TABLES DE MATIERES

Liste des figures

Liste des tableaux

Introduction

0.1 Introduction

Les méthodes statistiques sont aujourd'hui utilisées dans presque tous les secteurs de l'activité humaine et font partie des connaissances de base de l'ingénieur, du gestionnaire, de l'économiste. Parmi les innombrables applications citons dans le domaine industriel : la fiabilité des matériels, le contrôle de la qualité, l'analyse des résultats de mesure et leur planification, la prévision et dans le domaine de l'économie et des sciences de l'homme : les modèles économétriques, les sondages, les enquêtes d'opinion, les études quantitatives de marchés, etc.

0.1.1 Définitions et notations

A- Unité statistique (Individus): On appelle unité statistique l'élément de base de l'ensemble que l'on veut étudier.

Exemple: Si l'on s'intéresse à l'ensemble des petites et moyennes entreprises (PME). Une PME est considérée comme unité statistique.

B- Population: On appelle population l'ensemble des unités statistiques.

Exemple : si l'on s'intéresse aux salariés d'une entreprise, alors l'ensemble des salariés forme la population.

C- Caractère statistique: On appelle modalité, l'aspect du caractère retenu dans le cadre de l'analyse.

Exemple : Pour une personne : la couleur des cheveux, la taille

- **Caractère qualitatif :** C'est un caractère qui ne peut pas être mesuré, ni être repéré.

  Exemple: situation familiale : célibataire, marié, veuf

- **Caractère quantitatif :** Le caractère quantitatif peut être repérable ou mesurable, c'est-è-dire susceptible d'être soumise è une mesure donnant une valeur numérique.

  Exemple : Le revenu des ménages est mesurable en Dirhams, Taille d'un individu, poids .

D- Variable statistique : C'est l'expression numérique ou littérale retenue dans le cadre de l'analyse. On la note X_i.
Exemple :

- Taille d'une entreprise : la variable exprimera les différentes tailles d'une entreprise : petite, moyenne, grande
- Nombre d'enfants à charge des familles : la variable sera numérique: $0, 1, 2, \ldots, n$ enfants à charge.

E- Effectif: C'est le nombre d'individus pouvant être rattachés une variable. On le note n_i.
Exemple: Dans une classe de 20 étudiants : 7 sont redoublants et 13 ne le sont pas.

$$n_1 = 7 \quad \text{et} \quad n_2 = 13.$$

E- Fréquence: Elle peut être de deux types : soit absolue (c'est l'effectif lui-même n_i), soit relative (rapport d'un effectif à l'effectif total). On note la frquence relative :

$$f = \frac{n_i}{N} \quad \text{où} \quad N = \sum_{i=1}^{p} n_i$$

0.1.2 Nature de la variable

- **Variable quantitative :**
Une variable statistique quantitative peut être soit discrète soit continue.

A- **Variable discrte :**
Lorsque la variable ne peut prendre qu'une valeur numérique, on parle de variable discrète.
Exemple: Nombre d'automobiles possédées par les ménages.

B- **Variable continue :**
On parle de variable continue lorsque celle-ci peut prendre plusieurs valeurs dans un intervalle donné. L'intervalle des valeurs possibles est divisé en n petits intervalles.

$$[\alpha, \beta] = [A_0, A_1[\cup[A_1, A_2[\cup \ldots \cup [A_{n-1}, A_n],$$

où $A_0 = \alpha < A_1 < \ldots < A_{n-1} < A_n = \beta$
Exemple

Durée d'une conversation téléphonique, la variable pourrait être la suivante pour un effectif de 34 personnes :

Temps Ecoulés (en mn) X_i	Effectifs n_i
[0, 5[	10
[5, 15[	9
[15, 30[	15
Total	34

- **Variable qualitative :**
  Une variable qualitative peut être nominale ou ordinale.
  Elle est nominale lorsque l'ensemble des modalités ne possède pas de structure particulière.
  Exemple1 : Dans un sondage, on s'intéresse à l'avis des consommateurs envers un produit : la variable X prend les modalités suivantes : " satisfait " ; " non satisfaits " et " ne sait pas ".

  Une variable est ordinale lorsque l'ensemble des modalités est ordonné.
  Exemple: Si on prend la variable X " État des voitures ", X prend les modalités : " Neuf ", " Moyen " et " Médiocre ". X est une variable ordinale.

CHAPITRE 1

Inférence Statistique - Échantillonnage

1.1 Échantillon aléatoire :

1.1.1 Introduction :

Une étude statistique sur tous les éléments d'une population P (à mois que nous soyons en présence d'un recensement, mais peu d'institutions peuvent se permettre un tel luxe) est souvent physiquement irréalisable et s'avère également, dans bien des cas très coûteuse. Alors comment obtenir certaines indictions fiables sur diverses caractéristiques d'une population sans en examiner tous les éléments ? Dans ce cas, l'observation doit être limitées à un sous-ensemble de la population P, appelée échantillon. Le choix de cet échantillon peut être fait de manière raisonnée ou aléatoire. Ainsi dans les sondages, le choix raisonné est utilisé lorsque l'on dispose d'informations a priori sur la distribution du caractère dans la population. Notons d'autre part que le choix aléatoire est le seul qui permette d'utiliser les résultats du calcul de probabilités, fournissant le cadre où nous nous plaçons. La distribution du caractère X dans l'échantillon de taille n est la même que la distribution du caractère X dans la population P, si le tirage le tirage aléatoire dans l'échantillon est équiprobable. En appelant X_i la variable aléatoire égale à la valeur du caractère X pour le i^{eme} individu, on peut définir une deuxième fois un n-échantillon par toute suite $(X_1, X_2, \ldots, X_n)$ de variables aléatoires indépendantes et identiquement distribuées. La théorie de l'échantillonnage se propose d'étudier les propriétés du n-uple $(X_1, X_2, \ldots, X_n)$ et les caractéristiques le résumant.

1.1.2 Statistique:

Il est d'usage dans la pratique de résumer les n valeurs d'un échantillon $x_1, x_2, \ldots, x_n$ par quelques caractéristiques simple telles que moyenne, variance, plu grande valeur, etc. Ces caractéristiques sont elles même des réalisations de variables aléatoires issues de $(X_1, X_2, \ldots, X_n)$.

Définition 1.1.1 *Une statistique T est une variable aléatoire fonction de* $(X_1, X_2, \ldots, X_n)$

$$T = f(X_1, X_2, \ldots, X_n)$$

Une statistique peut être à valeur dans $\mathbb{R}$ ou dans $\mathbb{R}^p$; dans le cas de $\mathbb{R}^p$, on parlera de statistique vectorielle.
Parmi les exemples de statistiques usuelles, citons :

- **La moyenne empirique:**
Elle est notée $\overline{X}$ et elle est définie par :

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

C'est la moyenne arithmétique des variables aléatoires de l'échantillon.

- **La fréquence empirique :**
Soit un caractère qualitatif à deux modalités A et B que l'on peut coder par 0 et 1. On note p la proportion dans la population des individus qui possèdent la modalité A. Le nombre aléatoire Y_n d'individus de l'échantillon qui ont la modalité A, s'écrit :

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

où les v.a X_i sont des v.a de Bernoulli de paramètre p. On pose $F_n = \dfrac{Y_n}{N}$ appelée fréquence empirique.

- **La variance empirique:**
La variance empirique du caractère X dans l'échantillon est définie par :

$$S^2 = \frac{1}{N}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{N}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

Par ailleurs, une autre statistique utile en estimation est la variance empirique corrigée S^{*2} définie par:

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

- **Les valeurs extrêmes et l'étendue:**
La plus petite des valeurs prise par $X_1, X_2, \ldots, X_n$ est notée $Y_1 = \min(X_1, X_2, \ldots, X_n)$, la plus grande $Y_n = \max(X_1, X_2, \ldots, X_n)$. L'étendue est alors $R = Y_n - Y_1$. A noter que ce sont trois variables aléatoires.

1.1.3 Propriétés élémentaires

On note $m_r = E[X^r]$ le moment d'ordre r de la variable aléatoire X, $r = E[(X-m)^r]$ le moment centré correspondant et $\sigma^2 = \mu_2$ la variance de X. Un premier résultat donne l'espérance et la variance de la moyenne empirique $\overline{X}$.

Proposition 1.1.1 *Soit $(X_1, X_2, \ldots, X_n)$ un n-échantillon de la v.a. X. Si $E[X_2]$ existe, alors pour tout entier n, $n > 1$, la moyenne empirique $\overline{X}$ vérifie :*

$$E(\overline{X}) = m$$

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

Preuve. $E(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n.m$
$Var(\overline{X}) = \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} n\sigma^2$ (D'après l'indépendance des X_i). ∎

La signification de cette proposition est que " en moyenne ", la variable aléatoire X est égale à m avec une dispersion d'autant plus faible que n est grand.
Dans le cas particulier d'une fréquence empirique, on obtient :

$$E(F_n) = p$$

$$Var(F_n) = \frac{p(p-1)}{p};$$

où p est la proportion des individus dans la population qui possèdent la modalité A d'un caractère qualitatif à deux modalités A et B.

Proposition 1.1.2 *Soit X une variable aléatoire telle que $m_4 = E[X^4]$ existe. Alors, pour tout entier $n > 1$,*

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

$$E(S^{*2} = \sigma^2.$$

En outre, $Var[S^2]$ et $Var[S^{*2}]$ sont majorées par une expression de la forme $\frac{c}{n}$, où c est une constante dépendant de m_4.

Preuve.

$$\begin{aligned}
E(S^2) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}E(X_i^2)-E(\overline{X}^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}\big(Var(X_i)+E(X_i)^2\big)-\big(Var(\overline{X})+E(\overline{X})^2\big) \\
&= \frac{1}{n}\sum_{i=1}^{n}\big(Var(X)+E(X)^2\big)-\frac{\sigma^2}{n}-m^2 \\
&= \sigma^2+m^2-\frac{\sigma^2}{n}-m^2 \\
&= \frac{n-1}{n}\sigma^2.
\end{aligned}$$

De même

$$E(S^{*2}) = E\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}E(S^2) = \sigma^2$$

∎

1.2 Cas des grands échantillons:

1.2.1 La loi des grands nombres:

Définition 1.2.1 *La suite infinie $(Y_n)_n$ de variables aléatoires converge en probabilité vers la v.a. Y si, pour tout $\epsilon > 0$,*

$$Y_n \xrightarrow{P} Y \Leftrightarrow \lim_n \left[|Y_n - Y| < \epsilon\right] = 1$$

A partir de cette définition, le théorème suivant énonce la loi des grands nombres qui traduit la convergence de la moyenne empirique vers la moyenne théorique m = E(X), lorsque n tend vers l'infini.

Théorème 1.2.1 ((Loi des grands nombre)) *Soit X une v.a. de moyenne m finie, alors pour toute suite infinie $(X_n)_n$ de v.a. deux à deux indépendantes et de même loi que X, la suite des v.a.*

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

Converge en probabilité vers m lorsque n tend vers l'infini.

Ce théorème montre que la moyenne empirique $\overline{X}_n$ est proche de la moyenne théorique m avec une grande probabilité lorsque la taille de l'échantillon est suffisamment grande.
Application: Grâce à la loi des grands nombres, on peut montrer que :

$$S_n^2 \xrightarrow{P} \sigma^2,$$

de même

$$S_n^{*2} \xrightarrow{P} \sigma^2,$$

Quand n tend vers l'infini et à condition que $E(X^2)$ existe.

1.2.2 Le théorème central limite:

Le théorème central limite précise comment la suite de v.a. tend vers l'infini, en donnant la forme de la distribution limite.

Théorème 1.2.2 ((théorème central limite)) *Soit X une v.a. de moment $E[X^2]$ fini, alors pour toute suite infinie $(X_n)n$ de v.a. indépendantes et de même loi que X, la suite des v.a.*

$$U_n = \frac{\overline{X}_n - m}{\sigma/n},$$

tend en loi vers la loi normale centrée réduite $N(0,1)$ lorsque n tend vers l'infini.

Application: Cas d'une fréquence empirique
On prélève indépendamment et avec remise n individus d'une population séparée en deux sous-populations et de proportions p et $(1-p)$ (pièces défectueuses ou correctes dans une production industrielle par exemple). Soit F_n la fréquence empirique. On a

$$\begin{aligned} E[F_n] &= p \\ Var[F_n] &= \frac{p(1-p)}{n}. \end{aligned}$$

Et si n est grand,

$$F_n \approx N\left(p, \sqrt{\frac{p(1-p)}{n}}\right).$$

Exemple :
Pour un échantillon de 400 pièces issues d'une fabrication où 10% sont défectueuses. On peut s'attendre à trouver dans 95% des cas un pourcentage de pièces défectueuses dans l'échantillon comprise entre: $10\% \pm 1.96\sqrt{\dfrac{0.1 \times 0.9}{400}}$

Corollaire 1.2.1 *Soit X une v.a. de moment $E[X^2]$ fini, alors pour toute suite infinie $(X_n)_n$ de v.a indépendantes et de même loi que X, la moyenne empirique $\overline{X}_n$ tend vers la loi normale $N\left(m, \dfrac{\sigma}{\sqrt{n}}\right)$*

Exemple:
Soit un lot de 500 chocolats. Le poids d'un chocolat est une v.a. telle que $m = 5g$ et $\sigma = 0.5g$. Quelle est la probabilité qu'une boite de 50 chocolats issus de ce lot ait un poids total supérieur à $260g$?

Solution:
L'échantillon étant grand ($n = 50 > 30$), on peut appliquer le théorème central limite

$$\overline{X} \approx N\left(5, \frac{0.5}{\sqrt{50}}\right) \text{ approximativement.}$$

On pose $T = 50\overline{X}$; cette nouvelle v.a. suit approximativement

$$\begin{aligned} T &\approx N\left(5 \times 50, \frac{50 \times 0.5}{\sqrt{50}}\right) \\ &= N(250, 0.5\sqrt{50}). \end{aligned}$$

Calculons

$$\begin{aligned} P[T > 260] &= P\left[T > \frac{260 - 250}{0.5\sqrt{50}}\right] \\ &= P(U > 2.83) \\ &= 1 - P(U < 2.83) \\ &= 1 - 0.9977 \end{aligned}$$

Proposition 1.2.1 *Si X est une v.a. d'espérance m et d'écart-type σ, si g est une fonction numérique dérivable au voisinage de m et vérifiant $g'(m) \neq 0$, si enfin $\overline{X}_n$ désigne la moyenne empirique d'un n-échantillon de X, alors la suite des v.a.*

$$g'(m)\frac{g(\overline{X}_n) - g(m)}{\sigma/\sqrt{n}},$$

tend en loi vers la loi normale $N(0, 1)$ lorsque n tend vers l'infini.

Autrement dit, si n est suffisamment grand, la v.a. $g(\overline{X}_n) - g(m)$ suit approximativement la loi normale $N\left(g(m), \dfrac{\sigma}{\sqrt{n}g'(m)}\right)$.
En pratique, la loi U_n peut être assimilée à la loi $N(0, 1)$ dès que n est supérieur ou égale à 30.

1.3 Cas des petits échantillons:

Lorsque la taille de l'échantillon est petite (inférieur à 30 en pratique) les résultats précédents ne sont plus applicables. Dans ce paragraphe, on ne considère que le cas de la loi normale.
Soit X une v .a. normale d'espérance m et d'écart-type σ . La densité associées est définie pour tout réel par:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right].$$

De plus, on appelle échantillon normal tout échantillon de la v.a. X. Le résultat suivant sur la stabilité de la loi normale affirme que la somme de v.a. normales indépendantes est encore une v.a. normale.

Proposition 1.3.1 *Si $X_1, X_2, \ldots, X_n$ sont des v.a. indépendantes de lois respectives $N(m_i, \sigma_i^2)$, alors la variable aléatoire $Y = X_1 + \ldots + X_n$ suit la loi normale d'espérance mathématique $m = m_1 + \ldots + m_n$ et d'écart-type $\sigma = \sqrt{\sigma_1^2 + \ldots + \sigma_n^2}$.*

La proposition montre que lorsque l'échantillon est normal, la moyenne empirique suit une loi, quelque soit la taille de l'échantillon. Pour un échantillon de loi quelconque, cette propriété n'est pas vraie. Toutefois le théorème central limite permet d'affirmer qu'elle est approximativement vérifiée à condition que n soit assez grand.

CHAPITRE 2

Estimation: Estimation ponctuelle

2.1 Introduction

L'estimation consiste à donner des valeurs approchées aux paramètres d'une population (m, , etc.) à l'aide d'un échantillon de n observations issues de cette population. On suppose vérifiée l'hypothèse d'échantillonnage aléatoire simple.

2.1.1 Exemples élémentaires:

Les lois grands nombres justifient l'usage de $\overline{X}_n$ et de S_n^2 comme estimations de m et de σ^2 respectivement. On sait que $\overline{X}_n \xrightarrow{P} m$ et $S_n^2 \xrightarrow{P} \sigma^2$. De même la fréquence empirique f d'un événement est une estimation de sa probabilité p.
Les variables aléatoires $\overline{X}_n$, S_n^2, F_n sont appelées alors estimateurs de m, σ^2, p respectivement.
Cependant le même paramètre peut être estimé à l'aide d'estimateurs différents. Afin de choisir entre plusieurs estimateurs possibles d'un même paramètre, il faut définir les qualités d'un estimateur.

2.1.2 Qualité d'un estimateur:

Soit le paramètre à estimer θ et T_n un estimateur, c'est-à-dire une fonction des X_i à valeurs dans un domaine acceptable pourθ.
La première qualité d'un estimateur est d'être convergent. Il est souhaitable que si $n \longrightarrow \infty$, $T_n \longrightarrow \theta$, c'est le cas des estimateurs présentés précédemment. Deux estimateurs convergents ne convergent cependant pas nécessairement à la même vitesse, ceci est lié, pour une taille d'échantillon donnée, à la notion de précision d'un estimateur.
L'erreur d'estimation $T_n - \theta$ qui est une variable aléatoire se décompose de façon élémentaire en:

$$T_n - \theta = T_n - E(T_n) + E(T_n) - \theta,$$

où $E(T_n)$ représente l'espérance de l'estimateur.
$T_n - E(T_n)$ représente les fluctuations aléatoires de T_n autour de sa valeur moyenne tandis que $E(T_n) - \theta$ est assimilable à une erreur systématique due au fait que T_n varie autour de sa valeur centrale $E(T_n)$ et non autour de θ. La quantité $E(T_n) - \theta$ s'appelle le biais. Il est donc souhaitable d'utiliser des estimateurs sans biais, tels que $E(T_n) = \theta$. Ainsi, $\overline{X}_n$ est un estimateur sans biais de m, mais S_n^2 est biaisé pour σ^2. Il est donc souvent préférable d'utiliser $S_n^{*2} = \dfrac{n}{n-1} S_n^2$ pour estimer σ^2.

Cependant, il ne faut pas croire que S_n^* est un estimateur sans biais de σ.
En effet

$$E(\sqrt{S_n^{*2}}) \neq \sqrt{E(S_n^{*2})}$$

On mesure généralement la précision d'un estimateur T_n par l'erreur quadratique moyenne:

$$E|(T_n - \theta)^2|.$$

On peut écrire

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

Comme $E(T_n) - \theta$ est une constante et que $E[(T_n - E(T_n)] = 0$, on en déduit que:

$$E[(T_n - \theta)^2] = Var(T_n) + [E(T_n) - \theta]^2.$$

De deux estimateurs sans biais, le plus précis est donc celui de variance minimale.
Exemple:
Montrer que si m est connu, l'estimateur $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - m)^2$ est meilleur que S_n^{*2}.

Définition 2.1.1 *L'estimateur T_n est dit asymptotiquement sans biais si*

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

Exemple:
$S_n^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ est un estimateur asymptotiquement sans biais de σ^2.

Définition 2.1.2 *Un estimateur T_n du paramètre θ est dit convergent en moyenne quadratique (m.q) s'il vérifie:*

$$\lim_{n \to +\infty} E[(T_n - \theta)^2] = 0$$

Rappelons que la convergence en moyenne quadratique entraîne la convergence en probabilité, mais que la réciproque est fausse en général. En pratique, pour montrer le caractère convergent d'un estimateur, on essaie le plus souvent de montrer la convergence en moyenne quadratique.

Proposition 2.1.1 *Un estimateur T_n du paramètre est convergent en moyenne quadratique si et seulement si, il est asymptotiquement sans biais et si sa variance tend vers* 0 *lorsque n tend vers l'infini.*

Preuve. Il suffit de voir que

$$E[(T_n-\theta)^2] = Var(T_n) + [E(T_n)-\theta]^2,$$

pour conclure. ∎

Exemple:
Pour estimer la moyenne théorique $\theta = m$ d'une variable aléatoire X, on peut prendre $T_n = \overline{X}_n$. Cet estimateur est sans biais, si de plus sa variance est finie, la relation $Var(\overline{X}_n) = \dfrac{Var(X)}{n}$ montre que $Var(\overline{X}_n)$ tend vers 0 lorsque n tend vers l'infini. La proposition précédente permet alors d'affirmer que $\overline{X}_n$ est un estimateur convergent en moyenne quadratique de $m = E(X)$.

Le critère de convergence en moyenne quadratique donné dans la proposition précédente suppose le calcul de l'espérance mathématique et la variance de l'estimateur T. Comme, en pratique ce calcul n'est pas toujours aisé, on propose une méthode permettant dans certains cas, de démontrer la convergence en probabilité de T vers θ.

Proposition 2.1.2 *Soit $(Y_n)_n$ une suite infinie de v.a. qui converge en probabilité vers la v.a. Y et si g est une fonction d'une variable réelle, continue sur l'ensemble des valeurs possibles de Y et des Y_n, alors la suite des v.a. $g(Y_n)$ converge en probabilité vers $g(Y)$ lorsque n tend vers l'infini, soit*

$$g(Y) = \lim_{n\to+\infty} g(Y_n)$$

Exemple:
L'estimateur S_n^2 (et aussi S_n^{*2}) est convergent en moyenne quadratique vers σ^2. Il est donc convergent en probabilité, or la fonction g définie par $g(u) = \sqrt{u}$ est continue sur $\mathbb{R}$, d'après la proposition précédente, $S_n = \sqrt{S_n^2}$ tend donc en probabilité vers σ lorsque n tend vers l'infini. S_n est donc un estimateur convergent du paramètre σ.

Définition 2.1.3 *Soit T_n et T_n' deux estimateurs sans biais de θ. T_n est dit plus efficace que T_n' si:*

$$Var(T) \leq Var(T_n')$$

Définition 2.1.4 *L'estimateur sans biais et de convergence minimale est appelé estimateur efficace.*

2.2 Estimation ponctuelle des paramètres usuels:

2.2.1 Estimation de la moyenne:

Soit X une variable aléatoire dont on veut estimer la moyenne (ou espérance) $\mu = E(X)$ à partir d'un n-échantillon $(X_1, X_2, \ldots, X_n)$ de X. On ne suppose rein sur la loi de X

Proposition 2.2.1 $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$; *la moyenne empirique est un estimateur efficace de* μ

Preuve. $\overline{X}_n$ est sans biais de, car $E(\overline{X}_n) = \mu$ et de plus

$$Var(\overline{X}_n = \frac{Var(\overline{X}_n)}{n} \overset{n\to+\infty}{\longrightarrow} 0$$

∎

$\forall T$ un autre estimateur de μ,

$$Var(T) \geq Var(\overline{X}_n).$$

2.2.2 Estimateur de la variance d'une population Gaussienne:

Soit X une v.a. qui suit une loi normale $N(\sigma, \mu)$. On veut estimer la variance σ^2 de X.

A- μ est connu:

Proposition 2.2.2 $T_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ *est un estimateur efficace de* μ

En effet

$$\begin{aligned}
E(T_n^2) &= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\frac{1}{n}\sum_{i=1}^{n} X_i\mu + \mu^2\right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - \frac{2}{n}\mu\sum_{i=1}^{n} E(X_i) + \mu^2\\
&= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - \mu^2\\
&= \frac{1}{n}\sum_{i=1}^{n} \left[Var(X_i) + (E(X_i))^2\right] - \mu^2\\
&= \sigma^2 + \mu^2 - \mu^2\\
&= \sigma^2
\end{aligned}$$

Donc T_n^2 est sans biais. De plus

$$Var(T_n^2 = \frac{Var(\overline{X}_n)}{n} \overset{n\to+\infty}{\longrightarrow} 0,$$

et est minimale

B- μ est inconnu:

Proposition 2.2.3 $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ *est un estimateur biaisé de σ^2, mais asymptotiquement sans biais.*

Preuve. Déjà fait ∎

2.3 Méthode du maximum de vraisemblance

Les paragraphes précédents ont donné quelques exemples d'estimateurs. Ainsi, la moyenne empirique est un estimateur sans biais et convergent (en moyenne quadratique) de la moyenne théorique. La variance empirique corrigée S_n^{*2} est un estimateur sans biais et convergent de la variance théorique.
Au cas où le paramètre inconnu n'est ni la moyenne, ni la variance de la v.a. X associée au caractère étudié, on peut utiliser des méthodes permettant de construire les estimateurs.

Définition 2.3.1 *La vraisemblance d'un échantillon* $(X_1, X_2, \ldots, X_n)$ *de la v.a.* X *est la fonction* L *de* $n+1$ *variables définie par:*

$$L(x_1, x_2, \ldots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta),$$

où $x-1, x2, \ldots, x_n$ *sont des réalisations possibles de* X *et* θ *le paramètre inconnu. On rappelle que* $f(x, \theta)$ *définie la loi de la v.a.* X *par la formule:*

$$f(x, \theta) = p[X = x],$$

si X *est une variable aléatoire discrète, et dans le cas continue par:*

$$f(x, \theta) = f_X(x)$$

Remarque. Compte tenu de l'indépendance des v.a. $X_1, X_2, \ldots, X_n$, la vraisemblance $L(x_1, x_2, \ldots, x_n, \theta)$ n'est autre que la probabilité de réalisation de la suite des valeurs $x_1, x_2, \ldots, x_n$ lorsque le paramètre inconnu vaut θ.
La méthode du maximum de vraisemblance, consiste étant donné un échantillon $x_1, x_2, \ldots, x_n$ à prendre comme estimation de θ la valeur de θ qui rend maximale la vraisemblance $L(x_1, x_2, \ldots, x_n, \theta)$. Il s'agit là d'un problème d'optimisation: on cherche une valeur θ_0 de θ qui maximise la fonction $L(x_1, x_2, \ldots, x_n, \theta)$.

Si la fonction de vraisemblance est dérivable, on peut chercher ce maximum en calculant sa dérivée. Enfin, la fonction de vraisemblance étant positive, on peut aussi se borner à rechercher le maximum sur l'ensemble des où cette fonction est strictement positive. Sur cet ensemble, il est alors commode et équivalent de chercher le maximum de la fonction:

$$\log L(x_1, x_2, \ldots, x_n, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta).$$

Une condition nécessaire sur θ est:

$$\frac{\partial \log L(x_1, x_2, \ldots, x_n, \theta)}{\partial \theta} = 0,$$

pour $\theta = \theta_0$.
Et si elle remplie, une condition suffisante pour θ_0, alors on:

$$\frac{\partial^2 \log L(x_1, x_2, \ldots, x_n, \theta)}{\partial^2 \theta} < 0,$$

pour $\theta = \theta_0$.
L'estimateur du maximum de vraisemblance étant alors:

$$\widehat{\theta}_n = \theta_0(X_1, X_2, \ldots, X_n).$$

Exemple:
Estimation pour la moyenne et la variance d'une loi normale:
Dans ce cas:

$$\begin{array}{rcl} X & \rightsquigarrow & N(m, \sigma^2) \\ \theta & = & (\theta_1, \theta_2), \end{array}$$

où $\theta_1 = m$ et $\theta_2 = \sigma^2$.
De plus:

$$f(x, \theta) = f(x, m, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right].$$

La vraisemblance de l'échantillon s'écrit:

$$L(x_1, x_2, \ldots, x_n, \theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - m}{\sigma}\right)^2\right],$$

et le logarithme népérien est

$$\log L(x_1, x_2, \ldots, x_n, \theta) = n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - m}{\sigma}\right)^2 - n \log \sigma.$$

Les conditions nécessaires consistent en l'annulation des dérivées partielles premières par rapport à m et σ^2. Elles s'écrivent respectivement:

$$\left\{\begin{array}{l} \dfrac{\partial \log L}{\partial m} = -\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i - m) = 0, \\ \dfrac{\partial^2 \log L}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i - m)^2 = 0. \end{array}\right.$$

De ces deux relations, on conclut

$$\left\{\begin{array}{l} m_0 = \overline{x}, \\ \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2. \end{array}\right.$$

CHAPITRE 3

Estimation par intervalle de confiance

3.1 Principe de la méthode:

Il est souvent plus réaliste et plus intéressant de fournir un renseignement du type $a < \theta < b$ plutôt que d'écrire sèchement $\widehat{\theta} = c$.
Fournir un tel intervalle $[a, b]$ s'appelle donner une estimation par intervalle de confiance de θ.
Principe de la méthode:
La méthode des intervalles de confiance est la suivante: Soit T un estimateur de θ on prendra évidemment le meilleur estimateur possible), dont on connaît la loi de probabilité pour chaque valeur de θ. Étant donné une valeur θ_0 de θ, on peut déterminer un intervalle de probabilité de niveau $1-\alpha$ pour T, c'est-à-dire deux bornes t_1 et t_2 telles que

$$p[t_1 < T < t_2] = 1 - \alpha,$$

pour $\theta_0 = \theta$.
Ces bornes dépendent évidemment de θ_0.
On choisira dans la plus parts des cas un intervalle de probabilité à risque symétrique $\frac{\alpha}{2}$ et $\frac{\alpha}{2}$.
On adopte alors la règle de décision suivante:
Soit t la valeur observée de T:

- si t appartient à l'intervalle $[t_1, t_2]$, on conserve θ_0 comme valeur possible de θ;
- si t n'appartient pas à l'intervalle $[t_1, t_2]$, on élimine θ_0.

On répète cette opération pour toutes les valeurs de θ.

3.2 Estimation de la moyenne d'une variable de Laplace Gauss

3.2.1 L'écart-type σ est connu

$\overline{X}_n$ est le meilleur estimateur de m et $\overline{X}_n$ suit une loi $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$. L'intervalle de probabilité de $\overline{X}_n$ à $1-\alpha$ est:

$$m - u_\alpha \frac{\sigma}{\sqrt{n}} < \overline{X}_n < m + u_\alpha \frac{\sigma}{\sqrt{n}},$$

d'où l'intervalle de confiance:

$$\overline{x}_n - u_\alpha \frac{\sigma}{\sqrt{n}} < m < \overline{x}_n + u_\alpha \frac{\sigma}{\sqrt{n}}.$$

Si $1 - \alpha = 0.95$, on a $u_\alpha = 1.96$

3.2.2 L'écart-type σ est inconnu

On utilise le fait que $T = \dfrac{\overline{X}_n - m}{S_n^*/\sqrt{n}} = \dfrac{\overline{X}_n - m}{S_n/\sqrt{n-1}}$ suit une loi de student à n degrés de liberté.
L'intervalle de probabilité pour t est

$$-t_\alpha < \frac{\overline{X}_n - m}{S}\sqrt{n-1} < t_\alpha.$$

D'où l'intervalle de confiance:

$$\overline{x}_n - t_\alpha \frac{S}{\sqrt{n-1}} < m < \overline{x}_n + t_\alpha \frac{S}{\sqrt{n-1}},$$

où bien

$$\overline{x}_n - t_\alpha \frac{S^*}{\sqrt{n}} < m < \overline{x}_n + t_\alpha \frac{S^*}{\sqrt{n}}.$$

Le théorème central-limite a pour conséquence que les intervalles précédents sont valables pour estimer la moyenne théorique m d'une loi quelconque à condition que n soit assez grand.

3.3 Estimation de l'écart-type d'une loi de Laplace-Gauss

3.3.1 m est connu:

$T = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - m)^2$ est le meilleur estimateur de σ^2 et $\dfrac{nT}{\sigma^2}$ suit la loi χ^2_n (chi deux à n degrés de liberté) comme somme de n carrés de loi $N(0,1)$ indépendantes.
Soient k_1 et k_2 les bornes de l'intervalle de probabilité d'une variable χ^2_n:

$$p\left[k_1 < \frac{nT}{\sigma^2} < k_2\right] = 1 - \alpha$$

Figure 3.1: Fonction densité de la loi de χ^2_n

L'intervalle de confiance est:

$$\frac{nt}{k_1} < \sigma^2 < \frac{nt}{k_2},$$

où t est la valeur prise par T.

3.3.2 m est inconnu:

On utilise $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$ et on sait que $\frac{nS^2}{\sigma^2}$ suit la loi χ^2_{n-1}.
Soit l_1 et l_2 les bornes de l'intervalle de probabilité;

$$p\left[l_1 < \frac{nS^2}{\sigma^2} < l_2\right] = 1 - \alpha,$$

on a alors:

$$\frac{ns^2}{l_1} < \sigma^2 < \frac{ns^2}{l_2},$$

Remarque. Ces formules sont valables pour X suivant une loi normale exclusivement.

3.4 Intervalle de confiance pour le paramètre d'une loi binomiale quand n est grand

C'est le problème connu sous le nom d'intervalle de confiance pour une proportion p inconnu. Etant donnée une population infinie (ou fini si le tirage s'effectue avec remise) où une proportion p des individus possédant un certain caractère. Il s'agit de trouver un intervalle de confiance pour p à partir de f, proportion trouvée sur un échantillon de taille n.
On sait que nF suit une loi binomiale $B(n, p)$; et si n est assez grand, on sait que:

$$nF \rightsquigarrow N(np, \sqrt{np(1-p)}),$$

donc que

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

L'intervalle de probabilité symétrique est:

$$p - u_\alpha\sqrt{\frac{p(1-p)}{n}} < F < p + u_\alpha\sqrt{\frac{p(1-p)}{n}}.$$

Posons $u = k$ pour simplifier les notations.
Les bornes de l'intervalle de probabilité sont données par:

$$y = p \pm k\sqrt{\frac{p(1-p)}{n}}.$$

Soit

$$(y-p)^2 = \frac{k^2 p(1-p)}{n},$$

ou

$$y^2 + p^2\left(1 + \frac{k^2}{n}\right) - 2py - \frac{k^2}{n}p = 0,$$

équation d'une ellipse passant par l'origine et le point $(1, 1)$, points pour lesquels elle a une tangente verticale.
Étant donné une valeur f observée, l'intervalle de confiance s'obtient en résolvant en p l'équation:

$$f^2 + p^2\left(1 + \frac{k^2}{n}\right) - 2pf - \frac{k^2}{n}p = 0,$$

ou

$$p^2\left(1 + \frac{k^2}{n}\right) - p\left(\frac{k^2}{n} + 2f\right) + f^2 = 0$$

$$\begin{aligned}\Delta &= p\left(\tfrac{k^2}{n} + 2f\right)^2 - 4\left(1 + \tfrac{k^2}{n}\right)f \\ &= \frac{k^4}{n^2} + 4f\frac{k^2}{n} - 4f^2\frac{k^2}{n}.\end{aligned}$$

D'où

$$p = \frac{\left(2f + \frac{k^2}{n}\right) \pm \sqrt{\frac{k^4}{n^2} + 4f\frac{k^2}{n} - 4f^2\frac{k^2}{n}}}{2\left(1 + \frac{k^2}{n}\right)}.$$

Formule encombrante, mais dont on peut trouver une approximation en considérant que n est grand et en faisant un développement limité au premier ordre en $(\frac{1}{n})$; le premier terme

$$\frac{\left(2f + \frac{k^2}{n}\right)}{2\left(1 + \frac{k^2}{n}\right)} \approx f + o\left(\frac{1}{n^2}\right),$$

le second terme se réduit en simplifiant par n^2:

$$\sqrt{\frac{k^2 + fnk^2 - 4f^2nk^2}{4(n+k^2)^2}} = \sqrt{\frac{k^2 + fnk^2 - 4f^2nk^2}{4n^2 + 8k^2n + 4k^4}}.$$

Ce radical est équivalent au suivant (en écrivant que chaque terme est équivalent à celui de plus haut degré en n)

$$\sqrt{\frac{fnk^2 - f^2nk^2}{n^2}} = k\sqrt{\frac{f(1-f)}{n}}.$$

Donc, on a si n est grand, l'expression approchée suivante pour l'intervalle de confiance:

$$f - u_\alpha\sqrt{\frac{f(1-f)}{n}} < p < f + u_\alpha\sqrt{\frac{f(1-f)}{n}}.$$

Annexe: Tables

Tables des principales lois de probabilité

BIBLIOGRAPHIE

[1] G. Saporta, "Probabilités, Analyse des données et Statistique", Edition TECHNIP, 1990.

[2] D. Fourdrinier " Statistique inférentielle : Cours et exercices corrigés ", Collection Scienecs Sup, Dunod, 2002.

[3] P. Souvay, " Savoir utiliser la statistique. Outil d'aide à la décision et à l'amélioration de la qualité ", Afnor, 2002.

[4] M. Tenenbaus, " Méthodes statistiques en gestion ", Dunod (Entreprise), 1994.

[5] J.P Pouget, N. Boy, P. Bénichou, G. Demengel "Probabilités : Statistiques inférentielles, fiabilité, outils pour l'ingénieur", Editeur : Ellipses Marketing, 1997