
Statistique inférentielle
ACTUARIAT

PROF. ARMEL YODÉ

Il y'a des vides à combler dans ce fascicule.
Venez à l'école !

Table des matières

1 Rappels et compléments	4
1.1 Loi normale	4
1.2 Convergences	8
1.3 Théorèmes limites	9
1.3.1 Lois des grands nombres	9
1.3.2 Théorème Central limite	10
1.3.3 Théorème de Slutsky et généralisation	11
1.3.4 La méthode delta	12
2 Modélisation statistique	13
2.1 Population, échantillon	13
2.2 Echantillonnage aléatoire	14
2.2.1 Echantillonnage aléatoire simple	14
2.2.2 Echantillonnage systématique	14
2.2.3 Echantillonnage stratifié	14
2.2.4 Echantillonnage par grappes	15
2.3 Modèles statistiques	16
3 Exhaustivité et Information de Fisher	18
3.1 Vraisemblance	18
3.2 Exhaustivité	19
3.3 Information de Fisher	20
4 Estimateurs	22
4.1 Qu'est ce qu'un estimateur ?	22
4.2 Qu'est ce qu'un "bon" estimateur ?	22
4.2.1 Propriétés à distance finie	22
4.2.1.1 Biais	22
4.2.1.2 Risque quadratique	23
4.2.1.3 Borne de Cramer-Rao	24
4.2.2 Propriétés asymptotiques	25
4.2.2.1 Convergence ou consistance	25
4.2.2.2 Normalité asymptotique	26
5 Méthodes d'estimation	27
5.1 Méthode des moments	27
5.2 Méthode du maximum de vraisemblance	28
5.3 Méthode des moindres carrés ordinaire et régression linéaire	31

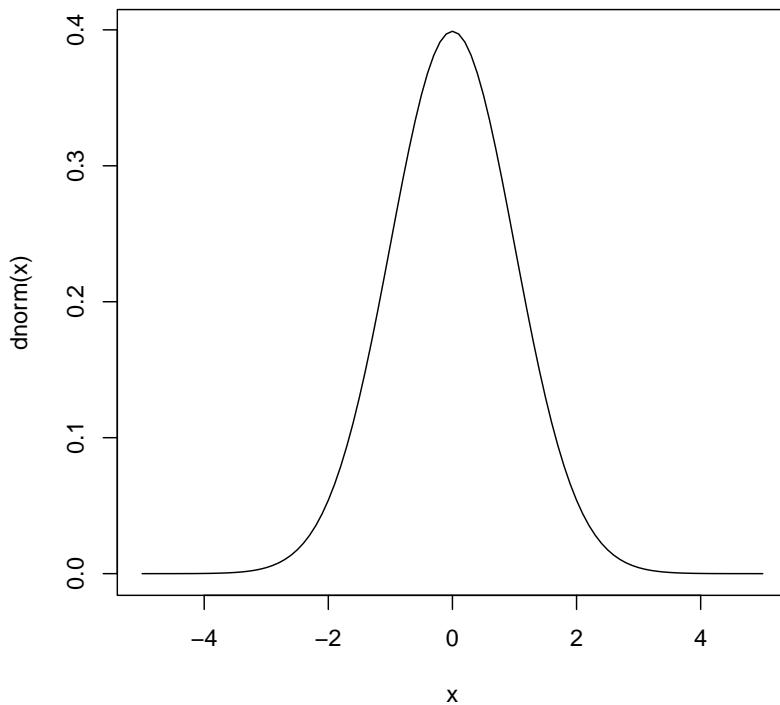
TABLE DES MATIÈRES	3
6 Estimation par intervalle de confiance	33
6.1 Introduction	33
6.2 Construction d'un intervalle de confiance	34
6.2.1 Fonction pivotale	34
6.2.2 Construction d'un intervalle de confiance bilatéral	34
6.2.2.1 Méthode non asymptotique	34
6.2.2.2 Méthode asymptotique	35
6.2.3 Densité de probabilité unimodale	35
6.3 Exemples	35
6.3.1 Intervalle de confiance pour la moyenne d'une loi normale	35
6.3.2 Intervalle de confiance pour la variance d'une loi normale	40
6.3.3 Intervalle de confiance pour une proportion	41
6.3.4 Intervalle de confiance pour la moyenne d'une loi quelconque	43
7 Généralités sur les tests d'hypothèses	45
7.1 Principe des tests	45
7.2 Etapes des tests	47
7.3 Construction d'un test d'hypothèses	47
7.4 La p -value	48
8 Tests de Student : un échantillon	49
8.1 Introduction	49
8.2 $H_0 : m \leq m_0$ contre $H_1 : m > m_0$	49
8.2.1 On suppose que la variance σ^2 est connue.	49
8.2.2 On suppose σ^2 est inconnue	51
8.3 $H_0 : m \geq m_0$ contre $H_1 : m < m_0$	53
8.3.1 On suppose que la variance σ^2 est connue.	53
8.3.2 On suppose que la variance σ^2 est inconnue.	54
8.4 $H_0 : m = m_0$ contre $H_1 : m \neq m_0$	56
8.4.1 On suppose que la variance σ^2 est inconnue.	57
9 Tests de Student : deux échantillons	59
9.1 Introduction	59
9.2 Test de Fisher de comparaison des variances	60
9.3 Test de Student de comparaison des moyennes	60
9.3.1 Résolution du test lorsque les variances connues	61
9.3.2 Résolution du test lorsque les variances sont inconnues	61
10 Tests de comparaison des proportions	63
10.1 Test sur la valeur d'une proportion	63
10.2 Test de comparaison de deux proportions	64
11 Tests du χ^2	68
11.1 Test d'adéquation à une loi donnée	68
11.1.1 Cas d'une loi discrète	68
11.1.2 Cas d'une loi continue	70
11.2 Test d'adéquation à une famille de lois	70
11.3 Test d'indépendance	71

1.1 Loi normale

On dit qu'une variable aléatoire X suit une loi normale ou gaussienne si sa densité de probabilité est

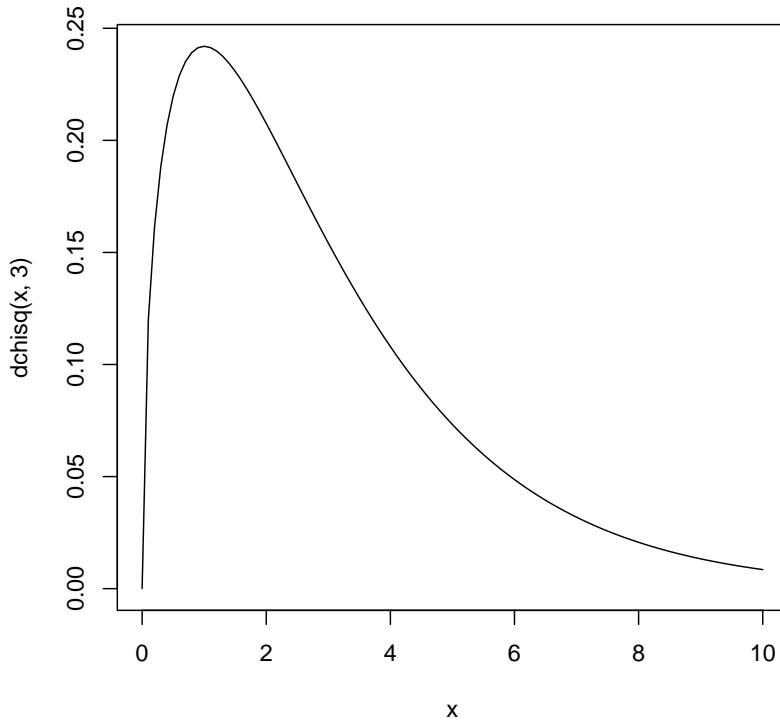
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

On note $\mathcal{N}(m, \sigma^2)$ la loi normale de moyenne m et de variance σ^2 . La loi normale est dite centrée-réduite si $m = 0$ et $\sigma^2 = 1$. Voici la courbe de la densité de la loi normale centrée-réduite. Ci-dessous, la courbe de la densité de la loi normale centrée-réduite.

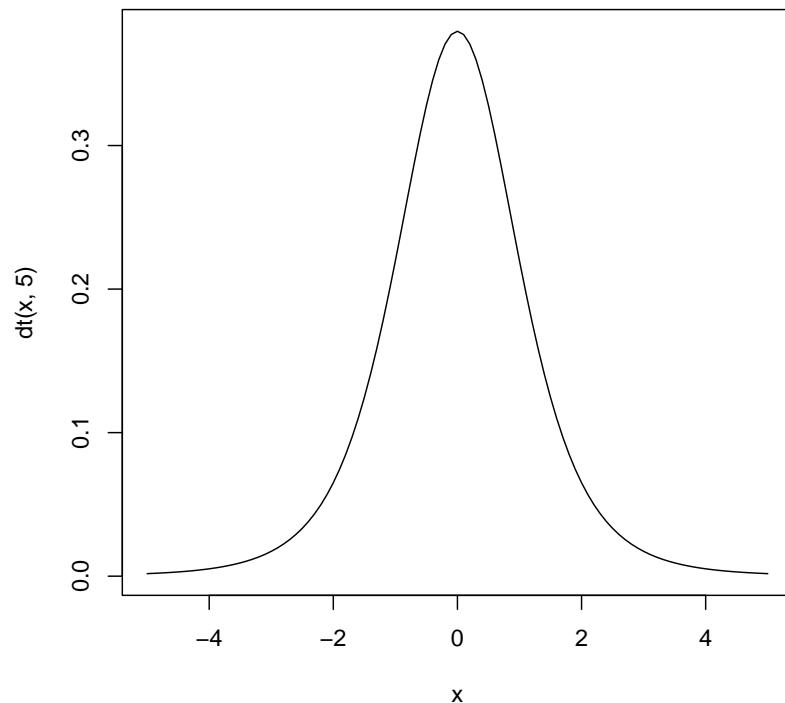


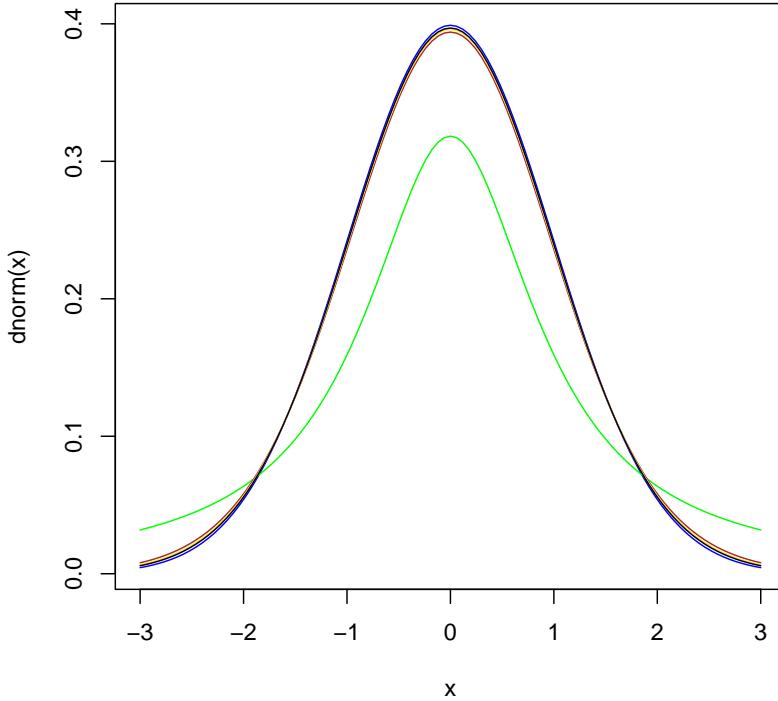
Ci-dessous, la courbe de la densité de la loi de khi-deux à 3 degrés de liberté

Loi de du khi-deux à n degrés de liberté.



Loi de Student à n degrés de liberté.





Théorème 1.1.1. Soient X_1, \dots, X_n des variables aléatoires indépendantes identiquement distribuées de loi normale $\mathcal{N}(m, \sigma^2)$ avec $\mu \in \mathbb{R}$ et $\sigma^2 > 0$. Posons

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

Alors nous avons :

1. S_n^2 et \bar{X}_n sont indépendantes
2. $\bar{X}_n \hookrightarrow \mathcal{N}\left(m, \frac{\sigma^2}{n}\right)$.
3. $\frac{(n-1)S_n^2}{\sigma^2} \hookrightarrow \chi^2(n-1)$ (loi de Khi-deux à $n-1$ degrés de liberté).
4. $\frac{\sqrt{n}(\bar{X}_n - m)}{S_n} \hookrightarrow T(n-1)$ (loi de Student à $n-1$ degrés de liberté).

1.2 Convergences

On considère une suite de variables aléatoires réelles $(X_n)_{n \geq 1}$ définies sur le même espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$. On note F_{X_n} la fonction de répartition de X_n et F_X celle de X .

Définition 1.2.1. On dit que la suite $(X_n)_{n \geq 1}$ converge en loi vers la variable aléatoire X et on note $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$ si

$$\lim_{n \rightarrow +\infty} F_{X_n}(x) = F_X(x)$$

en tout point x où F_X est continue.

Définition 1.2.2. On dit que la suite $(X_n)_{n \geq 1}$ converge en probabilité vers un réelle a et on note $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X$, si quelque soit $\varepsilon > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}\{|X_n - X| \geq \varepsilon\} = 0.$$

Remarque 1.2.1. La convergence en probabilité implique la convergence en loi :

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X \implies X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X.$$

Cependant si $X = a$ où a est une constante, alors il y a équivalence entre les deux modes de convergence

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} a \iff X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} a.$$

Proposition 1.2.1. Soit $a \in \mathbb{R}$. Si $\begin{cases} \mathbb{E}(X_n) \xrightarrow[n \rightarrow +\infty]{\longrightarrow} a \\ \text{Var}(X_n) \xrightarrow[n \rightarrow +\infty]{\longrightarrow} 0 \end{cases}$ alors $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} a$.

Démonstration. En appliquant l'inégalité de Markov, nous obtenons

$$\begin{aligned} 0 \leq \mathbb{P}(|X_n - a| > \varepsilon) &\leq \frac{1}{\varepsilon^2} \mathbb{E}(|X_n - a|^2) \\ &= \frac{1}{\varepsilon^2} (\mathbb{E}(X_n - \mathbb{E}(X_n))^2 + (\mathbb{E}(X_n) - a)^2) \\ &= \frac{1}{\varepsilon^2} (\text{Var}(X_n) + (\mathbb{E}(X_n) - a)^2). \end{aligned}$$

Ainsi, nous remarquons que si $\begin{cases} \mathbb{E}(X_n) \xrightarrow[n \rightarrow +\infty]{\longrightarrow} a \\ \text{Var}(X_n) \xrightarrow[n \rightarrow +\infty]{\longrightarrow} 0 \end{cases}$, alors d'après le Théorème des gendarmes

$$\lim_{n \rightarrow +\infty} \mathbb{P}\{|X_n - X| \geq \varepsilon\} = 0.$$

□

1.3 Théorèmes limites

Soient X_1, \dots, X_n des variables aléatoires indépendantes identiquement distribuées de moyenne m et de variance $\sigma^2 > 0$. Posons

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

Nous nous intéressons à deux résultats importants concernant la moyenne empirique \bar{X}_n de variables aléatoires indépendantes identiquement distribuées.

1.3.1 Lois des grands nombres

Théorème 1.3.1. Soient $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes identiquement distribuées telles que $\mathbb{E}(X_1) = m < +\infty$ et $\text{Var}(X_n) = \sigma^2$. Alors, nous avons

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} m.$$

Démonstration. Nous utilisons la proposition 1.2.1. En effet, nous avons $\mathbb{E}(\bar{X}_n) = m$ et $\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$.

□

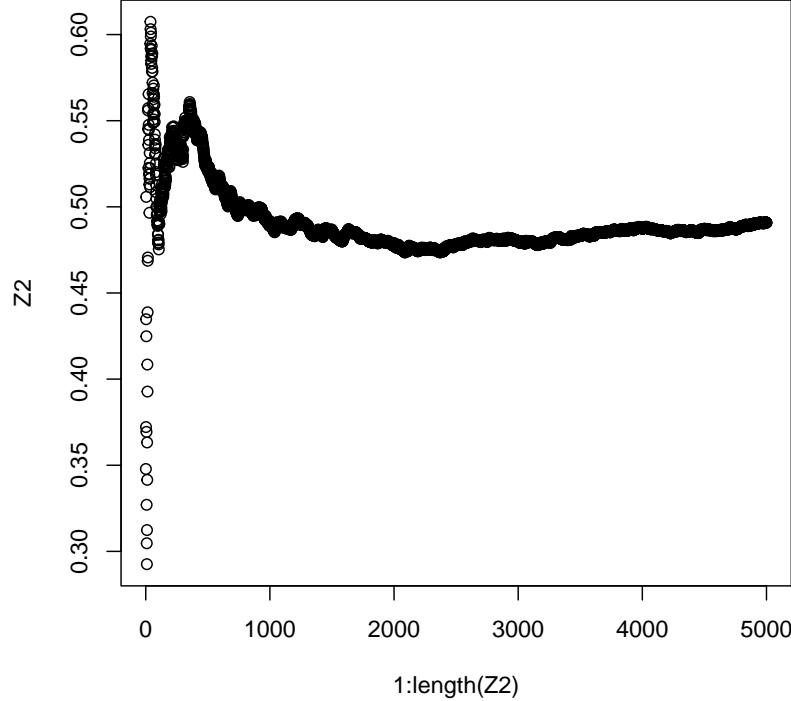
Ce résultat signifie que lorsque n devient grand, la moyenne empirique \bar{X}_n se réduit "presque" à la moyenne théorique m .

Illustration de la loi des grands nombres

Dans cet exemple $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes identiquement de loi exponentielle $\mathcal{E}(2)$. D'après la loi des grands nombres

$$\bar{X}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \frac{1}{2}.$$

Ce résultat est illustré par le graphique ci-dessous.



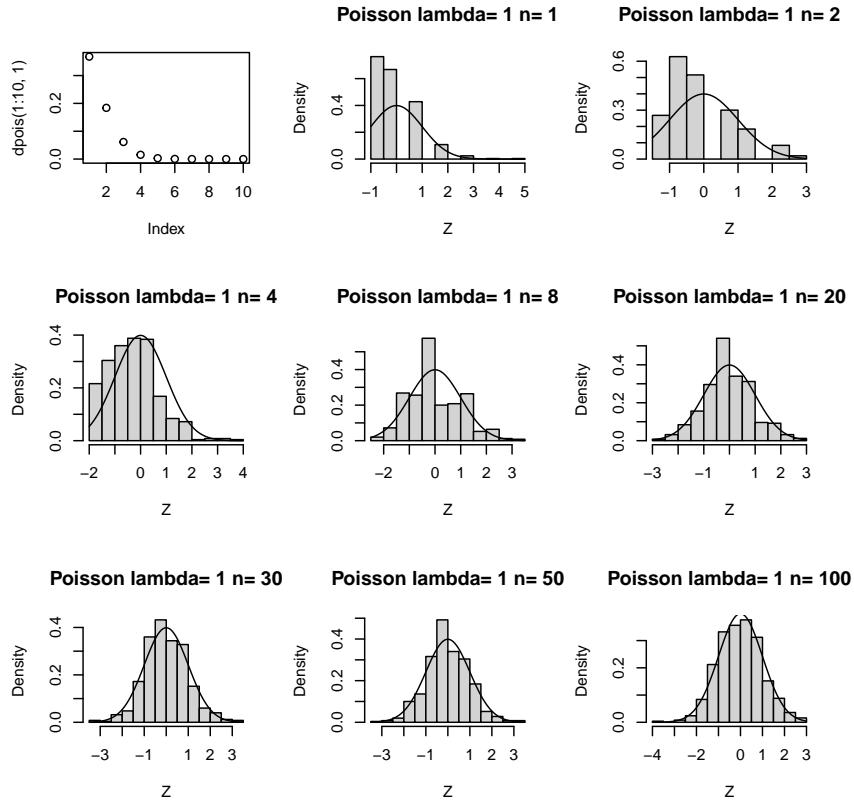
1.3.2 Théorème Central limite

Le théorème central limite permet d'étudier la convergence en loi de la moyenne empirique \bar{X}_n .

Théorème 1.3.2. Soient $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles indépendantes identiquement distribuées telles que $\mathbb{E}(X_1) = m < +\infty$ et $\sigma^2 = \text{var}(X_1) > 0$. Alors, nous avons

$$\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1) \iff \sqrt{n}(\bar{X}_n - m) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2).$$

Illustration du théorème central limite



Autrement dit, quand n est assez grand $\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ converge vers la loi normale centrée réduite $\mathcal{N}(0,1)$, c'est à dire que la moyenne empirique \bar{X}_n suit approximativement une loi normale $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. En pratique, l'approximation est fréquemment réalisée dès que $n \geq 30$.

1.3.3 Théorème de Slutsky et généralisation

Théorème 1.3.3. Soient X_n et Y_n deux suites de variables aléatoires telles que :

$$X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X$$

$$Y_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} c$$

où c est une constante. Alors

$$X_n + Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X + c$$

$$X_n Y_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} cX$$

$$\frac{X_n}{Y_n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \frac{X}{c} \quad \text{si } c \neq 0.$$

L'on peut généraliser ces résultats. Quelle condition doit vérifier une fonction g pour que $g(X_n)$ converge en loi (ou en probabilité) vers $g(X)$ dès que X_n converge en loi (ou en probabilité) vers X . Le résultat suivant permet de répondre à cette question.

Théorème 1.3.4. Soit g est une fonction continue. Alors

- $X_n \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} X \Rightarrow g(X_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} g(X).$
- $X_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} X \Rightarrow g(X_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} g(X).$

1.3.4 La méthode delta

Si

$$\sqrt{n}(Y_n - y) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_y^2),$$

quelle est la loi asymptotique de la variable aléatoire $\sqrt{n}(g(Y_n) - g(y))$? C'est à dire,

$$\sqrt{n}(g(Y_n) - g(y)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} ?$$

Quelles sont les conditions sur la fonction g ? La méthode delta permet de répondre à ce type de préoccupations.

Théorème 1.3.5. Si la suite de variables aléatoires (Y_n) est asymptotiquement normale, telle qu'il existe y et σ_y^2 avec

$$\sqrt{n}(Y_n - y) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_y^2)$$

et si g est une fonction de classe \mathcal{C}^1 alors $g(Y_n)$ est asymptotiquement normal

$$\sqrt{n}(g(Y_n) - g(y)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_y^2(g'(y))^2).$$

2.1 Population, échantillon

Exemple 2.1.1. Supposons que nous voulions caractériser la distribution des poids de naissance de tous les enfants nés vivants en Côte d'Ivoire en 2021. Supposons que cette distribution du poids de naissance a une moyenne m et une variance σ^2 . Idéalement, nous souhaitons estimer m et σ^2 exactement, sur la base de l'ensemble de la population des enfants nés vivants en Côte d'Ivoire en 2021. Mais cette tâche est difficile avec un groupe aussi important. Au lieu de cela, nous décidons de sélectionner un échantillon aléatoire de n nourrissons qui sont représentatifs de ce grand groupe et d'utiliser les poids de naissance x_1, \dots, x_n de cet échantillon pour nous aider à estimer m et σ^2 .

Qu'est-ce qu'un échantillon aléatoire ?

La population est l'ensemble sur lequel porte l'étude statistique. Un élément de la population est appelé individu. Le terme individu est à prendre au sens large : il peut s'agir d'une personne physique mais aussi d'un logement, d'une entreprise, d'un mouton, etc. La définition précise de la population constitue une tâche préalable nécessaire et plutôt difficile. Formellement, deux possibilités permettent de décrire une population :

- en extension prenant la forme d'une liste complète d'individus
- en compréhension obtenue par une phrase descriptive.

Définition 2.1.1. Un échantillon aléatoire est une sélection de certains membres de la population de telle sorte que chaque membre est choisi indépendamment et a une probabilité connue non nulle d'être sélectionné.

Il n'y a pas d'intervention du chercheur. Seul le hasard regit l'inclusion ou non d'un individu dans l'échantillon. Les informations recueillies sur l'échantillon peuvent être inférées à la population.

Définition 2.1.2. Un échantillon aléatoire simple est un échantillon aléatoire dans lequel chaque membre du groupe a la même probabilité d'être sélectionné.

Comment selectionne-t-on un échantillon aléatoire ?

2.2 Echantillonnage aléatoire

L'échantillon doit être représentatif de la population. Il doit donc présenter, pour les caractéristiques qui sont importantes pour l'étude, des propriétés qui soient le plus proche possible de celles de la population dont il est extrait. Dans le cas contraire, l'échantillon est biaisé et les résultats de l'étude seront faussés.

2.2.1 Echantillonnage aléatoire simple

Cette méthode est appropriée lorsque la population est nombreuse et relativement homogène. Ce choix peut se faire avec remise ou sans remise :

- avec remise, un individu peut être choisi plusieurs fois
- sans remise, un individu déjà choisi ne peut l'être de nouveau.

En pratique, lorsque la population a un effectif très élevé, on tire seulement un faible nombre d'éléments et l'on assimile un tirage sans remise à un tirage avec remise.

Procédure à suivre pour l'échantillonnage aléatoire simple

1. Définir clairement la nature de la population
2. Assigner un numéro à chaque individu de la population (1 à N)
3. Sélectionner l'échantillon en choisissant n'importe quelle méthode qui donne une chance égale à tous les numéros d'être tirés.

2.2.2 Echantillonnage systématique

L'échantillonnage systématique est une méthode qui exige aussi l'existence d'une liste de la population où chaque individu est numéroté de 1 jusqu'à N . Notons n , la taille de l'échantillon. L'entier voisin de N/n sera noté r et appelé pas de sondage. Pour constituer l'échantillon il faut :

Procédure à suivre pour l'échantillonnage systématique

1. Numéroter de 1 à N les individus de la base de sondage.
2. Déterminer le pas de sondage en divisant N par la taille de l'échantillon n : $r = N/n$.
3. Sélectionner un entier naturel d au hasard entre 1 et r . Ce nombre s'appelle l'origine. L'individu dont le numéro correspond à d est le premier individu ;
4. pour sélectionner les autres, il suffit d'ajouter à d le pas de sondage r .

2.2.3 Echantillonnage stratifié

Cette méthode permet de représenter les sous-groupes d'une population hétérogène. Cette façon un peu plus complexe d'échantillonner garantit que chaque sous-groupe de la population est représenté d'une certaine manière dans l'échantillon. On subdivise la population en strates (sous-groupes relativement homogènes). Proportionnellement à son importance dans la population, on calcule combien il faut d'individus au sein de l'échantillon pour représenter chaque strate. On peut utiliser n'importe quelle des méthodes d'échantillonnage mentionnées ci-dessus pour sélectionner l'échantillon à l'intérieur de chaque strate. Pourquoi doit-on créer des strates ? Pour bien des raisons, la principale étant que leur création peut rendre la stratégie d'échantillonnage plus efficace. La stratification est des plus utiles lorsque les variables de stratification sont :

- simples à utiliser ;

- faciles à observer ;
- étroitement reliées au thème de l'enquête.

Procédure à suivre pour l'échantillonnage stratifié proportionnel à la taille des sous-groupes dans la populations

1. Définir clairement la nature de la population.
2. Déterminer les strates à représenter dans l'échantillon.
3. Assigner un numéro à chaque individu de chaque strate.
4. Déterminer le pourcentage que représente chaque strate dans la population.
5. Sélectionner l'échantillon en choisissant n'importe quelle méthode qui donne une chance égale à tous les numéros d'être tirés à l'intérieur d'une strate. Du coup il faut s'assurer que chaque strate est représentée proportionnellement à sa représentation dans la population.

2.2.4 Echantillonnage par grappes

Dans chacune des méthodes précédents, l'individu était choisi individuellement. L'échantillonnage par grappes consiste plutôt à choisir plusieurs individus en même temps. On choisit au hasard le nombre de grappes suffisant pour construire l'échantillon. On sélectionne tous les individus des grappes choisies.

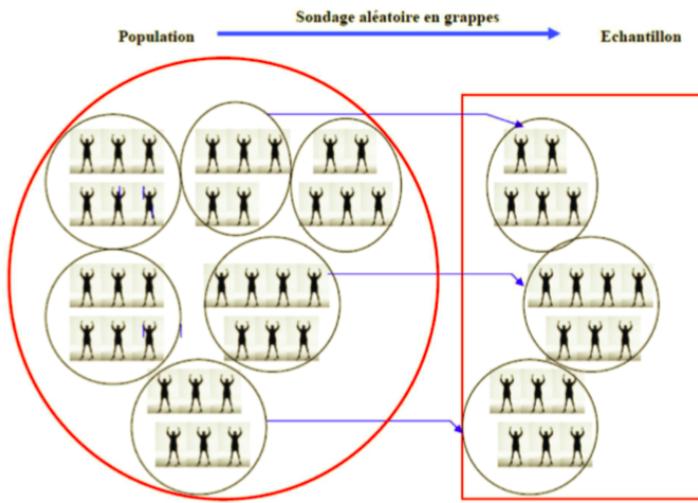


FIGURE 2.1 –

- **Les avantages**

- Echantillonnage aléatoire malgré l'absence de liste exhaustive
- Réduction des coûts par concentration.

- **Les inconvénients :**

- Les grappes : risque de ne pas représenter correctement la variabilité
- Les grappes utilisées doivent être de tailles à peu près équivalentes

Cette méthode permet d'obtenir un échantillon représentatif de la population si les grappes sont semblables et dans une grappe, les individus sont hétérogènes.

2.3 Modèles statistiques

Soit X une variable aléatoire réelle (discrète ou continue) dont la loi de probabilité \mathbb{P}_θ dépend d'un paramètre inconnu θ .

Définition 2.3.1. *On appelle modèle statistique la donnée d'une famille de lois de probabilité $\{\mathbb{P}_\theta, \theta \in \Theta \subset \mathbb{R}^d\}$; Θ est appelé espace des paramètres.*

Exemple.

Définition 2.3.2. *Un échantillon de X de taille n est un n -uplet (X_1, \dots, X_n) de variables aléatoires indépendantes de même loi que X .*

Exemple.

Remarque 2.3.1. *Attention ! Il ne faut pas confondre l'échantillon aléatoire (collection de variables aléatoires indiquées par une lettre majuscule) et la réalisation de cet échantillon (notée avec des lettres minuscules) :*

Echantillon : (X_1, \dots, X_n)

Réalisation : (x_1, \dots, x_n)

Définition 2.3.3. *On appelle statistique toute variable aléatoire ne dépendant que de l'échantillon (X_1, \dots, X_n) .*

Remarque 2.3.2. *Une statistique est un résumé de l'échantillon.*

La statistique inférentielle a pour objectif d'avoir des informations sur le paramètre inconnu θ en se basant sur l'échantillon (X_1, \dots, X_n) . On part de l'échantillon pour avoir une meilleure connaissance de la population.

Si X est une variable aléatoire réelle, alors on note :

- $f(x, \theta)$ si X est une variable aléatoire à densité
- $f(x, \theta) = \mathbb{P}_\theta(X = x)$ si X est une variable aléatoire discrète.

Définition 2.3.4. *Le support de \mathbb{P}_θ est l'ensemble $\{x : f(x, \theta) > 0\}$.*

Définition 2.3.5. *Si toutes les lois \mathbb{P}_θ , $\theta \in \Theta$ ont un support commun alors le modèle est dit homogène. Cela signifie que pour chaque $\theta \in \Theta$, $\{x : f(x, \theta) > 0\}$ ne dépend pas de θ .*

Définition 2.3.6. *Le modèle statistique $\{\mathbb{P}_\theta, \theta \in \Theta\}$ est identifiable lorsque l'application $\theta \mapsto \mathbb{P}_\theta$ est injective.*

On considère un échantillon (X_1, \dots, X_n) issu d'une loi de probabilité dépendant d'un paramètre inconnu $\theta \in \mathbb{R}$.

3.1 Vraisemblance

Définition 3.1.1. On appelle *vraisemblance* d'un échantillon (X_1, \dots, X_n) la fonction définie par

$$L(x_1, \dots, x_n, \cdot) : \Theta \rightarrow \mathbb{R}^+$$

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

Exemple.

3.2 Exhaustivité

Un échantillon nous apporte une certaine information sur le paramètre θ . Lorsque l'on résume cet échantillon par une statistique, il s'agit de ne pas perdre cette information. Une statistique qui conserve l'information contenue dans l'échantillon sera dite exhaustive.

Définition 3.2.1. *La statistique $T(X_1, \dots, X_n)$ est dite exhaustive pour θ si la loi conditionnelle de (X_1, \dots, X_n) sachant $T(X_1, \dots, X_n)$ ne dépend pas de θ .*

Le théorème ci-dessus appelé théorème de factorisation permet de trouver une statistique exhaustive ou de justifier qu'une statistique est exhaustive.

Théorème 3.2.1. *La statistique $T(X_1, \dots, X_n)$ est exhaustive pour θ si et seulement si la vraisemblance peut se factoriser sous la forme*

$$L(x_1, \dots, x_n, \theta) = g(T(x_1, \dots, x_n), \theta)h(x_1, \dots, x_n).$$

Exemple.

3.3 Information de Fisher

On considère un échantillon (X_1, \dots, X_n) issu d'une loi de probabilité \mathbb{P}_θ admettant une densité ou de fonction de masse $f(\cdot, \theta)$ avec $\theta \in \Theta \subset \mathbb{R}$. On note

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

la vraisemblance de l'échantillon. Pour mesurer l'information contenue dans un échantillon (X_1, \dots, X_n) , Ronald Aylmer Fisher (1890-1962) a défini la quantité ci-dessous.

Définition 3.3.1. *On appelle information de Fisher au point θ apportée par l'échantillon (X_1, \dots, X_n) la quantité*

$$I_n(\theta) = \mathbb{E}_\theta \left[\left(\frac{\partial \ln(L(X_1, \dots, X_n, \theta))}{\partial \theta} \right)^2 \right]$$

La proposition ci-dessus donne quelques propriétés de l'information de Fisher.

Proposition 3.3.1. *Nous avons :*

1. $I_n(\theta) \geq 0, \forall \theta \in \Theta$.
2. Si X et Y sont indépendantes de lois respectives \mathbb{P}_θ et \mathbb{Q}_θ . Notons $I_X(\theta)$, $I_Y(\theta)$ et $I_{(X,Y)}(\theta)$ les informations de Fisher au point θ respectivement apportées par X , Y , et (X, Y) . Alors, nous avons : Alors, nous avons :

$$I_{(X,Y)}(\theta) = I_X(\theta) + I_Y(\theta).$$

Comme conséquence, l'information de Fisher $I_n(\theta)$ au point θ fournie par l'échantillon (X_1, \dots, X_n) vérifie

$$I_n(\theta) = n I_{X_1}(\theta)$$

où $I_{X_1}(\theta)$ l'information de Fisher au point θ fournie par X_1 .

3. $T(X_1, \dots, X_n)$ est exhaustive $\iff I_n(\theta) = I_T(\theta) \quad \forall \theta \in \Theta$ où $I_T(\theta)$ est l'information de Fisher au point θ fournie par $T(X_1, \dots, X_n)$. Cette propriété permet donc d'établir l'exhaustivité d'une statistique.

Théorème 3.3.1. Si le support de X_1 ne dépend pas de θ et si la vraisemblance $\theta \mapsto L(x_1, \dots, x_n, \theta)$ est deux fois dérivable, alors

$$I_n(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2 \ln(L(X_1, \dots, X_n, \theta))}{\partial \theta^2} \right].$$

4.1 Qu'est ce qu'un estimateur ?

On considère un échantillon (X_1, \dots, X_n) issu d'une loi de probabilité \mathbb{P}_θ où $\theta \in \Theta \subset \mathbb{R}$ est inconnu. L'objectif est d'estimer θ en se basant sur l'échantillon (X_1, \dots, X_n) .

Définition 4.1.1. *Un estimateur $\hat{\theta}_n$ du paramètre θ est une statistique*

$$\hat{\theta}_n = T(X_1, \dots, X_n)$$

à valeurs dans un domaine acceptable pour θ .

- Si (x_1, \dots, x_n) est une observation de (X_1, \dots, X_n) , $T(x_1, \dots, x_n)$ est appelée estimation de θ .
- Il faut faire la distinction entre l'estimateur de θ (qui est une variable aléatoire réelle) et l'estimation de θ qui est une grandeur numérique.

Bien évidemment, cette statistique $T(X_1, \dots, X_n)$ n'est pas choisie au hasard ! L'idée est de trouver une statistique de sorte à fournir une bonne estimation du paramètre d'intérêt θ .

4.2 Qu'est ce qu'un "bon" estimateur ?

Quelles propriétés doit satisfaire un estimateur pour être considéré comme "bon" ? Nous devons distinguer deux cas suivant la taille d'échantillon n :

- propriétés à distance finie (pour n fixé)
- propriétés asymptotiques (pour $n \rightarrow +\infty$).

4.2.1 Propriétés à distance finie

4.2.1.1 Biais

Définition 4.2.1. *Le biais d'un estimateur $\hat{\theta}_n$ de θ est défini par*

$$b_n(\theta) = \mathbb{E}_\theta(\hat{\theta}_n) - \theta = \mathbb{E}_\theta(\hat{\theta}_n - \theta).$$

Le biais de l'estimateur est la moyenne des écarts systématiques entre $\hat{\theta}_n$ et θ . L'absence d'un écart systématique entre $\hat{\theta}_n$ et θ se traduit par un biais nul.

Définition 4.2.2. Un estimateur $\hat{\theta}_n$ de θ est dit sans biais lorsque pour tout $\theta \in \Theta$

$$\mathbb{E}_{\theta}(\hat{\theta}_n) = \theta.$$

Dans le cas contraire, l'estimateur $\hat{\theta}_n$ est dit biaisé.

Exemple.

4.2.1.2 Risque quadratique

On mesure la précision d'un estimateur par son risque quadratique.

Définition 4.2.3. Pour un estimateur $\hat{\theta}_n$ de θ , le risque quadratique est défini par

$$\begin{aligned} R(\hat{\theta}_n, \theta) &= \mathbb{E}_{\theta}(\hat{\theta}_n - \theta)^2 \\ &= \text{var}_{\theta}(\hat{\theta}_n) + (b_n(\theta))^2 \end{aligned}$$

Définition 4.2.4. Soient $\hat{\theta}_n$ et $\tilde{\theta}_n$ deux estimateurs de θ . On dit que $\hat{\theta}_n$ est préférable à $\tilde{\theta}_n$ si

$$R(\hat{\theta}_n, \theta) \leq R(\tilde{\theta}_n, \theta) \quad \forall \theta \in \Theta \iff R(\hat{\theta}_n, \theta) - R(\tilde{\theta}_n, \theta) \leq 0 \quad \theta \in \Theta.$$

Un estimateur optimal au sens du risque quadratique est l'estimateur qui a le plus petit risque quadratique pour toute valeur de $\theta \in \Theta$. Il est souvent difficile, voire impossible, de trouver un estimateur optimal.

Remarque 4.2.1. Pour un estimateur sans biais $\hat{\theta}_n$ de θ , le risque quadratique est défini par

$$R(\hat{\theta}_n, \theta) = \text{var}_\theta(\hat{\theta}_n)$$

Définition 4.2.5. Soient $\hat{\theta}_n$ et $\tilde{\theta}_n$ deux estimateurs sans biais de θ . On dit que $\hat{\theta}_n$ est préférable à $\tilde{\theta}_n$ si

$$\text{var}_\theta(\hat{\theta}_n) \leq \text{var}_\theta(\tilde{\theta}_n) \quad \forall \theta \in \Theta \iff \text{var}_\theta(\hat{\theta}_n) - \text{var}_\theta(\tilde{\theta}_n) \leq 0 \quad \theta \in \Theta.$$

Exemple.

4.2.1.3 Borne de Cramer-Rao

Le résultat suivant indique que le risque quadratique d'un estimateur sans biais (i.e. sa variance) ne peut être inférieure à une certaine borne qui dépend de l'information de Fisher.

Théorème 4.2.1. On suppose que l'information de Fisher sur θ apportée par (X_1, \dots, X_n) existe et est strictement positive pour tout θ . Soit $\hat{\theta}_n$ un estimateur sans biais de θ . Alors nous avons

$$\text{var}_\theta(\hat{\theta}_n) \geq \frac{1}{I_n(\theta)} \quad \forall \theta \in \Theta.$$

La borne $BCR(\theta) = \frac{1}{I_n(\theta)}$ est appelée borne de Cramer-Rao.

Remarque 4.2.2. Si $\hat{\theta}_n$ est un estimateur sans biais de $h(\theta)$ alors

$$\text{var}_\theta(\hat{\theta}_n) \geq \frac{(h'(\theta))^2}{I_n(\theta)}.$$

Dans ce cas, la borne de Cramer-Rao pour l'estimation sans biais de $h(\theta)$ est :

$$BCR(\theta) = \frac{(h'(\theta))^2}{I_n(\theta)}.$$

Définition 4.2.6. Un estimateur $\hat{\theta}_n$ de θ est dit efficace si

- $\hat{\theta}_n$ est sans biais
- $\text{var}_{\theta}(\hat{\theta}_n) = BCR(\theta)$.

Exemple.

4.2.2 Propriétés asymptotiques

4.2.2.1 Convergence ou consistance

Définition 4.2.7. Un estimateur $\hat{\theta}_n$ de θ est dit asymptotiquement sans biais lorsque pour tout θ ,

$$\mathbb{E}_{\theta}(\hat{\theta}_n) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta.$$

Définition 4.2.8. $\hat{\theta}_n$ est un estimateur convergent (ou consistant) de θ si

$$\hat{\theta}_n \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta \quad \text{lorsque } n \rightarrow +\infty$$

c'est à dire

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow +\infty} \mathbb{P}(|\hat{\theta}_n - \theta| \geq \varepsilon) = 0.$$

Interprétation : La convergence est une des propriétés les plus importantes pour un estimateur. On a la garantie qu'à un rang n assez grand et avec grande probabilité, $\hat{\theta}_n$ soit proche du paramètre θ .

Exemple.

4.2.2.2 Normalité asymptotique

Définition 4.2.9. Un estimateur $\hat{\theta}_n$ de θ est dit *asymptotiquement normal* si

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma_\theta^2) \quad n \rightarrow +\infty$$

où σ_θ^2 est à déterminer.

Interprétation : La normalité asymptotique est une propriété plus précise qui indique que la fluctuation de l'estimateur autour de θ est approximativement normale.

Exemple.

On considère un échantillon (X_1, \dots, X_n) issu d'une loi de probabilité \mathbb{P}_θ avec θ inconnu.

5.1 Méthode des moments

Principe de la méthode :

- Trouver des fonctions g et q telles que

$$\mathbb{E}(g(X_1)) = q(\theta). \quad (5.1.1)$$

Il faudrait choisir de préférence q bijective.

- Remplacer dans (5.1.1), la moyenne théorique par la moyenne empirique :

$$\frac{1}{n} \sum_{i=1}^n g(X_i) = q(\theta) \quad (5.1.2)$$

- Résoudre (5.1.2) ; si q est bijective alors l'estimateur par la méthode des moments est donné par :

$$\hat{\theta}_n = q^{-1} \left(\frac{1}{n} \sum_{i=1}^n g(X_i) \right).$$

Exemple.

5.2 Méthode du maximum de vraisemblance

La vraisemblance de l'échantillon (X_1, \dots, X_n) est donnée par

$$L_n(x_1, \dots, x_n, \theta) = \prod_{i=1}^n f(x_i, \theta).$$

Dans le cas d'une loi discrète

$$L_n(x_1, \dots, x_n, \theta) = \prod_{i=1}^n \mathbb{P}_\theta(X_i = x_i).$$

Pour un échantillon de taille 1

$$L_1(x, \theta) = \mathbb{P}_\theta(X_1 = x).$$

Principe de la méthode : Choisir comme estimateur la statistique $\hat{\theta}_n$, la valeur de θ qui maximise la vraisemblance $L_n(X_1, \dots, X_n, \theta)$:

Définition 5.2.1. $\hat{\theta}_n$ est un estimateur du maximum de vraisemblance de θ si

$$\forall \theta \in \Theta \quad L_n(X_1, \dots, X_n, \hat{\theta}_n) \geq L_n(X_1, \dots, X_n, \theta).$$

Log-vraisemblance.

Exemple.

5.3 Méthode des moindres carrés ordinaire et regression linéaire

Regression linéaire simple.

Regression linéaire multiple.

Estimation par intervalle de confiance

En estimation ponctuelle, on ne propose qu'une seule valeur pour le paramètre d'intérêt. Il n'y a quasiment aucune chance que cette valeur soit la vraie valeur. L'objectif de ce chapitre est de proposer une fourchette de valeurs possibles, tout un intervalle, ni trop gros, pour qu'il soit assez informatif, ni trop petit, pour qu'on soit raisonnablement sûr qu'il contienne la vraie valeur.

6.1 Introduction

Définition 6.1.1. Soit $\alpha \in]0, 1[$; on appelle *intervalle de confiance* pour le paramètre θ de niveau de confiance égale à $1 - \alpha$, un intervalle aléatoire $I(X_1, \dots, X_n) \subset \Theta$ tel que

$$\mathbb{P}_\theta(I(X_1, \dots, X_n) \ni \theta) = 1 - \alpha.$$

Définition 6.1.2. On dira que un intervalle aléatoire $I(X_1, \dots, X_n)$ est un *intervalle de confiance pour le paramètre θ de niveau de confiance asymptotique* égale à $1 - \alpha$ si

$$\lim_{n \rightarrow +\infty} \mathbb{P}_\theta(I(X_1, \dots, X_n) \ni \theta) = 1 - \alpha.$$

Lorsque

$$I(X_1, \dots, X_n) = [T_n^*(X_1, \dots, X_n), T_n^{**}(X_1, \dots, X_n)]$$

où $T_n^*(X_1, \dots, X_n)$ et $T_n^{**}(X_1, \dots, X_n)$ sont des statistiques à valeurs dans Θ , on parle d'*intervalle de confiance bilatéral*. Dans le cas où

$$I(X_1, \dots, X_n) = [T_n^*(X_1, \dots, X_n), +\infty[$$

ou

$$I(X_1, \dots, X_n) =]-\infty, T_n^*(X_1, \dots, X_n)],$$

on parle d'*intervalle de confiance unilatéral*.

Remarque 6.1.1. Dans l'univers des échantillons possibles, pour une proportion au moins $1 - \alpha$ d'entre eux, on obtient un intervalle qui contient θ .

Remarque 6.1.2. A α fixé, l'intervalle de confiance est d'autant meilleur que sa longueur est petite.

Remarque 6.1.3. On doit comprendre un intervalle de confiance de niveau $1 - \alpha$ comme un intervalle aléatoire qui a une probabilité $1 - \alpha$ de contenir le vrai paramètre θ .

Définition 6.1.3. Soit X une variable aléatoire réelle de fonction de répartition $F(x) = \mathbb{P}(X \leq x)$. Pour $\alpha \in]0, 1[$, on appelle quantile (ou fractile) d'ordre α de la loi de X le nombre

$$q_\alpha = \inf\{x \in \mathbb{R}, F(x) \geq \alpha\}.$$

Lorsque la fonction de répartition F est continue et strictement croissante, elle est inversible d'inverse F^{-1} et pour tout $\alpha \in]0, 1[$, on a $q_\alpha = F^{-1}(\alpha)$.

6.2 Construction d'un intervalle de confiance

1. Construction de la fonction pivot (ou pivotale)
2. Détermination des constantes
3. Pivotement

6.2.1 Fonction pivotale

Définition 6.2.1. On appelle fonction pivotale pour θ toute fonction de l'échantillon et de θ , $\phi(X_1, \dots, X_n, \theta)$ dont la loi ne dépend pas de θ .

Définition 6.2.2. Une fonction asymptotiquement pivotale pour θ est une variable aléatoire, $\phi(X_1, \dots, X_n, \theta)$ qui converge en loi vers une variable aléatoire dont la loi ne dépend pas de θ .

6.2.2 Construction d'un intervalle de confiance bilateral

6.2.2.1 Méthode non asymptotique

1. Soit $\phi(X_1, \dots, X_n, \theta)$ une fonction pivotale pour θ .
2. Pour un seuil $\alpha \in]0, 1[$ fixé, soient q_1 et q_2 tels que

$$\mathbb{P}_\theta [q_1 \leq \phi(X_1, \dots, X_n, \theta) \leq q_2] = 1 - \alpha$$

c'est à dire

$$\begin{aligned}\mathbb{P}_\theta [\phi(X_1, \dots, X_n, \theta) \leq q_1] &= \alpha_1 \\ \mathbb{P}_\theta [\phi(X_1, \dots, X_n, \theta) \geq q_2] &= \alpha_2\end{aligned}$$

avec $\alpha_1 + \alpha_2 = \alpha$.

3. La double inéquation

$$q_1 \leq \phi(X_1, \dots, X_n, \theta) \leq q_2 \tag{6.2.1}$$

peut se résoudre (ou "pivoter") en θ selon

$$T_1(X_1, \dots, X_n) \leq \theta \leq T_2(X_1, \dots, X_n),$$

on en déduit immédiatement un intervalle de confiance bilatéral pour θ de niveau de confiance $1 - \alpha$.

6.2.2.2 Méthode asymptotique

- Soit T_n un estimateur de θ tel que

$$\frac{T_n - \theta}{s_n(\theta)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

où $s_n(\theta)$ est une fonction continue de θ .

- Si la fonction $\frac{T_n - \theta}{s_n(\theta)}$ pivote pour isoler θ , on obtient l'intervalle de confiance approchée.
- Sinon T_n étant convergent, moyennant la continuité de s_n (quelque soit n), on obtient

$$\frac{T_n - \theta}{s_n(T_n)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

Le pivotement est alors immédiat.

Remarque 6.2.1. Pour les intervalles de confiance unilatéraux, on utilise la méthode ci-dessus.

6.2.3 Densité de probabilité unimodale

Définition 6.2.3. Une densité de probabilité f sur \mathbb{R} est unimodale autour d'un mode si il existe x^* un mode tel que f croissante sur $]-\infty, x^*]$ et f décroissante sur $[x^*, +\infty[$.

Proposition 6.2.1. Soit f une densité unimodale et $[a, b]$ un intervalle satisfaisant

$$i) \int_a^b f(x)dx = 1 - \alpha$$

$$ii) f(a) = f(b) > 0$$

$$iii) a \leq x^* \leq b \text{ où } x^* \text{ est le mode de } f.$$

Alors $[a, b]$ est l'intervalle le plus court parmi tous les intervalles satisfaisant i).

6.3 Exemples

6.3.1 Intervalle de confiance pour la moyenne d'une loi normale

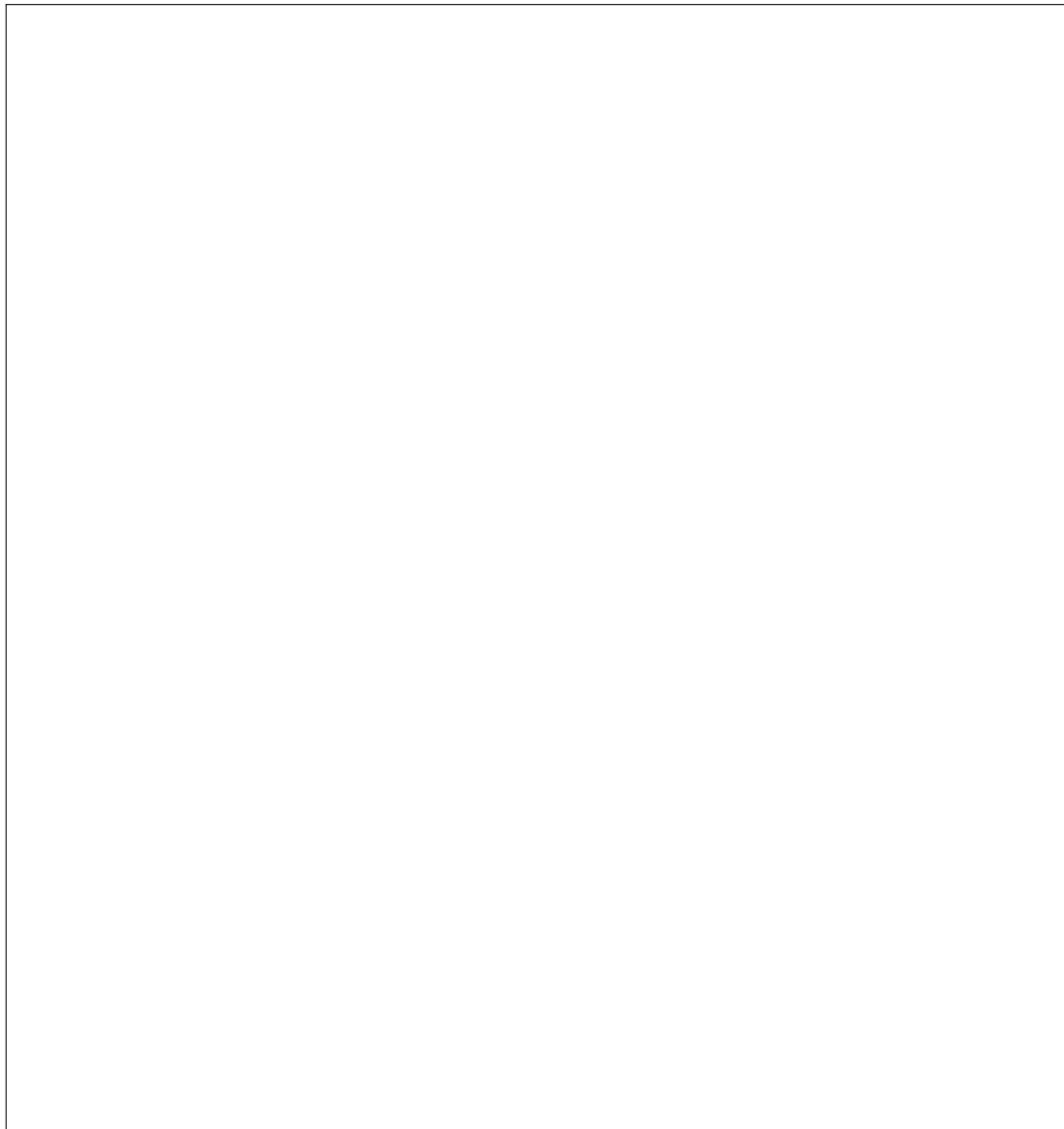
Considérons un échantillon (X_1, \dots, X_n) issu d'une loi normale $\mathcal{N}(\mu, \sigma^2)$ avec $\theta = (\mu, \sigma^2)$.

Si $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ alors

1. **σ^2 connue et estimation de μ .** Nous savons que \bar{X}_n est un estimateur efficace de μ .
De plus

$$\bar{X}_n \hookrightarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \hookrightarrow \mathcal{N}(0, 1).$$

Par suite $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ est une fonction pivot. Ainsi, nous obtenons



Remarque 6.3.1. On appelle marge d'erreur la quantité

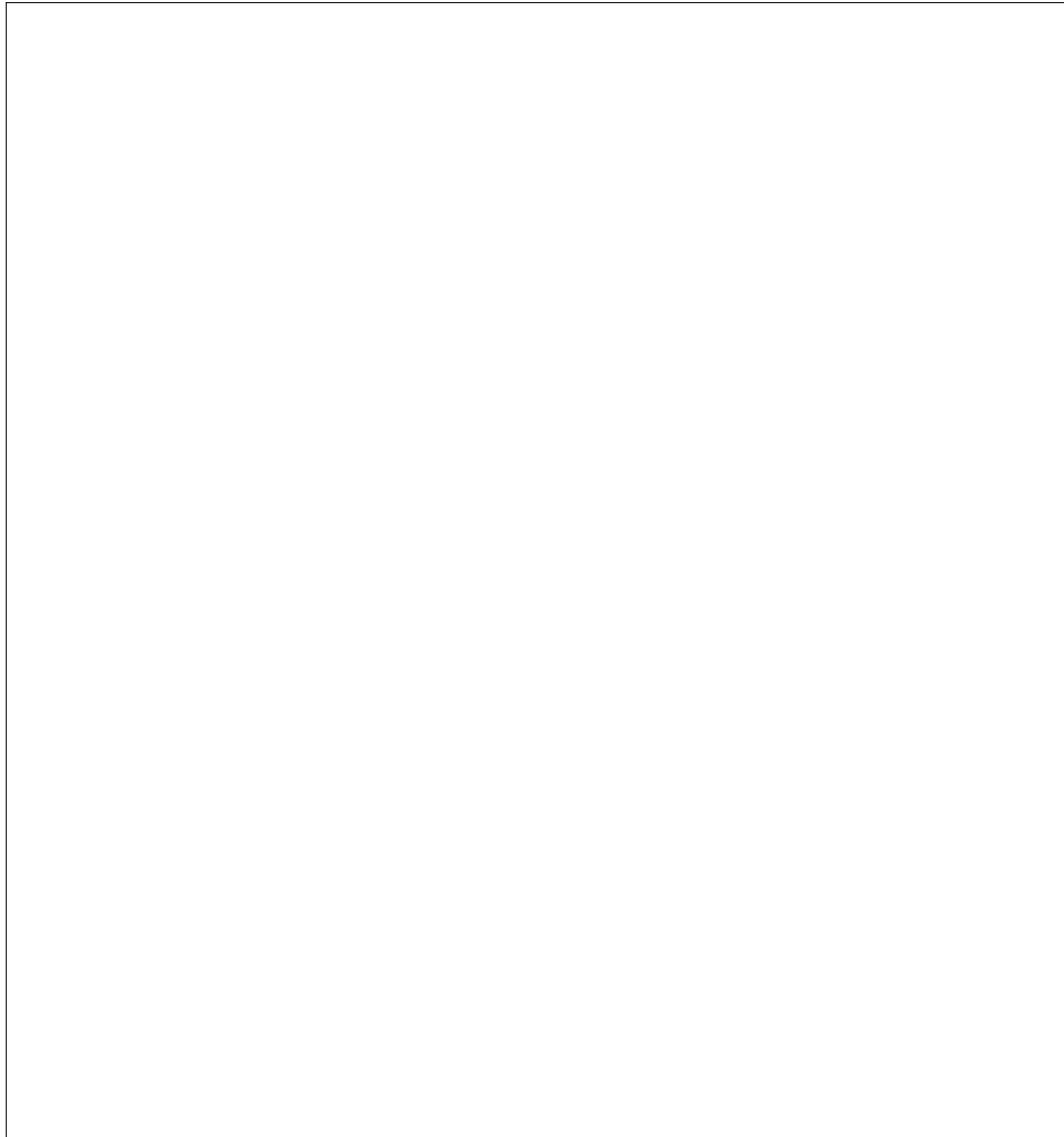
$$ME = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}.$$

Marge d'erreur et taille d'échantillon

2. σ^2 inconnue et estimation de μ . Nous avons le résultat suivant

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S} \hookrightarrow T(n-1) \quad \text{avec} \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

Cette variable aléatoire est une fonction pivotale pour μ .



Marge d'erreur et taille d'échantillon

6.3.2 Intervalle de confiance pour la variance d'une loi normale

1. **μ connue et estimation de σ^2 .** Nous savons que $V^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ est un bon estimateur de σ^2 . On déduit alors que

$$\frac{nV^2}{\sigma^2} \hookrightarrow \chi^2(n).$$

Ainsi, nous avons

$$\mathbb{P}\left(a \leq \frac{nV^2}{\sigma^2} \leq b\right) = 1 - \alpha$$

$$\mathbb{P}\left(\frac{nV^2}{\sigma^2} < a\right) + \mathbb{P}\left(\frac{nV^2}{\sigma^2} > b\right) = \alpha.$$

Ainsi $a = \chi_{\alpha_2}^{(n)}$ et $b = \chi_{1-\alpha_1}^{(n)}$ avec $\alpha_1 + \alpha_2 = \alpha$. On déduit que

2. **μ inconnue et estimation de σ^2 .** Nous avons

$$\frac{(n-1)S^2}{\sigma^2} \hookrightarrow \chi^2(n-1).$$

Ainsi, nous avons

$$\begin{aligned} \mathbb{P}\left(q_1 \leq \frac{(n-1)S^2}{\sigma^2} \leq q_2\right) &= 1 - \alpha \\ \mathbb{P}\left[\frac{(n-1)S^2}{\sigma^2} < q_1\right] + \mathbb{P}\left[\frac{(n-1)S^2}{\sigma^2} > q_2\right] &= \alpha. \end{aligned}$$

Ainsi $q_1 = \chi_{\alpha_2}^{(n-1)}$ et $q_2 = \chi_{1-\alpha_1}^{(n-1)}$ avec $\alpha_1 + \alpha_2 = \alpha$. On déduit que

6.3.3 Intervalle de confiance pour une proportion

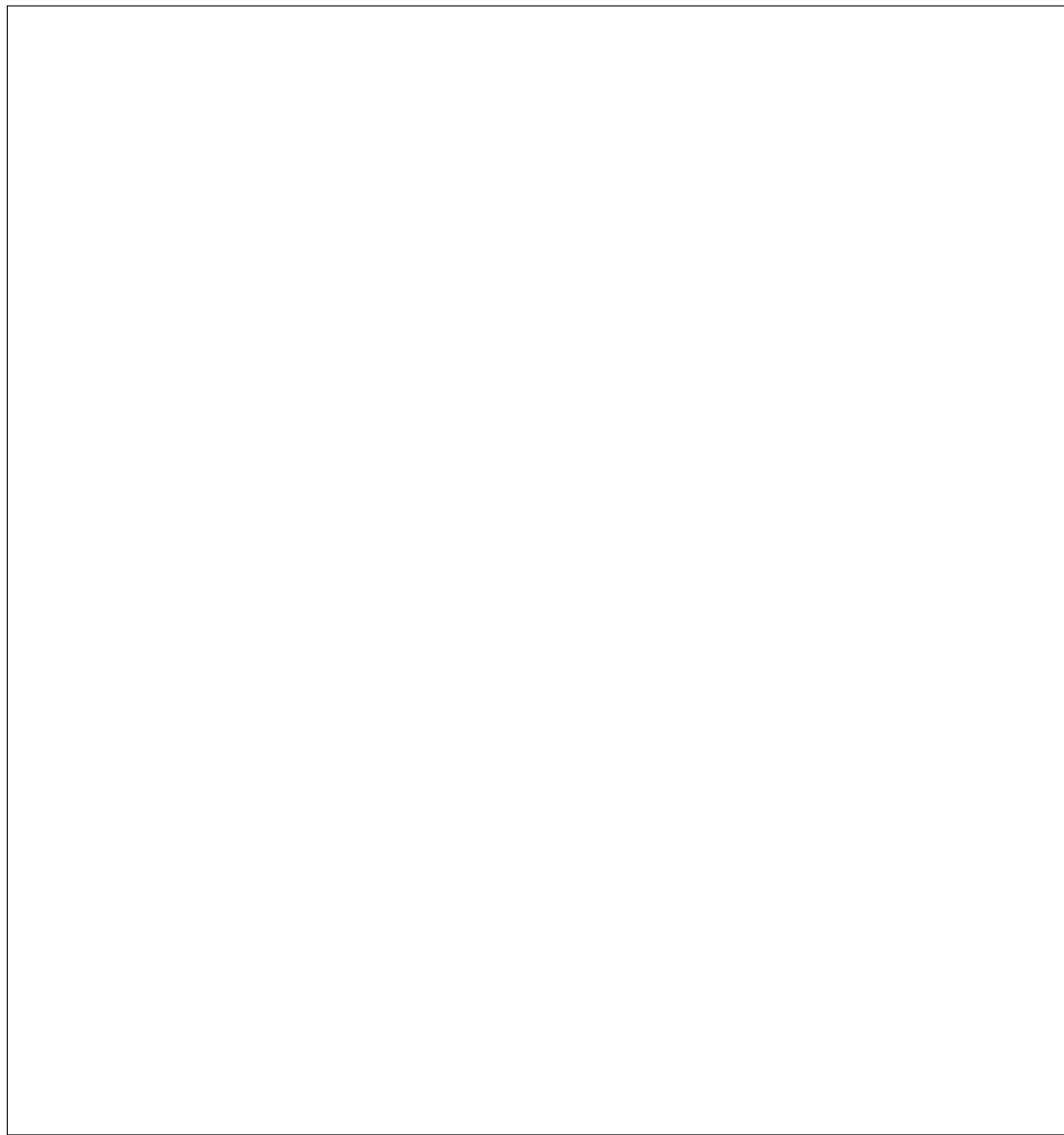
On considère un échantillon (X_1, \dots, X_n) issu de la loi de Bernouilli $\mathcal{B}(1, p)$, $p \in]0, 1[$. D'après le Théorème Central limite, nous avons :

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

On remplace alors le numérateur $\sqrt{p(1-p)}$ et $\sqrt{\bar{X}_n(1-\bar{X}_n)}$ et on obtient toujours

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

Pour n assez grand,

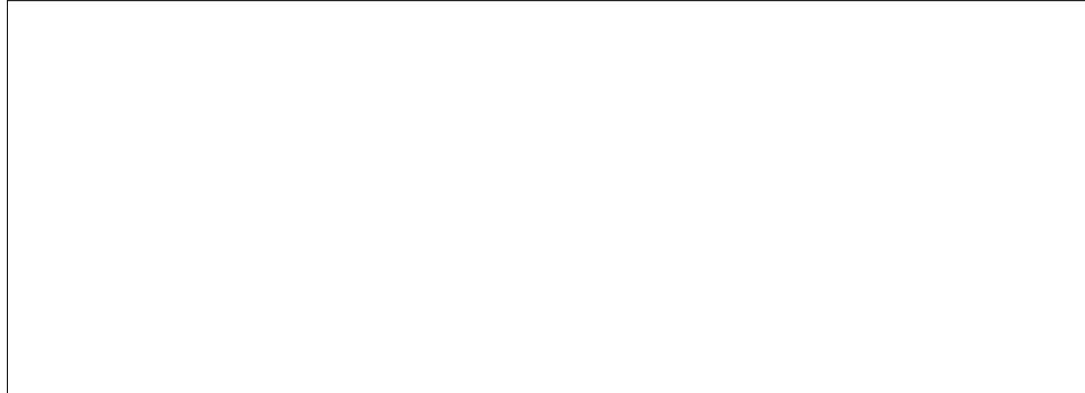


Marge d'erreur et taille d'échantillon

6.3.4 Intervalle de confiance pour la moyenne d'une loi quelconque

On considère un échantillon (X_1, \dots, X_n) issu d'une loi de probabilité admettant une moyenne m et une variance σ^2 . D'après le Théorème central limite, nous avons le résultat suivant :

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



7.1 Principe des tests

On considère un échantillon (X_1, \dots, X_n) issu d'une loi \mathbb{P}_θ avec $\theta \in \Theta$. Soient Θ_0 et Θ_1 deux sous-ensembles de Θ tels que $\Theta = \Theta_0 \cup \Theta_1$ et $\Theta_0 \cap \Theta_1 = \emptyset$. Soient les hypothèses :

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_1$$

L'hypothèse H_0 est appelée hypothèse nulle et H_1 , hypothèse alternative. Une hypothèse est dite simple si elle est réduite à un singléton. Les deux hypothèses sont telles que une et une seule est vraie.

Un test statistique est un mécanisme qui permet de trancher entre deux hypothèses à partir des résultats d'un échantillon. La décision consiste à choisir H_0 ou H_1 . Il y a quatre cas qui sont reproduits dans le tableau ci-dessous

	H_0 vraie	H_1 vraie
H_0 décidée	Bonne décision	Erreur de deuxième espèce
H_1 décidée	Erreur de première espèce	Bonne décision

Exemple 7.1.1. Contrôle de qualité. Une machine produit des pièces classées soit "bonnes" codées par 0, soit "défectueuses" codées par 1. Le nombre de pièces fabriquées étant gigantesque et l'examen de chaque pièce étant relativement coûteux, on ne peut évaluer la qualité de sa production que sur un lot de taille n faible au regard de la production. On observe alors ce lot de n pièces et on note (x_1, \dots, x_n) les observations.

Modélisation : on suppose que x_i est la réalisation d'une variable aléatoire X_i de loi de Bernouilli $\mathcal{B}(1, p)$, $p \in]0, 1[$; nous faisons les hypothèses suivantes :

- X_1, \dots, X_n sont indépendantes : on admet que des petites variations aléatoires pouvant influer sur la qualité des pièces ne se repercutent pas d'une pièce à une autre.
- X_1, \dots, X_n sont identiquement distribuées : on admet que la production a été stable durant la période d'observation ; cette stabilité est caractérisée par la constance de la probabilité p pour chaque pièce produite d'être défectueuse.

Nous considérons le problème de test de H_0 : la machine est aux normes contre H_1 : la machine n'est pas aux normes.

- Erreur de première espèce : décider que la machine n'est pas aux normes alors qu'en réalité elle est aux normes : dépenses inutiles de réparation ou de changement de matériels.

- Erreur de deuxième espèce : décider que la machine est aux normes alors qu'en réalité elle n'est pas aux normes : production de mauvaises pièces pouvant aboutir à un mécontentement de la clientèle, voire à des problèmes de sécurité.

Définition 7.1.1. On appelle test une statistique $\psi(X_1, \dots, X_n)$ à valeurs dans $\{0, 1\}$ telle que

$$\begin{aligned}\psi(X_1, \dots, X_n) = 0 &\implies \text{on accepte } H_0 \\ \psi(X_1, \dots, X_n) = 1 &\implies \text{on accepte } H_1.\end{aligned}$$

Définition 7.1.2. On appelle région critique la région d'acceptation de l'hypothèse alternative H_1 :

$$W = \{(X_1, \dots, X_n) : \psi(X_1, \dots, X_n) = 1\}.$$

Un test est caractérisé par sa région critique.

Définition 7.1.3. On appelle risque de première espèce du test $\psi(X_1, \dots, X_n)$ la probabilité de l'erreur de première espèce :

$$\begin{aligned}\alpha_\psi : \Theta_0 &\longrightarrow [0, 1] \\ \theta &\longmapsto \mathbb{P}_\theta(W).\end{aligned}$$

Définition 7.1.4. On appelle niveau du test $\psi(X_1, \dots, X_n)$ la quantité

$$\sup_{\theta \in \Theta} \alpha_\psi(\theta).$$

Le test $\psi(X_1, \dots, X_n)$ est dit de niveau $\alpha \in (0, 1)$ si

$$\sup_{\theta \in \Theta} \alpha_\psi(\theta) = \alpha.$$

Remarque 7.1.1. Le niveau du test est le plus gros risque de première espèce possible.

Définition 7.1.5. On appelle risque de deuxième espèce du test $\psi(X_1, \dots, X_n)$ la probabilité de l'erreur de deuxième espèce :

$$\begin{aligned}\beta_\psi : \Theta_1 &\longrightarrow [0, 1] \\ \theta &\longmapsto \mathbb{P}_\theta(\overline{W}).\end{aligned}$$

L'idéal serait de diminuer les deux risques d'erreur en même temps. Malheureusement, on montre qu'ils varient en sens inverse. Dans la pratique des tests statistiques, il est de règle de se fixer α , ce qui fait jouer à H_0 un rôle prééminent.

Un test est déterminé par sa région critique W . La région critique dépend du niveau α et d'une statistique appelée variable de décision. Pour la déterminer, il est indispensable de connaître la loi de la variable de décision sous l'hypothèse H_0 . Lorsque (x_1, \dots, x_n) sont des valeurs observées de cet échantillon,

- si $(x_1, \dots, x_n) \in W$, alors on rejette H_0 et on accepte H_1 ;
- si $(x_1, \dots, x_n) \notin W$, alors on accepte H_0 et on rejette H_1 .

Définition 7.1.6. On appelle puissance du test $\psi(X_1, \dots, X_n)$ la probabilité d'accepter H_1 quand H_1 est vraie :

$$\begin{aligned}\gamma_\psi : \Theta_1 &\longrightarrow [0, 1] \\ \theta &\longmapsto \mathbb{P}_\theta(W).\end{aligned}$$

La puissance

- croît avec le niveau de signification α .
- croît avec la taille de l'échantillon
- dépend de la région critique.

Remarque 7.1.2. Nous avons $\forall \theta \in \Theta_1, \gamma_\psi(\theta) = 1 - \beta_\psi(\theta)$.

Remarque 7.1.3. Un bon test est un test qui, pour un niveau α donné, maximise la puissance.

Définition 7.1.7. Un test $\psi(X_1, \dots, X_n)$ est sans biais lorsque la puissance du test est supérieure au niveau α sur Θ_1 :

$$\gamma(\theta) \geq \alpha \quad \forall \theta \in \Theta_1.$$

7.2 Etapes des tests

1. Etape préliminaire : modélisation du problème.
2. Formulation des hypothèses H_0 et H_1 .
3. Choix du seuil du test α .
4. Choix d'une statistique de test T_n , dont on connaît la loi sous H_0
5. Etude du comportement de T_n sous H_1 et déduction de la forme de la zone critique.
6. Calcul de cette zone pour le niveau α fixé puis confrontation aux données; et / ou calcul de la p-valeur du test sur les données
7. Conclusion statistique : conservation ou rejet de l'hypothèse de départ H_0 et commentaire éventuel sur la p-valeur.
8. Conclusion stratégique : décision que l'on va prendre une fois éclairé par le résultat statistique.

7.3 Construction d'un test d'hypothèses

Pour construire un test d'hypothèses portant sur la valeur d'un paramètre θ , l'on peut se fier au bon sens. Si on connaît un estimateur $\hat{\theta}_n$ de θ , on pourrait procéder de la façon suivante : soit θ_0 une valeur possible de θ .

- Test de $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$.
On rejette H_0 si $\hat{\theta}_n$ est "trop grand" i.e. la région critique est

$$W = \{\hat{\theta}_n - \theta_0 > l_\alpha\}.$$

- Test de $H_0 : \theta \geq \theta_0$ contre $H_1 : \theta < \theta_0$.
On rejette H_0 si $\hat{\theta}_n$ est "trop petit" i.e. la région critique est

$$W = \{\hat{\theta}_n - \theta_0 < l_\alpha\}.$$

- Test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta \neq \theta_0$.
On rejette H_0 si $|\hat{\theta}_n - \theta_0|$ est "trop grand" i.e. la région critique est

$$W = \{|\hat{\theta}_n - \theta_0| > l_\alpha\}.$$

- Test de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$.
 - $W = \{\hat{\theta}_n > l_\alpha\}$ si $\theta_1 > \theta_0$
 - $W = \{\hat{\theta}_n < l_\alpha\}$ si $\theta_1 < \theta_0$.

Pour déterminer l_α , il faut résoudre l'équation $\mathbb{P}_{\theta_0}(W) = \alpha$.

7.4 La *p*-value

En pratique, plutôt que de calculer la région critique en fonction de α , on préfère donner un seuil critique de α^* appelée *p*-value, qui est telle que

- si $\alpha^* < \alpha$, on rejette H_0
- si $\alpha < \alpha^*$, on accepte H_0 .

Les logiciels statistiques calculent et présentent les *p*-valeurs qui sont difficiles à obtenir sans moyen de calcul approprié.

8.1 Introduction

On appelle test de Student un test de comparaison de la moyenne dans un échantillon gaussien, c'est à dire un échantillon (X_1, \dots, X_n) issu de la loi normale $\mathcal{N}(m, \sigma^2)$. Soit m_0 une valeur possible de m . La moyenne empirique \bar{X}_n est un estimateur efficace de m . Deux résultats importants :

$$\bar{X}_n \hookrightarrow \mathcal{N}\left(m, \frac{\sigma^2}{n}\right) \iff \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \hookrightarrow \mathcal{N}(0, 1).$$

$$\frac{\sqrt{n}(\bar{X}_n - m)}{S_n} \hookrightarrow \mathcal{T}(n-1)$$

qui est la loi de Student à $n-1$ dégrés de liberté avec

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

8.2 $H_0 : m \leq m_0$ contre $H_1 : m > m_0$

8.2.1 On suppose que la variance σ^2 est connue.

En se référant à la Section 8.3, nous obtenons une première forme de la région critique

$$W = \left\{ \bar{X}_n - m_0 > l_\alpha \right\},$$

où la constante l_α est déterminée par (le test étant de niveau α)

$$\mathbb{P}_{m_0}(\bar{X}_n - m_0 > l_\alpha).$$

Sous l'hypothèse H_0 ,

$$\bar{X}_n \hookrightarrow \mathcal{N}\left(m_0, \frac{\sigma^2}{n}\right) \iff \frac{\sqrt{n}(\bar{X}_n - m_0)}{\sigma} \hookrightarrow \mathcal{N}(0, 1).$$

Ce qui implique alors

$$\mathbb{P}_{m_0} \left(\frac{\sqrt{n}(\bar{X}_n - m_0)}{\sigma} > \frac{\sqrt{n}l_\alpha}{\sigma} \right) = \alpha.$$

Ainsi, on en déduit que

$$\frac{\sqrt{n}l_\alpha}{\sigma} = q_{1-\alpha} \Leftrightarrow l_\alpha = \frac{\sigma}{\sqrt{n}}q_{1-\alpha}$$

où $q_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de $\mathcal{N}(0,1)$.

La région critique au niveau α du test $H_0 : m \leq m_0$ contre $H_1 : m > m_0$ lorsque σ^2 est connue est

$$\begin{aligned} W &= \left\{ \bar{X}_n - m_0 > \frac{\sigma}{\sqrt{n}}q_{1-\alpha} \right\} \\ &= \left\{ \frac{\sqrt{n}(\bar{X}_n - m_0)}{\sigma} > q_{1-\alpha} \right\} \end{aligned} \quad (8.2.1)$$

où $q_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de la loi normale centrée-réduite.

Remarque 8.2.1. On accepte H_1 au niveau α lorsque la différence $\bar{X}_n - m_0$ est significative, c'est à dire strictement supérieure à $\frac{\sigma}{\sqrt{n}}q_{1-\alpha}$.

Exercice 8.2.1. Une marque de tablettes de chocolat annonce que ses tablettes contiennent une teneur en cacao supérieure à 430 g par kg. On effectue un contrôle de qualité sur un échantillon de 10 tablettes et on obtient les teneurs suivantes en g/kg : 505.1 423.5 462.0 391.9 412.1 487.2 439.0 434.1 441.1 474.2. On admet que chaque mesure suit une loi normale $\mathcal{N}(m, \sigma^2)$. On admet dans un premier temps (au vu de contrôles antérieurs) que $\sigma = 24$. Que peut-on conclure au niveau $\alpha = 0.05$?

Solution.

8.2.2 On suppose σ^2 est inconnue

Nous allons remplacer dans (8.2.1), σ par l'écart-type empirique modifié S_n .

La région critique au niveau α du test $H_0 : m \leq m_0$ contre $H_1 : m > m_0$ lorsque σ^2 est inconnue est

$$W = \left\{ \frac{\sqrt{n}(\bar{X}_n - m_0)}{S_n} > t_{1-\alpha, n-1} \right\}$$

où $t_{1-\alpha, n-1}$ est le quantile d'ordre $1-\alpha$ de la loi de Student à $n-1$ degrés de liberté $\mathcal{T}(n-1)$.

Exercice 8.2.2. Une marque de tablettes de chocolat annonce que ses tablettes contiennent

une teneur en cacao supérieure à 430 g par kg. On effectue un contrôle de qualité sur un échantillon de 10 tablettes et on obtient les teneurs suivantes en g/kg : 505.1 423.5 462.0 391.9 412.1 487.2 439.0 434.1 441.1 474.2. On admet que chaque mesure suit une loi normale $\mathcal{N}(m, \sigma^2)$. Que peut-on conclure au niveau $\alpha = 0.05$?

Solution.

8.3 $H_0 : m \geq m_0$ contre $H_1 : m < m_0$

8.3.1 On suppose que la variance σ^2 est connue.

La région critique au niveau α du test $H_0 : m \geq m_0$ contre $H_1 : m < m_0$ lorsque σ^2 est connue est

$$\begin{aligned} W &= \left\{ \bar{X}_n < m_0 + \frac{\sigma}{\sqrt{n}} q_\alpha \right\} \\ &= \left\{ \frac{\sqrt{n}(\bar{X}_n - m_0)}{\sigma} < q_\alpha \right\} \end{aligned} \quad (8.3.1)$$

où q_α est le quantile d'ordre α de la loi normale centrée-réduite.

Exercice 8.3.1. Le département de contrôle de la qualité d'une entreprise détermine que le poids moyen net d'une boîte de céréales ne devrait pas être inférieur à 200 g. L'expérience a montré que les poids sont approximativement distribués normalement avec un écart-type de 15 g. Un échantillon de 15 boîtes prélevé aléatoirement sur la ligne de production donne un poids moyen de 195 g. Cela est-il suffisant pour pouvoir affirmer que le poids moyen des boîtes est inférieur à 200 g ?

Solution.

8.3.2 On suppose que la variance σ^2 est inconnue.

La région critique au niveau α du test $H_0 : m \geq m_0$ contre $H_1 : m < m_0$ lorsque σ^2 est inconnue est

$$W = \left\{ \frac{\sqrt{n}(\bar{X}_n - m_0)}{S_n} < t_{\alpha, n-1} \right\} \quad (8.3.2)$$

où $t_{\alpha, n-1}$ est le quantile d'ordre α de la loi de Student à $n-1$ degrés de liberté $\mathcal{T}(n-1)$.

Exercice 8.3.2. Le département de contrôle de la qualité d'une entreprise détermine que le poids moyen net d'une boîte de céréales ne devrait pas être inférieur à 200 g. L'expérience a

montré que les poids sont approximativement distribués normalement. Un échantillon de 15 boîtes prélevé aléatoirement sur la ligne de production donne un poids moyen de 195 g avec un écart-type estimé égal à 15 kg. Cela est-il suffisant pour pouvoir affirmer que le poids moyen des boîtes est inférieur à 200 g ?

Solution.

8.4 $H_0 : m = m_0$ contre $H_1 : m \neq m_0$

La région critique au niveau α du test $H_0 : m = m_0$ contre $H_1 : m \neq m_0$ lorsque σ^2 est connue est

$$W = \left\{ \left| \frac{\sqrt{n}(\bar{X}_n - m_0)}{\sigma} \right| > q_{1-\frac{\alpha}{2}} \right\} \quad (8.4.1)$$

où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale centrée-réduite.

Exercice 8.4.1. Une entreprise de vente par correspondance demande un montant fixe pour les frais d'envoi, indépendamment du poids du colis. Une étude réalisée il y a quelques années a montré que le poids moyen d'un colis était de 17.5 kg avec un écart-type de 3.6 kg. La comptabilité soupçonne que le poids moyen est maintenant différent de 17.5 kg. Un échantillon aléatoire de 100 colis est prélevé et fournit un poids moyen de $\bar{x} = 18.4$ kg. On suppose que les poids des colis sont distribués normalement. Que conclure au niveau $\alpha = 0.05$

Solution.

8.4.1 On suppose que la variance σ^2 est inconnue.

La région critique au niveau α du test $H_0 : m = m_0$ contre $H_1 : m \neq m_0$ lorsque σ^2 est inconnue est

$$W = \left\{ \left| \frac{\sqrt{n}(\bar{X}_n - m_0)}{S_n} \right| > t_{1-\frac{\alpha}{2}, n-1} \right\} \quad (8.4.2)$$

où $t_{1-\frac{\alpha}{2}, n-1}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student à $n-1$ degrés de liberté $\mathcal{T}(n-1)$.

Exercice 8.4.2. Une entreprise de vente par correspondance demande un montant fixe pour les frais d'envoi, indépendamment du poids du colis. Une étude réalisée il y a quelques années

a montré que le poids moyen d'un colis était de 17.5 kg. La comptabilité soupçonne que le poids moyen est maintenant différent de 17.5 kg. Un échantillon aléatoire de 100 colis est prélevé et fournit un poids moyen de $\bar{x} = 18.4$ kg avec un écart-type estimé égal à 3.6. On suppose que les poids des colis sont distribués normalement. Que conclure au niveau $\alpha = 0.05$

Solution.

9.1 Introduction

Soient P_1 et P_2 deux populations. On étudie un caractère (rendement, chiffre d'affaire, seuil de perception, etc.) sur ces deux populations. Le caractère a pour espérance m_1 et pour variance σ_1^2 dans la population P_1 et a pour espérance m_2 et pour variance σ_2^2 dans la population P_2 . Pour des raisons techniques, on supposera que le caractère est distribué selon une loi normale. On dispose alors de deux échantillons (X_1, \dots, X_{n_1}) et (Y_1, \dots, Y_{n_2}) issus respectivement de P_1 et P_2 , tels que X_i et Y_j sont indépendantes :

- (X_1, \dots, X_{n_1}) est issu de $\mathcal{N}(m_1, \sigma_1^2)$
- (Y_1, \dots, Y_{n_2}) est issu de $\mathcal{N}(m_2, \sigma_2^2)$.

Dans cette section, on comparera les moyennes et les variances des deux échantillons. Les moyennes empiriques, variances empiriques modifiées des deux échantillons sont notées respectivement \bar{X}_{n_1} , S_1^2 , \bar{Y}_{n_2} et S_2^2 .

Exemple 9.1.1. Deux groupes d'étudiants de tailles respectives $n_1 = 25$ et $n_2 = 31$ ont suivi le même cours de statistique et passe le même examen. Les moyennes et écarts-types empiriques des notes obtenues dans les deux groupes sont respectivement :

	moyenne	Variance S^2
Groupe 1	12.8	3.4
Groupe 2	11.3	2.9

On suppose que les notes sont reparties dans les deux groupes selon des lois normales et qu'elles sont toutes indépendantes. Peut-on considérer que le premier groupe est meilleur que le deuxième, c'est-à-dire qu'un point et demi d'écart entre les moyennes est significatif d'une différence de niveau ? La procédure à suivre consiste à tester d'abord l'égalité des variances, puis l'égalité des moyennes.

Exemple 9.1.2. Deux variétés de blé ont été cultivées chacune sur 8 parcelles ($n_1 = n_2 = 8$). Les rendements observés (en quintaux/hectare) sont regroupés dans le tableau ci-dessus :

	moyenne	variance σ^2
Echantillon 1	80.0	1.00
Echantillon 2	81.5	1.00

Si l'on considère que les 16 parcelles, la variété 2 présente en moyenne un rendement supérieur (de 1.5q/ha) à celui de la variété 1. Peut-on généraliser ce résultat ? Autrement dit, la différence observée (de 1.5q/ha) doit être considérée comme une conséquence d'un rendement moyen différent selon la variété ou, au contraire, est-il fortuit ? Selon un autre point de vue, la question peut être posée ainsi : la différence de moyenne observée doit être imputée au hasard (c'est-à-dire à la variété "naturelle" dite aussi "résiduelle" pour exprimer que l'on ne sait l'expliquer par la statistique) ?

9.2 Test de Fisher de comparaison des variances

Comparer les variances des deux échantillons revient à résoudre par exemple le problème de test suivant : $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Au niveau $\alpha \in]0, 1[$, la région critique du test $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$ est

$$W = \left\{ \frac{S_1^2}{S_2^2} < f_{\frac{\alpha}{2}}^* \right\} \cup \left\{ \frac{S_1^2}{S_2^2} > f_{1-\frac{\alpha}{2}}^* \right\}$$

où $f_{\frac{\alpha}{2}}^*$ est le quantile d'ordre $\frac{\alpha}{2}$ de la loi de Fisher à $(n_1 - 1, n_2 - 1)$ degrés de liberté, $f_{1-\frac{\alpha}{2}}^*$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Fisher à $(n_1 - 1, n_2 - 1)$ degrés de liberté et

$$\begin{aligned} S_{n_1} &= \left(\frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X}_{n_1})^2 \right)^{1/2} \\ S_{n_2} &= \left(\frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y}_{n_2})^2 \right)^{1/2}. \end{aligned}$$

9.3 Test de Student de comparaison des moyennes

On désire maintenant comparer les moyennes. Le test d'égalité des moyennes est :

$$H_0 : m_1 = m_2 \text{ contre } H_1 : m_1 \neq m_2.$$

Lorsque H_0 est vraie, on observe très rarement une parfaite égalité des moyennes. La question est donc de savoir à partir de quel écart de moyenne va-t-on choisir H_1 ?

La région critique est de la forme

$$W = \left\{ |\bar{X}_{n_1} - \bar{Y}_{n_2}| > l_\alpha \right\}.$$

Pour déterminer l_α , l'on a besoin de la loi de $\bar{X}_{n_1} - \bar{Y}_{n_2}$ sous l'hypothèse H_0 . Nous savons que

$$\begin{aligned} \bar{X}_{n_1} &\hookrightarrow \mathcal{N}\left(m_1, \frac{\sigma_1^2}{n_1}\right) \\ \bar{Y}_{n_2} &\hookrightarrow \mathcal{N}\left(m_2, \frac{\sigma_2^2}{n_2}\right). \end{aligned}$$

Comme ces deux variables sont indépendantes, on en déduit que

$$\bar{X}_{n_1} - \bar{Y}_{n_2} \hookrightarrow \mathcal{N}\left(m_1 - m_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

Ainsi nous avons

$$V = \frac{(\bar{X}_{n_1} - \bar{Y}_{n_2}) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$

Par suite, sous H_0 , nous obtenons

$$V = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1).$$

9.3.1 Résolution du test lorsque les variances connues

$$W = \left\{ \left| \bar{X}_{n_1} - \bar{Y}_{n_2} \right| > u_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right\}$$

Exemple 9.3.1. Revenons à l'exemple 9.1.2. Les variances sont connues, $\sigma_1^2 = \sigma_2^2 = 1$, $n_1 = n_2 = 8$ et les rendements moyens observés $\bar{x}_8 = 80q/h$ et $\bar{y}_8 = 81.5q/h$. On suppose que le seuil du test est $\alpha = 0.05$. De ce fait, $u_{0.975} = 1.96$. Nous avons donc

$$u_{0.975} \sqrt{\frac{1}{8} + \frac{1}{8}} = 0.98 \quad \bar{x}_8 - \bar{y}_8 = -1.5 < -0.98.$$

Nous décidons donc de rejeter H_0 . La variété 2 a un rendement moyen différent de celui de la variété 1.

9.3.2 Résolution du test lorsque les variances sont inconnues

Posons

$$Z = \frac{(n_1 - 1)S_{n_1}^2}{\sigma_1^2} + \frac{(n_2 - 1)S_{n_2}^2}{\sigma_2^2}.$$

Comme $\frac{(n_1 - 1)S_{n_1}^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$ et $\frac{(n_2 - 1)S_{n_2}^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$ et que ces deux variables sont indépendantes, nous obtenons $Z \sim \chi^2(n_1 + n_2 - 2)$. De plus, les variables aléatoires Z et V sont indépendantes. Par la définition de la loi de Student, nous déduisons que

$$T_{n_1, n_2} = \frac{V}{\sqrt{\frac{Z}{n_1+n_2-2}}} = \frac{\sqrt{n_1 + n_2 - 2}(\bar{X}_{n_1} - \bar{Y}_{n_2}) - (m_1 - m_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) \left(\frac{(n_1-1)S_{n_1}^2}{\sigma_1^2} + \frac{(n_2-1)S_{n_2}^2}{\sigma_2^2} \right)}} \sim \mathcal{T}(n_1 + n_2 - 2).$$

Sous l'hypothèse $H_0 : m_1 = m_2$, nous avons

$$T_{n_1, n_2} = \frac{\sqrt{n_1 + n_2 - 2}(\bar{X}_{n_1} - \bar{Y}_{n_2})}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) \left(\frac{(n_1-1)S_{n_1}^2}{\sigma_1^2} + \frac{(n_2-1)S_{n_2}^2}{\sigma_2^2} \right)}} \sim \mathcal{T}(n_1 + n_2 - 2).$$

On note que lorsque n_1 et n_2 sont grands, le caractère gaussien des observations n'est plus requis, et que T_{n_1, n_2} suit approximativement, sous H_0 , une loi $\mathcal{N}(0, 1)$.

Supposons que $\sigma_1^2 = \sigma_2^2$.

Si le test de Fisher accepte l'égalité des variances (H_0), nous avons

$$T_{n_1, n_2} = \sqrt{\frac{(n_1 + n_2 - 2)n_1 n_2}{n_1 + n_2}} \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{(n_1 - 1)S_{n_1}^2 + (n_2 - 1)S_{n_2}^2} \sim \mathcal{T}(n_1 + n_2 - 2)$$

La région critique au niveau $\alpha \in]0, 1[$ est

$$W = \left\{ \left| T_{n_1, n_2} \right| > t_{1-\frac{\alpha}{2}, n_1+n_2-2} \right\}$$

où $t_{1-\frac{\alpha}{2}, n_1+n_2-2}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student $\mathcal{T}(n_1 + n_2 - 2)$.

Supposons que $\sigma_1^2 \neq \sigma_2^2$.

A priori, si le test de Fisher rejette l'égalité des variances, on ne peut pas appliquer le test. On estime séparément σ_1^2 et σ_2^2 par leurs estimateurs S_1^2 et S_2^2 . Posons

$$T_{n_1, n_2} = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\frac{S_{n_1}^2}{n_1} + \frac{S_{n_2}^2}{n_2}}}.$$

Sous H_0 , $T_{n_1, n_2} \approx T([v])$

$$v = \frac{\left(\frac{S_{n_1}^2}{n_1} + \frac{S_{n_2}^2}{n_2} \right)^2}{\frac{S_{n_1}^4}{n_1^2(n_1-1)} + \frac{S_{n_2}^4}{n_2^2(n_2-1)}}.$$

La région critique au niveau $\alpha \in]0, 1[$ est

$$W = \left\{ \left| T_{n_1, n_2} \right| > q_{1-\frac{\alpha}{2}} \right\}$$

où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de Student $[v]$ degrés de liberté.

10.1 Test sur la valeur d'une proportion

Soient un échantillon (X_1, \dots, X_n) issu d'une loi de Bernouilli $\mathcal{B}(1, p)$ et p_0 une valeur possible de p . Nous savons que $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est un estimateur efficace de p . De plus, d'après le théorème central-limite, pour n assez grand, nous avons l'approximation en loi suivante

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \hookrightarrow \mathcal{N}(0, 1).$$

Au niveau $\alpha \in]0, 1[$, la région critique du test $H_0 : p \leq p_0$ contre $H_1 : p > p_0$ est :

$$W = \left\{ \bar{X}_n > \sqrt{\frac{p_0(1-p_0)}{n}} q_{1-\alpha} + p_0 \right\}$$

où $q_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de loi normale centrée-réduite $\mathcal{N}(0, 1)$.

Au niveau $\alpha \in]0, 1[$, la région critique du test $H_0 : p \geq p_0$ contre $H_1 : p < p_0$ est :

$$W = \left\{ \bar{X}_n < \sqrt{\frac{p_0(1-p_0)}{n}} q_\alpha + p_0 \right\}$$

où q_α est le quantile d'ordre α de loi normale centrée-réduite $\mathcal{N}(0, 1)$.

Au niveau $\alpha \in]0, 1[$, la région critique du test $H_0 : p = p_0$ contre $H_1 : p \neq p_0$ est :

$$W = \left\{ \bar{X}_n < p_0 - \sqrt{\frac{p_0(1-p_0)}{n}} q_{1-\frac{\alpha}{2}} \right\} \cup \left\{ \bar{X}_n > p_0 + \sqrt{\frac{p_0(1-p_0)}{n}} q_{1-\frac{\alpha}{2}} \right\}$$

où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de loi normale centrée-réduite $\mathcal{N}(0, 1)$.

10.2 Test de comparaison de deux proportions

Le problème se pose quand on veut comparer deux populations selon un critère qui est une proportion :

- Comparer les performances deux machines au vu de la proportion de pièces défectueuses qu'elles produisent.
- Comparer les proportions de soulards à Yopougon et Cocody pour vérifier les idées reçues.

Mathématiquement, on a une première population de taille n_1 et une seconde de taille n_2 . On veut comparer les deux population selon un critère. On note X_i et Y_i les variables aléatoires définies respectivement par

$$X_i = \begin{cases} 1 & \text{si le } i\text{ème individu de la population 1 présente la caractéristique} \\ 0 & \text{sinon} \end{cases}$$

$$Y_i = \begin{cases} 1 & \text{si le } i\text{ème individu de la population 2 présente la caractéristique} \\ 0 & \text{sinon.} \end{cases}$$

On note p_1 la probabilité qu'un individu de la population 1 possède la caractéristique et p_2 la probabilité qu'un individu de la population 2 possède la caractéristique. On souhaite comparer p_1 et p_2 . On suppose que

- X_1, \dots, X_{n_1} sont indépendantes
- Y_1, \dots, Y_{n_2} sont indépendantes
- (X_1, \dots, X_{n_1}) et (Y_1, \dots, Y_{n_2}) sont indépendants.

Alors $\sum_{i=1}^{n_1} X_i$ suit la loi binomiale $\mathcal{B}(n_1, p_1)$ et $\sum_{i=1}^{n_2} Y_i$ suit la loi binomiale $\mathcal{B}(n_2, p_2)$.

On se contentera ici de supposer que les tailles d'échantillons sont suffisamment grandes pour que l'on puisse faire l'approximation de la loi binomiale par la loi normale :

- $n_1 p_1 > 5$, $n_1(1-p_1) > 5$,
- $n_2 p_2 > 5$ et $n_2(1-p_2) > 5$.

Alors on peut considérer que $\sum_{i=1}^{n_1} X_i$ et $\sum_{i=1}^{n_2} Y_i$ sont des variables aléatoires indépendantes et approximativement de lois normales, respectivement $\mathcal{N}(n_1 p_1, n_1 p_1(1-p_1))$ et $\mathcal{N}(n_2 p_2, n_2 p_2(1-p_2))$.

Comme les estimateurs optimaux de p_1 et p_2 sont respectivement $\bar{X}_{n_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$ et $\bar{Y}_{n_2} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$, la région critique du test

$$H_0 : p_1 = p_2 \text{ contre } H_1 : p_1 \neq p_2$$

est donnée par

$$W = \left\{ \left| \bar{X}_{n_1} - \bar{Y}_{n_2} \right| > l_\alpha \right\}$$

où l_α est déterminé par l'équation

$$\mathbb{P}_{H_0}(W) = \alpha.$$

Sous les conditions ci-dessus, nous avons alors

$$\bar{X}_{n_1} \sim \mathcal{N}\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$$

$$\bar{Y}_{n_2} \hookrightarrow \mathcal{N}\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

Comme \bar{X}_{n_1} et \bar{Y}_{n_2} sont indépendantes, nous déduisons que

$$\bar{X}_{n_1} - \bar{Y}_{n_2} \hookrightarrow \mathcal{N}\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right).$$

Sous $H_0: p_1 = p_2 = p$, nous avons

$$\bar{X}_{n_1} - \bar{Y}_{n_2} \hookrightarrow \mathcal{N}\left(0, p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

et

$$\bar{X}_{n_1} - \bar{Y}_{n_2} \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \hookrightarrow \mathcal{N}(0, 1).$$

Comme p est inconnu, en remplaçant p par son estimateur $\hat{p} = \frac{n_1\bar{X}_{n_1} + n_2\bar{Y}_{n_2}}{n_1 + n_2}$ le résultat ci-dessus reste approximativement vrai. En posant

$$\hat{\sigma} = \sqrt{\frac{n_1\bar{X}_{n_1} + n_2\bar{Y}_{n_2}}{n_1 + n_2} \left(1 - \frac{n_1\bar{X}_{n_1} + n_2\bar{Y}_{n_2}}{n_1 + n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

sous l'hypothèse nulle H_0 la statistique

$$U = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\hat{\sigma}} \hookrightarrow \mathcal{N}(0, 1).$$

Au niveau $\alpha \in]0, 1[$, la région critique du test $H_0: p_1 \leq p_2$ contre $H_1: p_1 > p_2$ est :

$$W = \{U > q_{1-\alpha}\}$$

où $q_{1-\alpha}$ est le quantile d'ordre $1-\alpha$ de loi normale centrée-réduite $\mathcal{N}(0, 1)$.

Au niveau $\alpha \in]0, 1[$, la région critique du test $H_0: p_1 \geq p_2$ contre $H_1: p_1 < p_2$ est :

$$W = \{U < q_\alpha\}$$

où q_α est le quantile d'ordre α de loi normale centrée-réduite $\mathcal{N}(0, 1)$.

Au niveau $\alpha \in]0, 1[$, la région critique du test $H_0: p_1 = p_2$ contre $H_1: p_1 \neq p_2$ est :

$$W = \{|U| > q_{1-\frac{\alpha}{2}}\}.$$

où $q_{1-\frac{\alpha}{2}}$ est le quantile d'ordre $1 - \frac{\alpha}{2}$ de loi normale centrée-réduite $\mathcal{N}(0, 1)$.

Exercice 10.2.1. La machine 1 a produit 96 pièces dont 12 défectueuses. La machine 2 a produit 55 pièces dont 10 défectueuses. Peut-on en conclure que la machine 1 est significativement plus performante que la machine 2 ?

Solution.

Exercice 10.2.2. Une étude des décisions rendues par des jurys dans des cas de vols par effraction où l'accusé était de race noire a révélé les faits suivants : parmi les 28 cas où les victimes étaient de race noire, l'accusé a été trouvé coupable dans 12 cas ; parmi les 36 cas où la victime était de race blanche, l'accusé a été trouvé coupable dans 23 cas. Peut-on conclure que les jurys ont une plus forte tendance à déclarer coupables ceux qui sont accusés d'avoir commis des vols contre des Blancs ?

Solution.

11.1 Test d'adéquation à une loi donnée

11.1.1 Cas d'une loi discrète

On observe une variable aléatoire discrète X susceptible de prendre k valeurs

$$a_1, \dots, a_k.$$

On note $P = (p_1, \dots, p_k)$ le vecteur des probabilités définies par

$$p_j = \mathbb{P}(X = a_j), \quad j \in \{1, \dots, k\}.$$

On suppose que le vecteur P est inconnu. Soit $P^* = (p_1^*, \dots, p_k^*)$ un vecteur de probabilités connu ($\sum_{j=1}^k p_j^* = 1$). On veut résoudre le problème de test suivant :

$$H_0 : P = P^* \quad \text{contre} \quad H_1 : P \neq P^*.$$

Pour $j = 1, \dots, k$, on note

$$\hat{p}_j = \frac{N_j}{n}$$

la fréquence empirique de a_j ; N_j représente le nombre d'observations de la modalité a_j dans l'échantillon observé de taille n . Le vecteur des fréquences empiriques est

$$\hat{P} = (\hat{p}_1, \dots, \hat{p}_k).$$

Définition 11.1.1. On appelle distance du χ^2 , la quantité

$$T_n = n \sum_{j=1}^k \frac{(\hat{p}_j - p_j^*)^2}{p_j^*} = \sum_{j=1}^k \frac{(N_j - np_j^*)^2}{np_j^*}.$$

T_n mesure l'écart entre les effectifs observés et les effectifs "théoriques" sous l'hypothèse H_0

Au niveau $\alpha \in]0, 1[$, la région critique du test

$$W = \left\{ T_n > \chi_{1-\alpha, k-1}^2 \right\}$$

où $\chi_{1-\alpha, k-1}^2$ est le quantile d'ordre $1-\alpha$ de la loi de khi-deux $\chi(k-1)$ à $k-1$ degrés de liberté.

Remarque 11.1.1. En pratique, ce test marche bien si $n \geq 30$ et $np_j^* \geq 5$ pour tout j . Si cette condition n'est pas satisfaite, on peut regrouper les valeurs de a_j pour lesquelles p_j^* est trop faible.

Exercice 11.1.1. Lors de cent lancers d'un dé à six faces, on observe les résultats suivants :

x	1	2	3	4	5	6
Effectif observé	20	13	17	12	23	15
Effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Tester au niveau 5% l'hypothèse $H_0 = \{\text{le dé n'est pas pipé}\}$ contre l'hypothèse $H_1 = \{\text{le dé est pipé}\}$.

Solution.

11.1.2 Cas d'une loi continue

On observe X_1, \dots, X_n i.i.d. de même loi issue d'une loi P inconnue, continue. Etant donnée P^* une loi continue, on considère le problème de test d'hypothèses suivant

$$H_0 : P = P^* \quad \text{contre} \quad H_1 : P \neq P^*.$$

Dans cette situation, on doit partitionner \mathbb{R} en k classes A_j , $j = 1, \dots, k$. Pour appliquer les mêmes idées que plus haut, d'une part, k doit être assez grand pour que les lois discrètes, c'est-à-dire $\{p_j = P(A_j)\}$ et $\{p_j^* = P^*(A_j)\}$, soient assez proches des lois continues P et P^* . D'autre part, les probabilités $P(A_j)$ doivent être suffisamment grandes, pour que l'approximation asymptotique soit valable.

11.2 Test d'adéquation à une famille de lois

On veut tester si la loi de probabilité inconnue $P = (p_1, \dots, p_k)$ sur $\{a_1, \dots, a_k\}$ est égale à une loi $P^*(\theta) = (p_1^*(\theta), \dots, p_k^*(\theta))$, où $\theta = (\theta_1, \dots, \theta_s)$ est inconnu. On considère donc le problème de test suivant

$$H_0 : P = P^*(\theta) \quad \text{contre} \quad H_1 : P \neq P^*(\theta).$$

1. Comme précédemment, nous avons

$$T_n(\theta) = \sum_{j=1}^k \frac{(N_j - np_j^*(\theta))^2}{np_j^*(\theta)}$$

mais la quantité $T_n(\theta)$ n'est plus une statistique car θ est inconnu.

2. On estime θ par l'estimateur du maximum de vraisemblance $\hat{\theta}_n$. On obtient

$$T_n(\hat{\theta}_n) = \sum_{j=1}^k \frac{(N_j - np_j^*(\hat{\theta}_n))^2}{np_j^*(\hat{\theta}_n)}.$$

Sous H_0 , nous avons

$$T_n(\hat{\theta}_n) \xrightarrow{\mathcal{L}} \chi^2(k-s-1).$$

Au niveau $\alpha \in]0, 1[$, la région critique du test

$$W = \left\{ T_n(\hat{\theta}_n) > \chi^2_{1-\alpha, k-s-1} \right\}$$

où $\chi^2_{1-\alpha, k-s-1}$ est le quantile d'ordre $1-\alpha$ de la loi de khi-deux $\chi^2(k-s-1)$ à $k-s-1$ degrés de liberté.

Exercice 11.2.1. En se référant aux dates de début du pontificat (dates de consécration) et de fin (par décès, démission ou inaptitude), la durée d'exercice de chacun des 265 précédents papes (excepté François) a été calculée en nombre d'années. Les résultats groupés en cinq tranches sont présentés dans le tableau suivant :

Pontificat	Nombre de papes
moins d'une année	46
1 an - 5 ans	76
5 ans - 10 ans	68
10 ans - 20 ans	63
20 ans et plus	12

Que penser, au seuil de signification de 5%, de l'hypothèse selon laquelle la distribution du pontificat des papes serait une distribution exponentielle ?

Solution.

11.3 Test d'indépendance

On observe un couple (X, Y) à valeurs dans $\{c_1, \dots, c_r\} \times \{d_1, \dots, d_s\}$ et on veut tester si Y et Z sont indépendantes. On considère un échantillon de taille $((X_1, Y_1), \dots, (X_n, Y_n))$ de même loi que (X, Y) .

$$X \text{ et } Y \text{ sont indépendantes} \iff N_{ij} = \frac{N_{i\bullet} N_{\bullet j}}{n}$$

où

$$N_{i\bullet} = \sum_{j=1}^s N_{ij} \quad N_{\bullet j} = \sum_{i=1}^r N_{ij}.$$

La statistique de test est définie par

$$T_n = \sum_{j=1}^r \sum_{l=1}^s \frac{\left(N_{jl} - \frac{N_{j\bullet} N_{\bullet l}}{n} \right)^2}{\frac{N_{j\bullet} N_{\bullet l}}{n}}.$$

Sous l'hypothèse H_0 , la statistique T_n converge en loi vers $\chi^2((r-1)(s-1))$.

Au niveau $\alpha \in]0, 1[$, la région critique du test

$$W = \left\{ T_n > \chi^2_{1-\alpha, (r-1)(s-1)} \right\}$$

où $\chi^2_{1-\alpha, (r-1)(s-1)}$ est le quantile d'ordre $1-\alpha$ de la loi de khi-deux $\chi^2(r-1)(s-1)$ à $(r-1)(s-1)$ degrés de liberté.

Exercice 11.3.1. Une enquête sur l'influence de la ceinture de sécurité a donné les résultats suivants : sur 10.779 conducteurs ayant subit un accident l'enquête rapporte les effectifs dans le tableau qui suit selon la gravité et le port au non de la ceinture de sécurité :

Nature des blessures	Port de la ceinture	Pas de ceinture
Graves ou fatales	5	141
Blessures sérieuses	25	330
Peu ou pas de blessures	1229	9049

La ceinture de sécurité a-t'elle une influence sur la gravité des blessures lors d'un accident ?

Solution.