overline{X} - m_0\right|}{\sqrt{\frac{\hat{\sigma}^2}{n}}} = \frac{\left|\overline{X} - m_0\right|}{\sqrt{\frac{S^2}{n-1}}} = t_{obs}.$$

La valeur t_{obs} est comparée avec la valeur t_{seuil} lue dans la table de Student pour un risque d'erreur α fixé et $(n-1)$ degrés de liberté.

Règle de décision

1/ Si $t_{obs} > t_{seuil}$, l'hypothèse H_0 est rejetée au risque d'erreur α.

2/ Si $t_{obs} \leq t_{seuil}$, l'hypothèse H_0 est acceptée au risque d'erreur α.

Exemple 4.2.2 *Pour étudier un lot de fabrication de comprimés, on prélève au hasard* 10 *comprimés parmi les* 30000 *produits et on les pèse. On observe les valeurs de masse en grammes.*

0.80	0.81	0.84	0.83	0.85	0.86	0.85	0.83	0.84	0.80

Le poids moyen observée est-il compatible avec la valeur 0.83 *moyenne de la production au seuil* 98%?

Remarque 4.2.1 *Si $n < 30$, la variable aléatoire X étudiée doit impérativement suivre une loi normale $\mathcal{N}(0, \sigma)$. Pour $n \geq 30$, la variable de Student converge vers une loi normale centrée réduite.*

4.3 Comparaison d'une fréquence observée et une fréquence théorique

4.3.1 Principe du test

On considère la variable aléatoire X suivant la loi de bernoulli

X	0	1
P_X	$1-p$	p

observée sur une population. Un échantillon de taille n extrait de cette population suit la loi binomiale $\mathcal{B}(n,\ p)$.

Population connue		Population inconnue
$X \rightsquigarrow \mathcal{B}(n,\ p)$		$X \rightsquigarrow \mathcal{B}(n,\ p_0)$
$H_0 : p = p_0$		$H_1 : p \neq p_0$
	$\underset{n,\ \ f_i = \frac{n_i}{n}}{Echantillon}\ \ alétoire$	

Si la fréquence observée $f_i = \frac{n_i}{n}$ est un estimateur de la proportion p appartient à une population de référence connue de fréquence connue p_0, alors l'hypothse nulle H_0 est vraie. Sinon, f_i est en dehors de la population de fréquence connue $p_0 \Rightarrow f_i \neq p_0$, l'hypothèse alternative H_1 est vraie.

4.3.2 La distribution d'échantillonnage

On suppose que les variances dans la population de référence et dans population d'où on extrait l'échantillon sont égales. La fréquence f_i suit la loi normale $\mathcal{N}\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$, et la statistique étudiée est : $\mathcal{S} = f_i - p_0$ suit la loi normale $\mathcal{N}\left(0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$ de moyenne $E(\mathcal{S}) = 0$ et de variance

$$Var(\mathcal{S}) = \sqrt{\frac{p_0(1-p_0)}{n}}.$$

On pose maintenant

$$\begin{aligned} Z &= \frac{\mathcal{S} - E(\mathcal{S})}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &= \frac{f_i - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \\ &\rightsquigarrow \mathcal{N}(0,1) \end{aligned}$$

4.3.3 Règle de décision

La valeur calculée

$$\varepsilon_{obs} = \frac{f_i - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

est comparée avec la valeur ε_{seuil} trouvée sur la table de la loi normale standard pour un risque d'erreur α fixé au préalable.

Décision :

1/ Si $\varepsilon_{obs} > \varepsilon_{seuil}$, l'échantillon est en dehors de la population de référence $\rightarrow$ l'échantillon n'est pas représentatif et H_o est rejetée.

2/ Si $\varepsilon_{obs} \leq \varepsilon_{seuil}$, l'échantillon est représentatif et H_o est acceptée.

4.4 Exercices

Exercice 1 : Dans chacune des situations suivantes, indiquer s'il s'agit d'un problème de test d'hypothèse correctement énoncé. Justifier la réponse.

1/ $H_0 : m = 5,\ H_1 : m \neq 5.$

2/ $H_0 : m = 5,\ H_1 : m < 5.$

3/ $H_0 : \sigma < 2,\ H_1 : m = 2.$

4/ $H_0 : \bar{X} = 44,\ H_1 : \bar{X} \neq 44.$

5/ $H_0 : p = 0.1,\ H_1 : p = 0.19.$

6/ $H_0 : S_n = 153,\ H_1 : S_n > 153.$

Exercice 2 : Un fabricant a proposé des tubes à essai en déclarant que leur durée de vie est supérieure à 2000 h de chauffage. A l'aide d'un échantillon aléatoire de taille $n = 100$, on touve la durée de vie moyenne des tubes égale à 1975 h et l'écart-type est 130 h.

En considèrant un seuil de risque $\alpha = 0.05$, peut-on affirmer que la déclaration du fabriquant n'est pas vraie ?

Exercice 3 : Pour confirmer que la résistance moyenne des projecteurs LED produits dans une usine est 400 Ω. On mesure la résistance de 16 projecteurs, les valeurs obtenues sont :

396	403	392	387	388	397	386	401	406	400	394	389	406	400	391	402

1/ Donner les estimations ponctuelles de la moyenne et la variance.

2/ Sachant que la distribution des résistances est normale, peut-on considérer, au seuil de signification égal à 5%, que le lot respecte la norme de 400 Ω ? Même question avec un seuil de égal à 1%.

Exercice 4 : Un rivet doit être inséré dans un trou. On mesure le diamètre du trou d'un échantillon aléatoire de taille $n = 15$ pièces. L'écart type des mesures du diamètre du trou est $S = 0.008\ mm$.

Si la l'écart type du diamètre du trou dépasse 0.01 mm, il existe une probabilité inacceptablement élevée que le rivet ne convient pas.

1/ Si $\alpha = 0.01$, peut-on montrer que l'écart type du diamètre du trou dépasse 0.01 mm ?

2/ Trouvez la valeur P pour ce test.

Exercice 5 : On joue avec un dé qui semble tomber trop souvent sur la face 6. Dans une expérience, on a lancé 40 fois ce dé et obtenu 10 fois le 6.

1/ Au seuil de signification $\alpha = 5\%$, peut-on conclure que le dé est truqué par excès d'apparition du 6 ?

2/ Sur 40 lancers, à partir de quelle proportion de faces 6 peut-on conclure que le dé est truqué (au seuil de signification de 5%) ?

4.5 Corrigés

Exercice 1 :

1/ $H_0 : m = 5$, $H_1 : m \neq 5$. Oui, c'est un test d'hypothèse bilatéral, car l'hypothèse est formulée en termes de paramètre d'intérêt. On a l'égalité dans l'hypothèse nulle et l'inégalité bilatérale dans l'hypothèse alternative.

2/ $H_0 : m = 5$, $H_1 : m < 5$. Oui, c'est un test d'hypothèse unilatéral, car l'hypothèse est formulée en termes de paramètre d'intérêt. On a l'égalité dans l'hypothèse nulle et l'inégalité unilatérale dans l'hypothèse alternative.

3/ $H_0 : \sigma < 2$, $H_1 : \sigma = 2$. Non, car l'hypothèse nulle ne peut pas être une inégalité.

4/ $H_0 : \bar{X} = 44$, $H_1 : \bar{X} \neq 44$. Non, parce que l'hypothèse est énoncée en termes de statistique (la moyenne empirique des observations) plutôt que le paramètre.

5/ $H_0 : p = 0.1$, $H_1 : p = 0.19$. Non, car les valeurs dans les deux hypothèses (nulle et alternative) ne correspondent pas et les deux hypothèses sont des énoncés d'égalité.

6/ $H_0 : S_n = 153$, $H_1 : S_n > 153$. Non, parce que l'hypothèse est énoncée en termes de statistique (la variance empirique des observations) plutôt que le paramètre.

Exercice 2 : On veut tester au seuil 5% l'hypothèse H_0 : la durée de vie des tubes $>$ 2000 $h \rightarrow$ un test unilatéral. On va donc chercher la valeur z telle que

$$P\left(m - \frac{zS}{\sqrt{n}} \leq \bar{X}\right) = 0.95$$

sachant que la statistique étudiée est

$$\mathcal{S} = \bar{X} - m \rightsquigarrow \mathcal{N}\left(0,\ \frac{S}{\sqrt{n}}\right),$$

on pose la nouvelle variable

$$Z = \frac{\bar{X} - m}{\frac{S}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0,\ 1).$$

On considère le test d'hypothèse

$$H_0 : m > m_0 = \ 2000\ h \text{ contre } H_1 : m < \ 2000\ h.$$

On a

$$\begin{aligned} P\left(m_0 - z\frac{S}{\sqrt{n}} \leq \bar{X}\right) &= P\left(z \geq \frac{\sqrt{n}}{S}\left(\bar{X} - m_0\right)\right) \\ &= 1 - P\left(z \leq \frac{-\sqrt{n}}{S}\left(\bar{X} - m_0\right)\right) \\ &= 1 - P\left(z > \frac{\sqrt{n}}{S}\left(\bar{X} - m_0\right)\right) \\ &= 0.95. \end{aligned}$$

D'où

$$P\left(z \leq \frac{\sqrt{n}}{S}\left(\bar{X} - m_0\right)\right) = 0.95.$$

De la table de la loi normale standard, on trouve $z = 1.645$. Pour un échantillon de taille $n = 100$, l'intervalle d'acceptation de la moyenne est :

$$\begin{aligned} IC &= \left[2000 - 1.645\frac{130}{10},\ +\infty\right[\\ &= [1978.62,\ +\infty[\,. \end{aligned}$$

Comme $\hat{X} = 1975 \notin IC$, alors H_0 est rejetée et probablement la déclaration du fabriquant n'est pas vraie.

Exercice 3 : On a $m_0 = 400\ \Omega$ et $n = 16$.

1/ Estimateurs de la moyenne et de la variance :

i/ La moyenne est estimée par la moyenne empirique

$$\begin{aligned} \bar{X} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\ &= \frac{6338}{16} \\ &= 396.125\ \Omega \\ &= \hat{X}. \end{aligned}$$

ii/ L'estimateur de la variance :

$$\begin{aligned} Var(X) &= \overline{X^2} - \left(\bar{X}\right)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\bar{X}\right)^2 \\ &= \frac{2511322}{16} - (396.125)^2 \\ &= 42.609\ \Omega^2. \end{aligned}$$

On prend

$$\begin{aligned} Var(X) &= 42.61\ \Omega^2 \\ &= S^2. \end{aligned}$$

Comme la variance de la popultaion n'est pas connue, elle est estimée par

$$\widehat{\sigma^2} = \frac{n}{n-1}S^2$$

où S^2 est la variance empirique. Alors, l'estimateur de la vriance est :

$$\begin{aligned}\widehat{\sigma^2} &= \frac{16}{15} \times 42.61\ \Omega^2 \\ &= 45.45\ \Omega^2.\end{aligned}$$

Par conséquent, l'estimateur de l'écart type est : $\boxed{\widehat{\sigma} = 6.742\ \Omega}$. La statistique étudiée est

$$\mathcal{S} = \bar{X} - m_0 \rightsquigarrow \mathcal{N}\left(0, \sqrt{\frac{\widehat{\sigma^2}}{n}}\right)$$

où

$$E(\mathcal{S}) = 0 \text{ et } Var(\mathcal{S}) = \frac{\widehat{\sigma^2}}{n}.$$

La nouvelle variable

$$\begin{aligned}Z &= \frac{\mathcal{S} - E(\mathcal{S})}{\sqrt{Var(\mathcal{S})}} \\ &= \frac{\bar{X} - m_0}{\sqrt{\frac{\widehat{\sigma^2}}{n}}}\end{aligned}$$

suit la loi de Student à $(n-1)$ degrés de liberté.

2/ L'hypothèse nulle H_0 : Le lot respecte la norme de 400 $\Omega \to H_0 : m = m_0$. L'intervalle de confiance pour la moyenne m est :

$$IC = \left[m_0 - t_{\alpha,\ n-1}\frac{\hat{\sigma}}{\sqrt{n}},\ m_0 + t_{\alpha,\ n-1}\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

où $t_{\ n-1,\ \alpha}$ est la valeur lue sur la table de student à 15 degrés de liberté.

i/ Si $\alpha = 0.05$, on trouve

$$t_{\ 15,\ 0.05} = 2.131.$$

Alors,

$$\begin{aligned}IC &= \left[400 - 2.131\frac{6.742}{\sqrt{16}},\ 400 + 2.131\frac{6.742}{\sqrt{16}}\right] \\ &= [396.4\ \Omega,\ 403.59\ \Omega]\,.\end{aligned}$$

L'estimateur calculé de la moyenne $\hat{X} = 396.125\ \Omega \notin IC$, donc l'hypothèse H_0 est rejetée au risque 5%.

ii/ Pour un seuil $\alpha = 0.01$, on trouve

$$t_{\ 15,\ 0.01} = 2.947.$$

et l'intervalle de confiance devient

$$IC = [395.03\ \Omega,\ 404.97\ \Omega].$$

On remarque que l'estimateur calculé de la moyenne $\hat{X} = 396.125\ \Omega \in\ IC$, alors l'hypothèse H_0 est acceptée au risque 1%.

<u>**Exercice 4 :**</u> Afin d'utiliser la statistique χ^2_{n-1} dans les tests d'hypothèses et la construction d'intervalles de confiance, on doit supposer que la distribution parente est normale.

1/ Le paramètre d'intérêt est le véritable écart type du diamètre. Cependant, on effectue un test d'hypothèse sur σ^2.

$$H_0 : \sigma^2 = 0.0001 \text{ contre } H_0 : \sigma^2 > 0.0001$$

avec le seuil de risque $\alpha = 0.01$. On a

$$\begin{aligned} \chi^2_{n-1} &= \frac{(n-1)\,S^2}{\sigma^2} \\ &= \frac{14 \times (0.008)^2}{0.0001} \\ &= 8.96 \\ &= \varepsilon_{obs} \end{aligned}$$

et la valeur $\chi^2_{14,\ 0.01}$ lue sur la table de Khi-deux est $\chi^2_{14,\ 0.01} = 29.14$. Comme $\varepsilon_{obs} < \chi^2_{14,\ 0.01}$, on rejette pas H_0. Alors, on ne peut pas montrer que l'écart type du diamètre du trou dépasse $0,01\ mm$ au seuil $\alpha = 0.01$.

2/ La valeur P de ce test est donnée par :

$$P = P\left(\chi^2_{14} > 8.96\right)$$

d'où

$$0.5 < P < 0.9.$$

Exercice 5 : La taille de l'échantillon $n = 40$ et la proportion p est :

$$\begin{aligned} p &= \frac{n_i}{n} \\ &= \frac{1}{4}. \end{aligned}$$

1/ Test d'hypothèse

$$\begin{aligned} H_0 &: \text{ Le dé n'est pas truqué } p_0 = \frac{1}{6} \\ & \text{contre} \\ H_1 &: \text{ Le dé est truqué } p \neq p_0. \end{aligned}$$

$$\text{Echantillonnage : } n = 40,\ p = \frac{1}{4}.$$

On a la statistique $\mathcal{S} = p - p_0$ suit la loi normale $\mathcal{N}\left(0,\ \frac{\sigma}{n}\right)$. On pose

$$\begin{aligned} Z &= \frac{\mathcal{S} - E(\mathcal{S})}{\sigma_{\mathcal{S}}} \\ &= \frac{|p - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}}. \end{aligned}$$

Il est clair que $Z \rightsquigarrow \mathcal{N}(0,\ 1)$.

$$\begin{aligned} Z &= \frac{\left|\frac{1}{4} - \frac{1}{6}\right|}{\sqrt{\frac{\frac{1}{6} \times \frac{5}{6}}{40}}} \\ &= 1.41 \\ &= \varepsilon_{obs}. \end{aligned}$$

ε_{seuil} : est la valeur lue sur la table de la loi normale standard et qui correspond au seuil de signification $\alpha = 5\%$. De la table on trouve $\varepsilon_{seuil} = 1.645$.

Décision : $\varepsilon_{obs} < \varepsilon_{seuil}$, alors l'hypothèse nulle H_0 est acceptée $\longrightarrow$ on ne peut pas dire que le dé est truqué.

2/ Le dé est truqué si H_0 est rejetée $\Rightarrow$ $\varepsilon_{obs} > \varepsilon_{seuil}$.

$$\frac{|p - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} > 1.645$$

ce qui implique

$$\begin{aligned} p &> 1.645 \times \sqrt{\frac{p_0(1-p_0)}{n}} + p_0 \\ &> 1.645 \times \sqrt{\frac{\frac{1}{6}\times\frac{5}{6}}{40}} + \frac{1}{6} \\ &> 0.263. \end{aligned}$$

Comme $p = \frac{n_i}{n}$, alors

$$\frac{n_i}{40} > 0.263 \Rightarrow n_i > 10.52.$$

On prend donc, $n_i = 11$ fois.

4.6 Devoir N 3

Exercice 1 : On a soumis 15 copies d'examen à une double correction A et B, on veut tester si les deux séries de notes sont semblables (de même espérance mathématique).

A	$13,5$	12	$14,5$	15	$11,5$
B	$15,5$	$11,5$	14	$15,5$	13
n_i	1	2	3	4	5

Soit X la variable correspondante à la série A et Y la variable correspondante à B. La variable $X - Y$ suit la loi normale $\mathcal{N}(m_1 - m_2, \sigma)$ où m_1 et m_2 sont les moyennes de X et Y respectivement et l'écart type σ est inconnu. Son estimateur $\hat{\sigma} = \sqrt{\frac{n}{n-1}}S$, où S est la variance empirique de la variable $X_1 - X_2$.

Tester l'hypothèse

$$H_0: \ E(X) = E(Y),$$

au risque d'erreur $\alpha_1 = 2\%$ puis $\alpha_2 = 1\%$.

Exercice 2 : Un sondage a été effectué par un institut pour connaitre si les conducteurs favorisent l'utilisation d'essence sans plomb. Un échantillon de 500 personnes a participé à ce sondage.

Si plus de 400 personnes répondent positivement, on calcule qu'au moins 60% des conducteurs sont favorables à l'utilisation de ce carburant.

1/ Trouvez la probabilité d'erreur de type I si exactement 60% des conducteurs sont favorables à l'utilisation de ce carburant.

2/ Quelle est la probabilité d'erreur de type II si 75% des conducteurs favorisent cette action ?

Exercice 3 : Une entreprise de pharmaceutique fabrique un produit thérapeutique dont la concentration a un écart type de $\sigma = 4\ g/l$. Une nouvelle méthode de fabrication de ce produit a été proposée, bien que certains coûts supplémentaires soient impliqués. La gestion n'autorisera un changement de technique de production que si le l'écart type de la concentration dans le nouveau procédé est moins de 4 g/l. Les chercheurs ont choisi un échantillon de taille $n = 10$ et ils ont obtenu les données suivantes en g/l.

16.628	16.622	16.627	16.623	16.618	16.630	16.631	16.624	16.622	16.626

Effectuer l'analyse nécessaire pour déterminer si un changement dans la production technique doit être mise en œuvre.

Annexe 1 : Table de La loi normale Standard

Fonction de répartition de la loi normale standard

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx$$

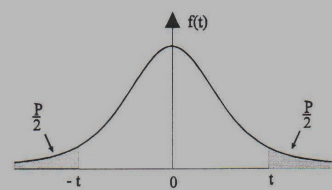
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

z	0.841	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
$\Phi(z)$	0.8000	0.9000	0.9500	0.9750	0.9800	0.9900	0.9950	0.9975	0.9990	0.9995

Annexe 2 : Table de Student

Table de la loi de Student

Valeurs de **T** ayant la probabilité **P** d'être dépassées en valeur absolue



ν	P =0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,260	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
∞	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,96	2,326	2,576

Nota. ν est le nombre de degrés de liberté.
Le quantile d'ordre $1 - \frac{\alpha}{2}$ se lit dans la colonne $P = \alpha$.
Le quantile d'ordre $1 - \alpha$ se lit dans la colonne $P = 2\alpha$.

Annexe 3 : Table de Khi-deux

Le tableau ci-dessous donne la valeur de x telle que $P(k > x) = p$.

p ddl	0,999	0,995	0,99	0,98	0,95	0,9	0,8	0,2	0,1	0,05	0,02	0,01	0,005	0,001
1	0,0000	0,0000	0,0002	0,0006	0,0039	0,0158	0,0642	1,6424	2,7055	3,8415	5,4119	6,6349	7,8794	10,8276
2	0,0020	0,0100	0,0201	0,0404	0,1026	0,2107	0,4463	3,2189	4,6052	5,9915	7,8240	9,2103	10,5966	13,8155
3	0,0243	0,0717	0,1148	0,1848	0,3518	0,5844	1,0052	4,6416	6,2514	7,8147	9,8374	11,3449	12,8382	16,2662
4	0,0908	0,2070	0,2971	0,4294	0,7107	1,0636	1,6488	5,9886	7,7794	9,4877	11,6678	13,2767	14,8603	18,4668
5	0,2102	0,4117	0,5543	0,7519	1,1455	1,6103	2,3425	7,2893	9,2364	11,0705	13,3882	15,0863	16,7496	20,5150
6	0,3811	0,6757	0,8721	1,1344	1,6354	2,2041	3,0701	8,5581	10,6446	12,5916	15,0332	16,8119	18,5476	22,4577
7	0,5985	0,9893	1,2390	1,5643	2,1673	2,8331	3,8223	9,8032	12,0170	14,0671	16,6224	18,4753	20,2777	24,3219
8	0,8571	1,3444	1,6465	2,0325	2,7326	3,4895	4,5936	11,0301	13,3616	15,5073	18,1682	20,0902	21,9550	26,1245
9	1,1519	1,7349	2,0879	2,5324	3,3251	4,1682	5,3801	12,2421	14,6837	16,9190	19,6790	21,6660	23,5894	27,8772
10	1,4787	2,1559	2,5582	3,0591	3,9403	4,8652	6,1791	13,4420	15,9872	18,3070	21,1608	23,2093	25,1882	29,5883
11	1,8339	2,6032	3,0535	3,6087	4,5748	5,5778	6,9887	14,6314	17,2750	19,6751	22,6179	24,7250	26,7568	31,2641
12	2,2142	3,0738	3,5706	4,1783	5,2260	6,3038	7,8073	15,8120	18,5493	21,0261	24,0540	26,2170	28,2995	32,9095
13	2,6172	3,5650	4,1069	4,7654	5,8919	7,0415	8,6339	16,9848	19,8119	22,3620	25,4715	27,6882	29,8195	34,5282
14	3,0407	4,0747	4,6604	5,3682	6,5706	7,7895	9,4673	18,1508	21,0641	23,6848	26,8728	29,1412	31,3193	36,1233
15	3,4827	4,6009	5,2293	5,9849	7,2609	8,5468	10,3070	19,3107	22,3071	24,9958	28,2595	30,5779	32,8013	37,6973
16	3,9416	5,1422	5,8122	6,6142	7,9616	9,3122	11,1521	20,4651	23,5418	26,2962	29,6332	31,9999	34,2672	39,2524
17	4,4161	5,6972	6,4078	7,2550	8,6718	10,0852	12,0023	21,6146	24,7690	27,5871	30,9950	33,4087	35,7185	40,7902
18	4,9048	6,2648	7,0149	7,9062	9,3905	10,8649	12,8570	22,7595	25,9894	28,8693	32,3462	34,8053	37,1565	42,3124
19	5,4068	6,8440	7,6327	8,5670	10,1170	11,6509	13,7158	23,9004	27,2036	30,1435	33,6874	36,1909	38,5823	43,8202
20	5,9210	7,4338	8,2604	9,2367	10,8508	12,4426	14,5784	25,0375	28,4120	31,4104	35,0196	37,5662	39,9968	45,3147
21	6,4467	8,0337	8,8972	9,9146	11,5913	13,2396	15,4446	26,1711	29,6151	32,6706	36,3434	38,9322	41,4011	46,7970
22	6,9830	8,6427	9,5425	10,6000	12,3380	14,0415	16,3140	27,3015	30,8133	33,9244	37,6595	40,2894	42,7957	48,2679
23	7,5292	9,2604	10,1957	11,2926	13,0905	14,8480	17,1865	28,4288	32,0069	35,1725	38,9683	41,6384	44,1813	49,7282
24	8,0849	9,8862	10,8564	11,9918	13,8484	15,6587	18,0618	29,5533	33,1962	36,4150	40,2704	42,9798	45,5585	51,1786
25	8,6493	10,5197	11,5240	12,6973	14,6114	16,4734	18,9398	30,6752	34,3816	37,6525	41,5661	44,3141	46,9279	52,6197
26	9,2221	11,1602	12,1981	13,4086	15,3792	17,2919	19,8202	31,7946	35,5632	38,8851	42,8558	45,6417	48,2899	54,0520
27	9,8028	11,8076	12,8785	14,1254	16,1514	18,1139	20,7030	32,9117	36,7412	40,1133	44,1400	46,9629	49,6449	55,4760
28	10,3909	12,4613	13,5647	14,8475	16,9279	18,9392	21,5880	34,0266	37,9159	41,3371	45,4188	48,2782	50,9934	56,8923
29	10,9861	13,1211	14,2565	15,5745	17,7084	19,7677	22,4751	35,1394	39,0875	42,5570	46,6927	49,5879	52,3356	58,3012
30	11,5880	13,7867	14,9535	16,3062	18,4927	20,5992	23,3641	36,2502	40,2560	43,7730	47,9618	50,8922	53,6720	59,7031
40	17,9164	20,7065	22,1643	23,8376	26,5093	29,0505	32,3450	47,2685	51,8051	55,7585	60,4361	63,6907	66,7660	73,4020
50	24,6739	27,9907	29,7067	31,6639	34,7643	37,6886	41,4492	58,1638	63,1671	67,5048	72,6133	76,1539	79,4900	86,6608
60	31,7383	35,5345	37,4849	39,6994	43,1880	46,4589	50,6406	68,9721	74,3970	79,0819	84,5799	88,3794	91,9517	99,6072
70	39,0364	43,2752	45,4417	47,8934	51,7393	55,3289	59,8978	79,7146	85,5270	90,5312	96,3875	100,4252	104,2149	112,3169
80	46,5199	51,1719	53,5401	56,2128	60,3915	64,2778	69,2069	90,4053	96,5782	101,8795	108,0693	112,3288	116,3211	124,8392
90	54,1552	59,1963	61,7541	64,6347	69,1260	73,2911	78,5584	101,0537	107,5650	113,1453	119,6485	124,1163	128,2989	137,2084
100	61,9179	67,3276	70,0649	73,1422	77,9295	82,3581	87,9453	111,6667	118,4980	124,3421	131,1417	135,8067	140,1695	149,4493
120	77,7551	83,8516	86,9233	90,3667	95,7046	100,6236	106,8056	132,8063	140,2326	146,5674	153,9182	158,9502	163,6482	173,6174
140	93,9256	100,6548	104,0344	107,8149	113,6593	119,0293	125,7581	153,8537	161,8270	168,6130	176,4709	181,8403	186,8468	197,4508
160	110,3603	117,6793	121,3456	125,4400	131,7561	137,5457	144,7834	174,8283	183,3106	190,5165	198,8464	204,5301	209,8239	221,0190
180	127,0111	134,8844	138,8204	143,2096	149,9688	156,1526	163,8682	195,7434	204,7037	212,3039	221,0772	227,0561	232,6198	244,3705
200	143,8428	152,2410	156,4320	161,1003	168,2786	174,8353	183,0028	216,6088	226,0210	233,9943	243,1869	249,4451	255,2642	267,5405
250	186,5541	196,1606	200,9386	206,2490	214,3916	221,8059	231,0128	268,5986	279,0504	287,8815	298,0388	304,9396	311,3462	324,8324
300	229,9634	240,6634	245,9725	251,8637	260,8781	269,0679	279,2143	320,3971	331,7885	341,3951	352,4246	359,9064	366,8444	381,4252
400	318,2596	330,9028	337,1553	344,0781	354,6410	364,2074	376,0218	423,5895	436,6490	447,6325	460,2108	468,7245	476,6064	493,1318
500	407,9470	422,3034	429,3875	437,2194	449,1468	459,9261	473,2099	526,4014	540,9303	553,1268	567,0698	576,4928	585,2066	603,4460
600	498,6229	514,5289	522,3651	531,0191	544,1801	556,0560	570,6680	628,9433	644,8004	658,0936	673,2703	683,5156	692,9816	712,7712
700	590,0480	607,3795	615,9075	625,3175	639,6130	652,4973	668,3308	731,2805	748,3591	762,6607	778,9721	789,9735	800,1314	821,3468
800	682,0665	700,7250	709,8969	720,0107	735,3623	749,1852	766,1555	833,4557	851,6712	866,9114	884,2789	895,9843	906,7862	929,3289
900	774,5698	794,4750	804,2517	815,0267	831,3702	846,0746	864,1125	935,4987	954,7819	970,9036	989,2631	1001,6296	1013,0364	1036,8260

Bibliographie

[1] D. P. Bertsekas and J. N. Tsitsiklis, Introduction to probability, Lecture notes Course, Massachusetts Institute of Technology, Cambridge, Massachusetts (2000).

[2] P. Hoel, Introduction to mathematical statistics, 3^{rd} edition, University of California, Sixth priting, USA, (1966).

[3] M. Lethilleux, Exercices de statistique et probabilités avec rappels de cours, $2^{ème}$ edition, ISBN 978-2-10-057979-2, Dunod, (2012).

[4] D. C. Montgomery et G. Runger, Applied statistics and probability for engineers, 3^{rd} edition, Arizona state university, ISBN 0-471-20454-4, USA (2002).

[5] T. Phan et J. P. Rowenczyk, Exervices et problèmes de statistique et probabilité, ISBN 978-2-10-056298-5, Dunod, (2012).

[6] Ve. Renée, Statistique et probabilités pour l'ingénieur, $2^{ème}$ edition, ISBN 210-049-9947 Dunod, (2014).