

LICENCE D'INGÉNIERIE MATHÉMATIQUES

Année Universitaire 2006-2007

COURS DE STATISTIQUE INFÉRENTIELLE

(Version Avril 2007)

JÉRÔME DUPUIS

UNIVERSITÉ PAUL SABATIER (Toulouse III)

Laboratoire de Statistique et Probabilités UMR CNRS C5583

SOMMAIRE

1. Exemples introductifs et généralités

1.1 Exemples introductifs

1.2 Généralités

2. Modèles statistiques

2.1 Le modèle statistique d'échantillonnage paramétrique

2.2 Autres modèles statistiques

3. Exhaustivité et information

3.1 Un exemple introductif

3.2 Exhaustivité.

3.3 Information de Fisher

4. L'estimation ponctuelle

4.1 Définition d'un estimateur

4.2 Propriétés d'un estimateur

4.3 Comparaison entre estimateurs

4.4 Estimateur du maximum de vraisemblance

4.5 Estimateur des moments

5. L'estimation par intervalle de confiance

5.1 Définition, exemple et commentaires

5.2 Quelques principes généraux

5.3 Intervalles de confiance classiques

6. Les tests

6.1 Exemples introductifs

6.2 La problématique d'un test

6.3 L'approche de Neyman-Pearson

6.4 Recherche de test uniformément plus puissant

6.5 La pratique des tests.

6.5.1 Tests portant sur la moyenne μ d'une v.a.r. X .

- la loi de X est normale (avec σ^2 connue, puis σ^2 inconnue).

- la loi de X est quelconque.

6.5.2 Tests portant sur la variance σ^2 d'une v.a.r. X .

- la loi de X est normale (avec μ connue, puis μ inconnue).

- la loi de X est quelconque

6.5.3 Comparaison des moyennes et variances pour 2 échantillons gaussiens indépendants.

- comparaison des variances (test de Fisher).

- comparaison des moyennes (test de Student).

6.5.4 Comparaison de proportions pour 2 échantillons indépendants.

6.6 Tests non paramétriques.

6.6.1 Test des signes et test de Wilcoxon.

6.6.2 Test du chi-deux d'adéquation à une loi donnée.

6.6.3 Test du chi-deux d'indépendance.

Annexe: rappels et compléments de probabilités

Chapitre 1. EXEMPLES INTRODUCTIFS ET GÉNÉRALITÉS

1.1 Exemples introductifs

Exemple N° 1.

Un industriel vient de recevoir un lot important de pièces. Afin d'avoir une idée de la proportion, notée θ , de pièces défectueuses celui-ci réalise un tirage aléatoire avec remise de n pièces au sein du lot (un examen exhaustif du lot étant trop coûteux). L'échantillon est extrait du lot selon le schéma de Bernoulli. Autrement dit, si on note X_i la variable aléatoire à valeurs dans $\{0, 1\}$ associée au tirage numéro i , on suppose que les n variables aléatoires $X_1, \dots, X_i, \dots, X_n$ sont indépendantes et de même loi; à savoir $P(X_i = 1) = \theta$ et $P(X_i = 0) = 1 - \theta$ (l'évènement ' $X_i = 1$ ' signifie donc que la pièce i est défectueuse). Le problème statistique est d'estimer la proportion inconnue θ à partir des observations, c'est à dire à partir des x_i (réalisations des variables aléatoires X_i).

Exemple N° 2.

On souhaite savoir si une pièce de monnaie est équilibrée ou non. La pièce est dite équilibrée si, quand on la lance, la probabilité d'avoir 'pile' est égale à la probabilité d'avoir 'face'. Si on note θ la probabilité d'avoir 'pile', la pièce est équilibrée si $\theta = \frac{1}{2}$. Pour répondre à la question posée on lance $n = 10$ fois la pièce. Le résultat de cette expérience aléatoire est: PPPFFPFPP (P=pile et F=face). Peut-on conclure sur la base de ce résultat que la pièce est équilibrée ?

Exemple N° 3.

Une entreprise E fabrique et commercialise des ampoules électriques de 100 Watts. Sur l'emballage de ces ampoules l'entreprise E a porté la mention: durée d'utilisation $T = 1000$ heures. Un organisme de défense du consommateur se propose de 'valider' la mention

portée par ce constructeur sur ces emballages. Il souhaite de plus pouvoir décider si la durée de vie (moyenne) d'une ampoule de 100 Watts fabriquée par l'entreprise E est supérieure ou égale T . Pour ce faire, il teste un grand nombre n d'ampoules fabriquées par cette entreprise, et note x_i la durée de vie observée de l'ampoule No i . La durée de vie d'une ampoule électrique étant 'par nature' une grandeur imprévisible - au sens où la valeur de x_i ne peut, avant de l'avoir observée, être prédite avec certitude - il est commode de la modéliser par une variable aléatoire X . La loi de probabilité généralement retenue pour X est la loi exponentielle. Rappelons qu'une variable aléatoire réelle X suit une loi exponentielle de paramètre θ si sa densité est:

$$f(x) = \frac{1}{\theta} \exp\left[-\frac{x}{\theta}\right] \mathbb{I}_{[0,+\infty[}(x).$$

On peut aisément vérifier que le paramètre θ de la loi exponentielle est égal à la durée de vie moyenne d'une ampoule; autrement dit, $E(X) = \theta$. Le problème statistique consiste: premièrement, à estimer le paramètre inconnu θ à partir des observations, c'est à dire à partir des x_i (réalisations des variables aléatoires X_i); et deuxièmement, à décider si $\theta \geq T$, ou si, à l'inverse $\theta > T$.

Commentaires

Ces trois exemples ont en commun de faire reposer la réponse au problème statistique considéré, soit sur une expérience aléatoire (cf exemples 1 et 2), soit une modélisation aléatoire de la grandeur étudiée (cf exemple 3). Ils ont également en commun de baser l'inférence sur le paramètre θ (qu'il s'agisse d'estimer θ ou de tester une valeur particulière de θ) sur n observations $x_1, \dots, x_i, \dots, x_n$ où x_i désigne la réalisation d'une variable aléatoire réelle X_i (dont la loi de probabilité est paramétrée par θ).

1.2 Généralités

Afin de préciser l'objet et la place de la Statistique inférentielle (au sein de la Statistique en

général) il convient de rappeler que le travail du Statisticien s'inscrit traditionnellement dans l'une des trois phases suivantes: production de données, exploration des données, et modélisation (aléatoire) des données.

La production de données inclut en particulier l'organisation de la collecte des données (conception de plans de sondage, par exemple). L'exploration des données repose essentiellement sur les méthodes de la Statistique descriptive et de l'Analyse des données.

La modélisation comporte classiquement plusieurs étapes:

- a) Au sein d'un ensemble de modèles probabilistes susceptibles d'avoir généré les observations, sélectionner l'un d'entre eux (cette étape étant réalisée à l'aide d'une procédure statistique de choix de modèles).
- b) Valider le modèle retenu.
- c) Estimer les paramètres du modèle retenu et éventuellement proposer des procédures de tests portant sur les paramètres du modèle.

Cette présentation présuppose qu'un modèle paramétrique a été retenu, mais d'autres alternatives existent. Dans ce cours on abordera essentiellement les questions relatives à la dernière étape de cette procédure.

La Statistique inférentielle (appelée encore Statistique mathématique ou Statistique inductive) consiste en un ensemble de concepts, de méthodes et d'outils élaborés en vue de réaliser chacune des étapes a) b) et c). Les outils mathématiques utilisés en Statistique inférentielle ou Statistique inductive) font largement appel à la Théorie des Probabilités, du simple fait que les observations sont considérées comme des réalisations de variables aléatoires.

Chapitre 2. MODÈLES STATISTIQUES

2.1 Le modèle statistique d'échantillonnage paramétrique.

Définition.

Soit X une variable aléatoire à valeurs dans un espace \mathcal{X} muni d'une tribu \mathcal{B} et de loi P_θ où $\theta \in \Theta \in \mathbb{R}^p$. La donnée de n variables aléatoires $X_1, \dots, X_i, \dots, X_n$ à valeurs également dans $(\mathcal{X}, \mathcal{B})$, indépendantes et de même loi que celle de X constitue ce qu'on appelle un modèle d'échantillonnage paramétrique. Il est noté $(\mathcal{X}, \mathcal{B}, P_\theta; \theta \in \Theta)^n$.

Commentaires.

- La famille de lois de probabilités $\{P_\theta; \theta \in \Theta\}$ à laquelle appartient la loi de X est supposée connue; seul le paramètre θ est inconnu.
- A noter que l'espace probabilisé sur lequel est défini la variable aléatoire X n'apparaît pas explicitement dans la définition. En fait, il n'est pas utile de l'introduire à partir du moment où l'on se donne la loi de X , à savoir P_θ .
- Les n variables aléatoires $X_1, \dots, X_i, \dots, X_n$ sont dites indépendantes et identiquement distribuées (en abrégé *i.i.d.*), et le vecteur aléatoire $(X_1, \dots, X_i, \dots, X_n)$ constitue ce qu'on appelle un échantillon *i.i.d.* de loi P_θ .
- Le vecteur $(x_1, \dots, x_i, \dots, x_n)$ s'appelle l'échantillon observé et l'espace \mathcal{X}^n est appelé l'espaces des observations.
- Du point de vue du praticien, envisager un modèle paramétrique pour un ensemble d'observations $x_1, \dots, x_i, \dots, x_n$ consiste à considérer les x_i comme des réalisations de variables aléatoires $X_1, \dots, X_i, \dots, X_n$ indépendantes et de même loi P_θ ; c'est en ce sens qu'il y a modélisation (et modèle).
- Le modèle statistique d'échantillonnage paramétrique $(\mathcal{X}, \mathcal{B}, P_\theta; \theta \in \Theta)^n$ est parfois noté $(\mathcal{X}^n, \mathcal{B}^{\otimes n}, P_\theta^{\otimes n}; \theta \in \Theta)$ où $P_\theta^{\otimes n}$ représente la probabilité produit des P_θ définie sur \mathcal{X}^n muni de la tribu produit $\mathcal{B}^{\otimes n}$.

- Dans le cadre de ce cours et sauf mention contraire, $\mathcal{X} \subseteq \mathbb{R}$ et X désignera donc une variable aléatoire réelle (discrète ou continue). D'autre part on supposera toujours que la variable aléatoire X possède une densité que l'on notera $f_X(x)$ (ou plus simplement $f(x)$ s'il n'y a pas d'ambiguïté); cette densité étant prise par rapport à la mesure de Lebesgue si X est une variable aléatoire continue ou par rapport à la mesure de comptage si X est une variable aléatoire discrète; autrement dit, dans le cas discret, $f(x) = P(X = x)$.

Notations.

Les vecteurs $(x_1, \dots, x_i, \dots, x_n)$ et $(X_1, \dots, X_i, \dots, X_n)$ seront notés par la suite \mathbf{x} et \mathbf{X} .

Exemples.

- Le modèle d'échantillonnage de l'exemple 1 (Section 1.1) s'écrit: $(\mathcal{X}, \mathcal{B}, b(\theta); \theta \in [0, 1])^n$ où $\mathcal{X} = \{0, 1\}$, où $\mathcal{B} = \mathcal{P}(\{0, 1\})$ désigne l'ensemble des parties de $\{0, 1\}$, et où $b(\theta)$ désigne la loi de Bernoulli de paramètre $\theta \in \Theta = [0, 1]$.

- Adopter le modèle d'échantillonnage pour le deuxième exemple revient à considérer que les n variables aléatoires $X_1, \dots, X_i, \dots, X_n$ sont *i.i.d.*. Si la variable aléatoire X (associé à un seul lancer) défini sur l'espace $\Omega = \{\text{Pile}, \text{Face}\}$ est telle que $X(\text{Pile})=1$ et $X(\text{Face})=0$ alors le modèle s'écrit alors comme précédemment $(\mathcal{X}, \mathcal{B}, b(\theta); \theta \in [0, 1])^n$ où $\mathcal{X} = \{0, 1\}$ et où $\mathcal{B} = \mathcal{P}(\{0, 1\})$

- Considérer le modèle d'échantillonnage $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(\theta); \theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+)^n$ où $\mathcal{B}(\mathbb{R})$ désigne la tribu des Boréliens de \mathbb{R} , et où $\mathcal{N}(\theta)$ désigne la loi normale de paramètre θ , revient à considérer un échantillon $X_1, \dots, X_i, \dots, X_n$ de loi $\mathcal{N}(\theta)$ (par défaut, quand on ne précise pas, l'échantillon est supposé être *i.i.d.*).

Remarques.

C'est essentiellement dans le cadre du modèle d'échantillonnage paramétrique qu'on abordera les problèmes d'estimation (cf Chapitre 3) et de tests (Chapitre 4). Le modèle d'échantillonnage paramétrique présente l'avantage d'être particulièrement simple à mani-

-puler du simple fait que l'expression de la vraisemblance y est immédiate. La vraisemblance (définie ci-dessous) constitue une grandeur clef de la statistique inférentielle (nous reviendrons en détails sur cette notion dans le Chapitre 3).

Définition. On appelle fonction de vraisemblance (ou plus simplement vraisemblance) la fonction $L: \Theta \rightarrow [0, 1]$ définie par

$$L(\theta; x_1, \dots, x_i, \dots, x_n) = f(x_1, \dots, x_i, \dots, x_n; \theta)$$

Dans le cadre du modèle d'échantillonnage

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

puisque les X_i sont *i.i.d.*. Il importe de noter que la vraisemblance est une fonction de θ (les x_i étant connus). Si le modèle d'échantillonnage a l'avantage de la simplicité, il présuppose, a contrario, un ensemble d'hypothèses probabilistes (les X_i sont indépendantes et suivent la loi P_θ) qu'il convient de valider en pratique (cf l'étape de validation du modèle).

2.3 Autres modèles statistiques.

Un cadre plus général que celui du modèle statistique d'échantillonnage paramétrique est celui dans lequel on se donne simplement une variable aléatoire à valeurs dans un espace (E, \mathcal{E}) et une famille \mathcal{P} de probabilités définie sur cet espace; le triplet $((E, \mathcal{E}); \mathcal{P})$ s'appelle alors un modèle statistique. Voici deux exemples de modèles statistiques (distincts du modèle statistique d'échantillonnage paramétrique).

Situation No 1.

On considère n variables $X_0, X_1, \dots, X_i, \dots, X_n$ à valeurs dans l'espace $\{1, 2\}$. On suppose que X_0 n'est pas aléatoire et qu'il est égal à 1. Par contre $X_1, \dots, X_i, \dots, X_n$ sont supposées aléatoires et sont telles que la loi conditionnelle de $X_i | x_{i-1}, \dots, x_1$ est égale à la loi conditionnelle de $X_i | x_{i-1}$ (avec $i \geq 1$). On note $\alpha = P(X_i = 1 | x_{i-1} = 1)$ et

$\beta = P(X_i = 2|x_{i-1} = 2)$. D'un point de vue statistique le problème consiste à estimer les paramètres α et β de ce modèle statistique sur la base des x_i .

Situation No 2.

On considère n variables aléatoires $X_1, \dots, X_i, \dots, X_n$ de densité f . On souhaite estimer f sur la base des x_i (réalisations des X_i supposées *i.i.d.*).

Commentaires.

Dans la situation No 1 on considère un modèle paramétrique (paramétré par $\theta = (\alpha, \beta)$) mais qui n'est pas un modèle d'échantillonnage puisque les X_i ne sont pas indépendants (les X_i formant une chaîne de Markov). Le problème statistique exposé dans la situation No 2 est typiquement un problème de statistique non paramétrique; le cadre étant celui du modèle d'échantillonnage puisque les X_i sont *iid*.

Chapitre 3. EXHAUSTIVITÉ ET INFORMATION DE FISHER

3.1 Un exemple introductif

Dans l'exemple No 1 (Section 1.1) il est clair que, s'agissant d'estimer θ et compte tenu des hypothèses probabilistes faites sur les X_i , la donnée du nombre de pièces défectueuses, qui est égale à $\sum_{i=1}^n x_i$, apporte autant d'information sur θ que l'échantillon $x_1, \dots, x_i, \dots, x_n$ tout entier. L'objet de ce chapitre est d'introduire des concepts permettant de formaliser cela.

3.2 Exhaustivité.

Définition 1.

On considère le modèle statistique d'échantillonnage $(\mathcal{X}, \mathcal{B}, P_\theta; \theta \in \Theta)^n$. On appelle statistique toute application mesurable T de \mathcal{X}^n dans un espace mesurable \mathcal{Y} .

Commentaires

Il convient de noter que T ne dépend pas de θ .

Exemples de statistique

$\mathcal{X} = \mathbb{R}$ et $\mathcal{Y} = \mathbb{R}$.

$$(x_1, \dots, x_i, \dots, x_n) \longmapsto T(x_1, \dots, x_i, \dots, x_n) = \bar{x}_n.$$

$$(x_1, \dots, x_i, \dots, x_n) \longmapsto T(x_1, \dots, x_i, \dots, x_n) = \sum_{i=1}^n x_i^2.$$

Par abus de langage on parlera des statistiques $T = \bar{X}_n$, $T = \sum_i^n X_i^2$, etc.

Définition 2.

On considère le modèle statistique d'échantillonnage $(\mathcal{X}, \mathcal{B}, P_\theta; \theta \in \Theta)^n$. Une statistique T est exhaustive si la loi conditionnelle de \mathbf{X} sachant $T(\mathbf{x}) = t$ ne dépend pas de θ .

Exemple

$\mathcal{X} = \{0, 1\}$ et P_θ est la loi de Bernoulli $b(\theta)$. La statistique $\sum_i^n X_i$ est exhaustive. On

vérifie en effet facilement que:

$$P(\mathbf{X} = \mathbf{x} | T = t) = \frac{\theta^t (1 - \theta)^{n-t}}{C_n^t \theta^t (1 - \theta)^{n-t}} = \frac{1}{C_n^t} \quad \text{si} \quad \sum_{i=1}^n x_i = t$$

et que $P(\mathbf{X} = \mathbf{x} | T = t) = 0$ sinon.

Le calcul de la loi conditionnelle est parfois fastidieux. Le théorème suivant (qui est particulièrement facile d'utilisation) permet d'éviter ce calcul.

Théorème de factorisation (admis).

Une statistique T est exhaustive si la densité $f(\mathbf{x})$ de \mathbf{X} peut se mettre sous la forme $g(\mathbf{x})h(T(\mathbf{x}), \theta)$ où g et h désignent des applications à valeurs dans \mathbb{R}^+ .

Exemple No 1

On considère le modèle d'échantillonnage dans lequel X suit une loi de Poisson paramétrée par θ . D'après le théorème de factorisation la statistique $T = \sum_{i=1}^n X_i$ est exhaustive puisque

$$f(\mathbf{x}; \theta) = \frac{1}{\prod_{i=1}^n x_i!} \theta^{T(\mathbf{x})} \exp[-n\theta]$$

factorise au travers de :

$$g(\mathbf{x}) = \frac{1}{\prod_{i=1}^n x_i!}$$

et de

$$h(T(\mathbf{x}, \theta)) = \theta^{T(\mathbf{x})} \exp[-n\theta]$$

Exemple No 2. On suppose que X suit loi normale $\mathcal{N}(\theta)$. Montrer que la statistique $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ est exhaustive pour $\theta = (\mu, \sigma^2)$ où $\mu = E(X)$ et $\sigma^2 = Var(X)$.

3.3 Information de Fisher

Soit X une variable aléatoire réelle de loi P_θ où $\theta \in \Theta$ et Θ désigne un ouvert de \mathbb{R} . On supposera dans cette section que les hypothèses suivantes sont satisfaites.

H₁: $\forall \theta \in \Theta, \forall x \in \mathcal{X}, f(x, \theta) > 0$.

H₂: $\forall \theta \in \Theta, \forall x \in \mathcal{X}, \frac{\partial}{\partial \theta} f(x, \theta)$ et $\frac{\partial^2}{\partial \theta^2} f(x, \theta)$ existent.

H₃: $\forall \theta \in \Theta, \forall B \in \mathcal{B}$ on peut dériver 2 fois $\int_B f(x, \theta) dx$ par rapport à θ sous le signe d'intégration; autrement dit on peut échanger les opérateurs d'intégration et de dérivation (à l'ordre 1 et 2 pour ce qui concerne la dérivation).

Remarques.

H₁, H₂ et H₃ sont vérifiées si P_θ appartient à la famille exponentielle qui comprend, entre autres: la loi normale, la loi binomiale, la loi de Poisson, et la loi gamma. Par contre, la loi uniforme sur $[0, \theta]$ ne les vérifie pas (cf TD). Rappelons que sa densité est donnée par $f(x, \theta) = \frac{1}{\theta} \mathbb{I}_{[0, \theta]}(x)$.

3.3.1 Définition

Soit le modèle $((\mathcal{X}, \mathcal{B}); P_\theta, \theta \in \Theta)$. La quantité

$$I(\theta) = E \left[\left(\frac{\partial}{\partial \theta} \log f(X, \theta) \right)^2 \right]$$

s'appelle l'information de Fisher au point θ .

A noter que l'espérance est prise par rapport à la loi P_θ de X .

3.3.2 Autres expressions de l'information de Fisher

On appelle score la quantité $S(x, \theta) = \frac{\partial}{\partial \theta} \log f(x, \theta)$. On a donc:

$$I(\theta) = E[S^2]$$

Proposition. Sous les hypothèses H₁, H₂ et H₃ on a:

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

L'intérêt de cette proposition est que cette expression de $I(\theta)$ est plus facile à manipuler que celle de la définition.

Exemple.

Si $X \sim b(\theta)$ on vérifie que $I(\theta) = \frac{1}{\theta(1-\theta)}$ en remarquant que $f(x, \theta) = \theta^x(1-\theta)^{1-x}$.

3.3.3 Propriétés de l'information de Fisher

Propriété 1.

$I(\theta) \geq 0$ pour tout θ

Propriété 2.

Soient X et Y deux *v.a.r.* à valeurs dans \mathcal{X} et \mathcal{Y} , indépendantes, et de loi respective P_θ et Q_θ . On note $I_{(X,Y)}(\theta)$, $I_X(\theta)$ et $I_Y(\theta)$ les informations de Fisher au point θ respectivement apportées par X , Y , et (X, Y) . Dans ces conditions on a:

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

Conséquence. Soit X une variable aléatoire à valeurs dans $(\mathcal{X}, \mathcal{B})$ de loi P_θ . L'information de Fisher (au point θ) fourni par le modèle d'échantillonnage $(\mathcal{X}, \mathcal{B}, P_\theta; \theta \in \Theta)^n$ vérifie:

$$I_{(X_1, \dots, X_i, \dots, X_n)}(\theta) = nI_X(\theta)$$

Propriété 3. (admise)

On considère le modèle d'échantillonnage $(\mathcal{X}, \mathcal{B}, P_\theta; \theta \in \Theta)^n$. Soit T une statistique définie sur \mathcal{X}^n à valeurs dans un espace mesurable \mathcal{Y} . On a l'équivalence suivante qui relie les notions d'exhaustivité et d'information de Fisher:

$$I_{(X_1, \dots, X_i, \dots, X_n)}(\theta) = I_T(\theta) \iff T \text{ est exhaustive;}$$

$I_T(\theta)$ représente l'information de Fisher au point θ dans le modèle $(\mathcal{Y}; P_\theta^T)$ où P_θ^T désigne la loi image de P_θ par T .

Cette propriété permet donc d'établir l'exhaustivité d'une statistique.

Exemple

$\mathcal{X} = \{0, 1\}$ et P_θ est la loi de Bernoulli $b(\theta)$. Montrons que la statistique $T = \sum_i^n X_i$ est

exhaustive. On a d'une part:

$$I_{(X_1, \dots, X_i, \dots, X_n)}(\theta) = nI_X(\theta) = \frac{n}{\theta(1-\theta)}.$$

D'autre part, la loi de $T = \sum_i^n X_i$ est la loi binomiale de paramètres (n, θ) . Comme elle a pour densité

$$f(t, \theta) = C_n^t \theta^t (1-\theta)^{n-t}; \quad t \in \{0, \dots, n\}$$

on en déduit que:

$$\frac{\partial^2}{\partial \theta^2} \log f(t, \theta) = - \left[\frac{t}{\theta^2} + \frac{n-t}{(1-\theta)^2} \right]$$

d'où $I_T(\theta) = \frac{n}{\theta(1-\theta)}$ puisque $E[T] = n\theta$ (cqfd).

Chapitre 4. L'ESTIMATION

Le cadre statistique est celui du modèle d'échantillonnage $(\mathcal{X}, \mathcal{B}, P_\theta; \theta \in \Theta)^n$ où, sauf mention contraire, $\mathcal{X} \subseteq \mathbb{R}$ et $\Theta \subseteq \mathbb{R}$. Le problème statistique considéré dans ce chapitre consiste à estimer le paramètre θ (ou plus généralement une fonction $h(\theta)$) à partir des observations $x_1, \dots, x_i, \dots, x_n$ (réalisations de l'échantillon $X_1, \dots, X_i, \dots, X_n$).

4.1 Définition d'un estimateur

Définition. On appelle estimateur de θ toute fonction mesurable T de \mathcal{X}^n dans Θ . Autrement dit:

$$\begin{aligned} T : \mathcal{X}^n &\longrightarrow \Theta \\ \mathbf{x} &\longmapsto T(\mathbf{x}) \end{aligned}$$

Commentaires

- $T(\mathbf{x}) = T(x_1, \dots, x_i, \dots, x_n)$ s'appelle l'estimation de θ .
- Il importe de faire la distinction entre l'estimateur de θ (qui est une variable aléatoire réelle) et l'estimation de θ qui est une grandeur numérique.
- La loi de l'estimateur T est la loi image de $P_{\mathbf{X}}$ par T , où $P_{\mathbf{X}}$ désigne la loi du vecteur aléatoire \mathbf{X} .
- On désignera souvent un estimateur quelconque de θ par le symbole $\hat{\theta}$ (ou par $\hat{\theta}_n$ pour rappeler que la taille de l'échantillon est n)

Exemple

L'estimateur, sans doute le plus utilisé en pratique, est l'estimateur du maximum de vraisemblance. Il est défini de la façon suivante:

$$\hat{\theta}_n(\mathbf{x}) = \text{Arg} \max_{\theta \in \Theta} L(\theta; \mathbf{x}).$$

Nous reviendrons en détails sur cet estimateur par la suite. Donnons simplement un exemple du calcul de cet estimateur. Si $X \sim \mathcal{N}(\theta, \sigma^2)$ où σ^2 est supposé connue, il est

facile de vérifier que $\hat{\theta}_n(\mathbf{x}) = \bar{x}_n$. Par abus d'écriture, on écrira $\hat{\theta}_n = \bar{X}$. On montre que $\hat{\theta}_n$ suit une loi $\mathcal{N}(\theta, \frac{\sigma^2}{n})$.

4.2 Propriétés d'un estimateur

4.2.1 Estimateur sans biais

Soit $\hat{\theta}$ un estimateur de θ .

Définition. Si $E[\hat{\theta}] = \theta$ alors l'estimateur $\hat{\theta}$ est dit sans biais.

Exemple

$\hat{\theta}_n = \bar{X}_n$ est un estimateur sans biais de $E[X]$.

Remarque

Plus généralement, soit T un estimateur de $h(\theta)$; T est sans biais si $E[T] = h(\theta)$

Définition. $\hat{\theta}$ est dit biaisé si $E[\hat{\theta}] \neq \theta$ et $b(\hat{\theta}) = E[\hat{\theta}] - \theta$ s'appelle le biais de $\hat{\theta}$.

Exemple

$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est un estimateur biaisé de σ^2 car $E[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2 \neq \sigma^2$.

Par contre $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est un estimateur sans biais de σ^2 (cf TD No 1).

Définition. Un estimateur $\hat{\theta}_n$ de θ est asymptotiquement sans biais si son biais tend vers zéro quand $n \rightarrow +\infty$.

Exemple

$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est un estimateur asymptotiquement sans biais de σ^2 .

4.2.2 Estimateur convergent

Définition. $\hat{\theta}_n$ est convergent s'il converge en probabilité vers θ (quand $n \rightarrow +\infty$).

Exemple

Considérons le modèle d'échantillonnage $((\mathcal{X}, \mathcal{B}); b(\theta); \theta \in [0, 1])^n$ où $\mathcal{X} = \{0, 1\}$ et $b(\theta)$ désigne la loi de Bernoulli de paramètre $\theta \in \Theta = [0, 1]$. Dans ce contexte, on considère l'estimateur $\hat{\theta}_n = \bar{X}_n$. D'après la loi faible des grands nombres $\hat{\theta}_n$ est convergent.

Propriété. Un estimateur $\hat{\theta}_n$ de θ asymptotiquement sans biais et dont la variance tend vers zéro est convergent.

4.3 Comparaison entre estimateurs

La relation d'ordre habituellement utilisée pour comparer deux estimateurs repose sur la notion de risque quadratique défini ci-dessous.

Définition. $R(\hat{\theta}) = E[\hat{\theta} - \theta]^2$ s'appelle le risque quadratique de $\hat{\theta}$ (ou erreur moyenne quadratique).

Remarque

$R(\hat{\theta}) = E[\hat{\theta} - \theta]^2$ s'interprète comme l'erreur (quadratique) moyenne (au sens probabiliste) que l'on commet en estimant θ par $\hat{\theta}$. A noter que l'espérance est prise sous la loi des X_i . A noter également que $R(\hat{\theta}) = Var[\hat{\theta}] + [b(\hat{\theta})]^2$ et que $R(\hat{\theta})$ dépend en général de θ . En tant que fonction de θ le risque quadratique sera également noté $\mathcal{R}(\theta)$.

Soient $\hat{\theta}_1$ et $\hat{\theta}_2$ deux estimateurs de θ . On dira que $\hat{\theta}_1$ est meilleur que $\hat{\theta}_2$ au sens du risque quadratique si $R(\hat{\theta}_1) \leq R(\hat{\theta}_2)$ ou de façon équivalente: $\mathcal{R}_1(\theta) \leq \mathcal{R}_2(\theta)$ pour tout $\theta \in \Theta$. A noter que cette relation n'est pas une relation d'ordre total. Par exemple $\hat{\theta}_1 = \bar{X}_n$ et $\hat{\theta}_2 = \frac{n\bar{X}_n + 1}{n+2}$ ne sont pas comparables au titre du risque quadratique. Le vérifiez en représentant le graphe des fonctions \mathcal{R}_1 et \mathcal{R}_2 . Il est naturel, une fois choisi un critère de comparaison entre estimateurs (ici le risque quadratique), de chercher s'il existe un estimateur optimal au sens de ce risque.

Définition. Un estimateur T de θ est dit optimal (au sens du risque quadratique) si $R(T) \leq R(\hat{\theta})$ pour tout estimateur $\hat{\theta}$ de θ .

Malheureusement, il n'existe pas (en général) d'estimateur optimal (au titre du risque quadratique). Par contre l'existence d'estimateur optimaux est assurée (sous réserve de certaines conditions de régularité) quand on se limite à l'ensemble des estimateurs sans

biais de θ . Il convient de noter que comparer deux estimateurs sans biais de θ (au titre du risque quadratique) revient à les comparer au titre de la variance.

4.3.1 La borne de Cramer-Rao

Dans toute cette section on se limite aux estimateurs sans biais de θ et l'on désigne par $\mathcal{B}_0(\theta)$ la classe de tels estimateurs. Dans toute cette section on dira que $\hat{\theta}_1$ est meilleur que $\hat{\theta}_2$ si $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$ et l'optimalité sera à entendre au sens de la variance.

Théorème. On suppose que Θ est un ouvert de \mathbb{R} . Sous les conditions de régularité H_1 , H_2 et H_3 (cf Section 3.2), on a, pour tout $\hat{\theta}_n \in \mathcal{B}_0(\theta)$:

$$Var[\hat{\theta}_n] \geq \frac{1}{nI(\theta)}$$

Cette inégalité s'appelle l'inégalité de Cramer-Rao et la quantité $\frac{1}{nI(\theta)}$ est notée $K_n(\theta)$.

Définition. $\hat{\theta}_n \in \mathcal{B}_0(\theta)$ est dit efficace si $Var[\hat{\theta}_n] = \frac{1}{nI(\theta)}$

Exemple

Si $X \sim b(\theta)$ il est facile de vérifier que $\hat{\theta}_n = \bar{X}_n$ est efficace puisque $I(\theta) = \frac{1}{\theta(1-\theta)}$ et que $Var[\hat{\theta}_n] = \frac{\theta(1-\theta)}{n}$.

Définition. $\hat{\theta}_n$ est asymptotiquement efficace si le rapport:

$$\frac{Var[\hat{\theta}_n]}{K_n(\theta)}$$

tend vers 1 quand $n \rightarrow +\infty$.

Commentaires et compléments

Il est clair qu'un estimateur efficace $\hat{\theta}_n = \bar{X}_n$ est optimal. Mais, a contrario, un estimateur optimal n'atteint pas nécessairement pas la borne de Cramer-Rao et n'est donc pas nécessairement efficace. La recherche des estimateurs optimaux sort du cadre de ce cours.

Signalons simplement que les estimateurs optimaux sont à rechercher parmi les estimateurs qui sont fonction d'une statistique exhaustive. Donnons enfin le résultat suivant qui montre qu'on améliore un estimateur en le conditionnant par une statistique exhaustive. Soit $T \in \mathcal{B}_0(\theta)$ et S une statistique exhaustive. L'estimateur $E[T|S]$ est sans biais, et est meilleur que T . Ce théorème s'appelle le théorème de Rao-Blackwell.

4.4 L'estimateur du maximum de vraisemblance

On rappelle que la vraisemblance d'un échantillon $x_1, \dots, x_i, \dots, x_n$ est donnée par:

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

4.4.1 Définition

L'estimateur du maximum de vraisemblance de θ est l'estimateur $\hat{\theta}_n$ défini par:

$$\hat{\theta}_n(\mathbf{x}) = \text{Arg max}_{\theta \in \Theta} L(\theta; \mathbf{x}).$$

où $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$.

4.4.2 Remarques

- L'estimation par maximum de vraisemblance consiste, dans le cas discret, à retenir la valeur de θ qui donne la probabilité maximale à l'échantillon observé.
- L'estimateur du maximum de vraisemblance est parfois noté $\hat{\theta}_{\text{ML}}$.
- L'estimateur du maximum de vraisemblance peut-être biaisé. Exemple. La loi de X est la loi uniforme sur $[0, \theta]$. Alors $\hat{\theta}_{\text{ML}} = \sup_i X_i$ et $E[\hat{\theta}_{\text{ML}}] = \frac{n}{n+1}\theta$ (cf TD No2).
- L'estimateur du maximum de vraisemblance peut ne pas être efficace.
- L'estimateur du maximum de vraisemblance peut ne pas être unique.

- Sous réserve d'existence des dérivées première et seconde (relativement à θ) de la fonction de vraisemblance, $\hat{\theta}_n(\mathbf{x})$ est solution de:

$$\frac{\partial}{\partial \theta} L(\theta, \mathbf{x}) = 0;$$

et est tel que:

$$\frac{\partial^2}{\partial \theta^2} L(\theta, \mathbf{x}) < 0$$

au point $\theta = \hat{\theta}_n(\mathbf{x})$.

Commentaire

Ces hypothèses de dérivabilité sont généralement vérifiées en pratique. Cependant la loi uniforme constitue une exception notable. Dans le cas de la loi uniforme, la recherche de l'E.M.V. doit se faire "à la main".

4.4.3 Propriétés de l'estimateur du maximum de vraisemblance

Dans toute cette Section l'estimateur du maximum de vraisemblance est noté $\hat{\theta}_n$. L'intérêt de l'estimateur $\hat{\theta}_n$ réside essentiellement dans ses propriétés asymptotiques (cad quand $n \rightarrow +\infty$). Commençons cependant par donner une propriété de $\hat{\theta}_{ML}$ à distance finie.

Théorème. Les conditions H_1 , H_2 et H_3 sont supposées remplies. S'il existe un estimateur efficace de θ alors il est égal à l'estimateur du maximum de vraisemblance.

Théorème. On fait les hypothèses suivantes.

- Θ est un ouvert de \mathbb{R} .
- l'application $\theta \mapsto P_\theta$ est injective (θ est dit identifiable).
- $\frac{\partial}{\partial \theta} f(x, \theta)$ existe $\forall x, \forall \theta$
- $\forall \theta, \forall x, f(x, \theta) > 0$
- à $n \geq 1$ fixé (mais quelconque) $\hat{\theta}_n$ est unique.

Alors sous ces hypothèses l'estimateur du maximum de vraisemblance est convergent.

Théorème. On suppose de plus que:

- $\frac{\partial^2}{\partial \theta^2} f(x, \theta)$ existe, et est continue en θ et en x (uniformément).

- $0 < I(\theta) < +\infty$

- $\forall A \in \mathcal{B}$ on peut dériver 2 fois $\int_A f(x, \theta) dx$ par rapport à θ .

Alors sous toutes ces hypothèses

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

4.5 L'estimateur des moments

En pratique cet estimateur est utilisé dès $\dim \Theta \geq 2$. On posera $\theta = (\theta_j; j = 1, \dots, J)$.

On note μ_k le moment théorique non centré d'ordre $k \geq 1$; on a donc $\mu_k = E[X^k]$. On

note M_k le moment empirique non centré d'ordre $k \geq 1$; on a donc $M_k = \frac{1}{n} \sum_{i=1}^n X_i^k$. La

méthode des moments exploite le fait que M_k est un estimateur convergent de μ_k

Le principe de la méthode est la suivante.

1. On choisit J moments théoriques. En pratique on choisit (par souci de simplicité) les J premiers moments. On exprime les μ_j en fonction des θ_j . On a donc J égalités du type:

$$\mu_j = g_j(\theta_1, \dots, \theta_j, \dots, \theta_J)$$

2. On résout les J équations obtenues (les inconnues étant les θ_j que l'on exprime en fonction des μ_j). On obtient ainsi J égalités du type:

$$\theta_j = h_j(\mu_1, \dots, \mu_j, \dots, \mu_J)$$

3. On en déduit de façon naturelle un estimateur θ_j en estimant μ_j par M_j . Autrement dit on prend:

$$\hat{\theta}_j = h_j(M_1, \dots, M_j, \dots, M_J) \quad (j = 1, \dots, J).$$

L'estimateur ainsi construit s'appelle l'estimateur obtenu par la méthode des moments. Si la fonction h est continue, cet estimateur est convergent (d'après le théorème de Slutsky).

Commentaire

la procédure ci-dessus peut être mise en oeuvre avec les moments centrés ou un mélange de moments centrés et non centrés.

Exemple: estimation des paramètres de la loi Beta.

Rappelons que $X \sim \text{Beta}(a, b)$ où $a > 0$ et $b > 0$ si sa densité s'écrit:

$$f(x) = \frac{1}{\text{B}(a, b)} x^{a-1}(1-x)^{b-1} \mathbb{I}_{[0,1]}(x)$$

On montre que:

$$E[X] = \frac{a}{a+b} \quad \text{et} \quad \text{Var}[X] = \frac{ab}{(a+b)^2(a+b+1)}$$

On vérifie facilement que:

$$a = E(X) \left[\frac{E(X)[1 - E(X)] - \text{Var}(X)}{\text{Var}(X)} \right] \quad \text{et} \quad b = [1 - E(X)] \left[\frac{E(X)[1 - E(X)] - \text{Var}(X)}{\text{Var}(X)} \right]$$

Les estimateurs de a et de b obtenues par la méthode des moments sont donc:

$$\hat{a} = \bar{X}_n \left[\frac{\bar{X}_n[1 - \bar{X}_n] - V_n^2}{V_n^2} \right] \quad \text{et} \quad \hat{b} = [1 - \bar{X}_n] \left[\frac{\bar{X}_n[1 - \bar{X}_n] - V_n^2}{V_n^2} \right]$$

où

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

désigne le moment empirique centré d'ordre 2.

Chapitre 5. L'ESTIMATION PAR INTERVALLE DE CONFIANCE

Le cadre statistique est celui du modèle d'échantillonnage $(\mathcal{X}, \mathcal{B}, P_\theta, \theta \in \Theta)^n$ où $\mathcal{X} \subseteq \mathbb{R}$ et $\Theta \subseteq \mathbb{R}$. Le problème statistique considéré dans ce chapitre consiste à proposer, non plus une estimation ponctuelle de θ , mais un intervalle (de nature aléatoire) basé sur l'échantillon $X_1, \dots, X_i, \dots, X_n$ et qui contient, avec une probabilité fixée à l'avance (typiquement 95%), le paramètre inconnu θ .

5.1 Définition, exemple et commentaires

Définition. Soit $\alpha \in]0, 1[$, on appelle intervalle de confiance pour le paramètre θ , de niveau de confiance $1 - \alpha$, l'intervalle $[T_1, T_2]$ tel que:

$$P([T_1, T_2] \ni \theta) = 1 - \alpha$$

où T_1 et T_2 désignent deux statistiques définies sur \mathcal{X}^n à valeurs dans Θ . On dira que $[T_1, T_2]$ est un intervalle de confiance de niveau asymptotique $1 - \alpha$, si $P([T_1, T_2] \ni \theta)$ tend vers $1 - \alpha$ quand la taille n de l'échantillon tend vers $+\infty$.

Exemple

Considérons le modèle d'échantillonnage $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(\theta, 1); \theta \in \mathbb{R})^n$ où $\mathcal{N}(\theta, 1)$ désigne la loi normale de moyenne θ . Partant de $\bar{X}_n \sim \mathcal{N}(\theta, \frac{1}{n})$ il est facile de vérifier que:

$$P\left(\left[\bar{X}_n - \frac{1}{\sqrt{n}}z_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{1}{\sqrt{n}}z_{1-\frac{\alpha}{2}}\right] \ni \theta\right) = 1 - \alpha$$

où $z_{1-\frac{\alpha}{2}}$ désigne le quantile d'ordre $1 - \alpha/2$ de la loi normale $\mathcal{N}(0, 1)$. Rappelons que le quantile (on dit aussi fractile) d'ordre a de la loi $\mathcal{N}(0, 1)$ est le réel noté z_a défini par $P(Z \leq z_a) = a$ où $Z \sim \mathcal{N}(0, 1)$. Donc

$$\left[\bar{X}_n - \frac{1}{\sqrt{n}}z_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{1}{\sqrt{n}}z_{1-\frac{\alpha}{2}}\right]$$

est un intervalle de confiance de niveau de confiance $1 - \alpha$, pour la moyenne d'une loi normale (de variance égale à 1). On écrira aussi:

$$P\left(\bar{X}_n - \frac{1}{\sqrt{n}}z_{1-\frac{\alpha}{2}} \leq \theta \leq \bar{X}_n + \frac{1}{\sqrt{n}}z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Commentaires

1) Les bornes T_1 et T_2 de l'intervalle de confiance sont des variables aléatoires, contrairement à celles de l'intervalle de confiance observé $[T_1(\mathbf{x}), T_2(\mathbf{x})]$. Ce dernier s'appelle, en langage courant, une fourchette. Il est important de réaliser qu'écrire:

$$P(\theta \in [T_1(\mathbf{x}), T_2(\mathbf{x})]) = 1 - \alpha$$

n'a aucun sens, puisque rien n'est aléatoire dans l'écriture $\theta \in [T_1(\mathbf{x}), T_2(\mathbf{x})]$. Pour néanmoins donner un sens au niveau de confiance $1 - \alpha$, quand on parle de l'intervalle de confiance observé, on peut toujours considérer que l'échantillon observé \mathbf{x} résulte d'un tirage au hasard dans une population virtuelle \mathcal{P} (de grande taille) d'échantillons similaires à \mathbf{x} (cad provenant comme \mathbf{x} d'un n -échantillonnage *iid* de loi P_θ). Dire que le niveau de confiance est égal à 0.95 (par exemple) revient alors à dire que la probabilité de 'tirer' un échantillon de \mathcal{P} qui contienne θ est proche de 0.95.

2) On peut également définir des intervalles de confiance de la forme $[0, T]$ ou $[T, 1]$ (par exemple pour le paramètre $\theta \in [0, 1]$ d'une loi binomiale, cf TD).

3) $l = T_2 - T_1$ s'appelle la longueur de l'intervalle de confiance observé. A α fixé, l'intervalle de confiance est d'autant meilleur que l est petit. Cette remarque débouche tout naturellement sur la recherche d'intervalles de confiances optimaux (cf Section suivante).

5.2 Construction d'un intervalle de confiance: quelques principes généraux

On commence d'abord par introduire la notion de fonction pivotale qui va s'avérer particulièrement utile pour la construction d'un intervalle de confiance.

Définition. On appelle fonction pivotale pour θ toute fonction des X_i et de θ dont la loi ne dépend pas de θ . Elle est dite asymptotiquement pivotale si c'est la loi limite qui ne dépend pas de θ .

L'intérêt de cette notion est illustré par la remarque qui suit. Soit $h(X_1, \dots, X_i, \dots, X_n, \theta)$ une fonction pivotale pour θ . Supposons qu'on puisse déterminer numériquement u_1 et u_2 tels que: $P(u_1 \leq h(X_1, \dots, X_i, \dots, X_n, \theta) \leq u_2) = 1 - \alpha$. A noter que u_1 et u_2 sont indépendants de θ puisque h est pivotale pour θ . Alors, en résolvant (en θ) la double inéquation

$$u_1 \leq h(X_1, \dots, X_i, \dots, X_n, \theta) \leq u_2 \quad (1)$$

de telle sorte que (1) soit équivalent à:

$$g_1(X_1, \dots, X_i, \dots, X_n) \leq \theta \leq g_2(X_1, \dots, X_i, \dots, X_n),$$

on en déduit immédiatement un intervalle de confiance pour θ .

Exemple No 1

$X \sim \mathcal{N}(\theta, \sigma^2)$ où $\mathcal{N}(\theta, \sigma^2)$ désigne la loi normale de moyenne θ et de variance σ^2 (supposée connue). La fonction $\frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}}$ est pivotale pour θ puisque:

$$\frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Exemple No 2

$X \sim \mathcal{N}(\mu, \theta)$ où μ est connue et $\theta = \sigma^2$. La fonction $\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est pivotale pour θ puisque:

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

Exemple No 3

$X \sim \text{Poisson}(\lambda)$. Rappelons que $E(X) = \text{Var}(X) = \lambda$. La fonction $\frac{\bar{X}_n - \lambda}{\sqrt{\frac{\lambda}{n}}}$ est asymptotique-

ment pivotale pour λ puisque, d'après le théorème central limite,

$$\frac{\bar{X}_n - \lambda}{\sqrt{\frac{\lambda}{n}}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1) \quad \text{quand } n \longrightarrow +\infty$$

Nous concluerons cette section en montrant sur un exemple comment construire effectivement un intervalle de confiance à partir de la fonction pivotale; puis nous énoncerons un résultat qui permet d'obtenir un intervalle de confiance optimal pour θ dans un cas particulier.

Soit $X \sim \mathcal{N}(\theta, \sigma^2)$ où σ^2 est supposée connue. Cherchons un intervalle de confiance pour θ . Partant de

$$\frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

on a:

$$P \left[z_{\alpha_1} \leq \frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}} \leq z_{1-\alpha_2} \right] = 1 - \alpha$$

où z_{α_1} et $z_{1-\alpha_2}$ désignent respectivement les quantiles d'ordre α_1 et $1 - \alpha_2$ de la loi $\mathcal{N}(0, 1)$ tels que $\alpha_1 + \alpha_2 = \alpha$. À noter que, pour des raisons de symétrie: $z_{\alpha_1} = -z_{1-\alpha_1}$. Il est clair que:

$$-z_{1-\alpha_1} \leq \frac{\bar{X}_n - \theta}{\sigma/\sqrt{n}} \leq z_{1-\alpha_2} \iff I = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha_2}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha_1} \right] \ni \theta$$

Donc I est un intervalle de confiance de niveau de confiance de $1 - \alpha$ pour θ . Sa longueur est $\frac{\sigma}{\sqrt{n}}(z_{1-\alpha_2} - z_{\alpha_1})$. Parmi tous ces intervalles de confiance peut-on en exhiber un qui soit meilleur que tous les autres (au sens où sa longueur serait minimale). La réponse est fournie par le résultat suivant.

Propriété. Soit X une variable aléatoire réelle de densité $f(x)$ symétrique par rapport à zéro et unimodale. Soit \mathcal{I} la classe des intervalles réels de type $[a, b]$ tels que $a < 0 < b$ et vérifiant $P(a \leq X \leq b) = 1 - \alpha$ où $\alpha \in]0, 1[$. L'intervalle de longueur minimale est celui

qui est symétrique par rapport à zéro, c'est à dire du type $[-t, t]$, où t est alors le quantile d'ordre $1 - \frac{\alpha}{2}$ de $f(x)$.

Comme la densité de la loi normale $\mathcal{N}(0, 1)$ est symétrique par rapport à zéro et unimodale on peut utiliser ce résultat. Ce qui conduit à prendre $z_{1-\alpha_2} = z_{1-\alpha_1}$ soit $\alpha_1 = \alpha_2 = \alpha/2$.

Par conséquent: c'est l'intervalle de confiance

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right]$$

qui sera retenu pour une moyenne d'une loi normale $\mathcal{N}(\theta, \sigma^2)$. Précisons que le quantile $z_{1-\frac{\alpha}{2}}$ est fourni par les tables statistiques.

5.3 Intervalles de confiance classiques

Dans un premier temps, nous construirons des intervalles de confiance pour la moyenne et la variance de la loi normale. Puis, dans un second temps, nous indiquerons comment obtenir des intervalles de confiance pour des lois 'quelconques' quand la taille n de l'échantillon est 'grand'.

5.3.1 La loi de X est normale $\mathcal{N}(\mu, \sigma^2)$

cas No 1: Intervalle de confiance pour la moyenne μ quand σ^2 est connue.

Ce cas a déjà été traité dans la section 5.2. Rappelons que l'intervalle de confiance de niveau de confiance $1 - \alpha$ pour μ est

$$\left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right]$$

cas No 2: Intervalle de confiance pour la moyenne μ quand σ^2 est inconnue.

Rappelons le résultat de probabilité suivant. Soient $Z \sim \mathcal{N}(0, 1)$ et $Y \sim \chi^2(p)$ alors, si Z et Y sont indépendantes, $T = \frac{Z}{\sqrt{Y/p}}$ suit une loi de Student à p degrés de liberté (cette loi est tabulée). Comme

$$Z = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

et que:

$$Y = (n - 1) \frac{S_n^2}{\sigma^2} \sim \chi^2(n - 1)$$

où

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

on en déduit que:

$$T = \frac{Z}{\sqrt{Y/(n - 1)}} = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

suit une loi de Student à $n - 1$ degrés de liberté. Par conséquent T est pivotale pour μ .

D'où l'intervalle de confiance de niveau de confiance $1 - \alpha$ pour μ :

$$\left[\bar{X}_n - \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}}, \bar{X}_n + \frac{S_n}{\sqrt{n}} t_{1-\frac{\alpha}{2}} \right]$$

où $t_{1-\frac{\alpha}{2}}$ désigne le fractile d'ordre $1 - \frac{\alpha}{2}$ d'une loi de Student à $n - 1$ degrés de liberté.

cas No 3: Intervalle de confiance pour la variance σ^2 quand μ est connue.

Rappelons le résultat suivant. On se donne n variables aléatoires $Z_1, \dots, Z_i, \dots, Z_n$ *i.i.d*

et de loi $\mathcal{N}(0, 1)$ alors $\sum_i Z_i^2 \sim \chi^2(n)$. Comme

$$\frac{\bar{X}_i - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

on en déduit que:

$$\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$$

et est pivotale pour σ^2 . On a donc:

$$n \frac{\tilde{S}_n^2}{\sigma^2} \sim \chi^2(n) \quad \text{si on pose} \quad \tilde{S}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

D'où l'intervalle de confiance de niveau de confiance $1 - \alpha$ pour le paramètre $\theta = \sigma^2$:

$$\left[n \frac{\tilde{S}_n^2}{\chi_2^2(n)}, n \frac{\tilde{S}_n^2}{\chi_1^2(n)} \right],$$

où $\chi_1^2(n)$ et $\chi_2^2(n)$ désignent respectivement le fractile d'ordre $1 - \alpha_2$ et α_1 d'une loi du $\chi^2(n)$ tels $\alpha_1 + \alpha_2 = \alpha$. En pratique on prend: $\alpha_1 = \alpha_2 = \alpha/2$.

cas No 4: Intervalle de confiance pour la variance σ^2 quand μ est inconnue.

Rappelons que:

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

et est pivotale pour σ^2 . D'où l'intervalle de confiance de niveau de confiance $1 - \alpha$ pour le paramètre σ^2 :

$$\left[(n-1) \frac{S_n^2}{\chi_2^2(n-1)}, (n-1) \frac{S_n^2}{\chi_1^2(n-1)} \right]$$

où $\chi_2^2(n-1)$ et $\chi_1^2(n-1)$ désignent respectivement le fractile d'ordre $1 - \alpha_2$ et α_1 d'une loi du $\chi^2(n-1)$ tels que $\alpha_1 + \alpha_2 = \alpha$. En pratique on prend à nouveau $\alpha_1 = \alpha_2 = \alpha/2$.

5.3.2 Le cas des grands échantillons

Des intervalles de confiance asymptotiques peuvent être obtenus, par exemple, en appliquant le théorème central limite, ou en utilisant la loi limite de l'estimateur du maximum de vraisemblance, car cette façon de procéder est susceptible de fournir une fonction asymptotiquement pivotale pour θ .

L'utilisation du théorème central limite.

Nous avons déjà donné un exemple d'utilisation du théorème central limite pour obtenir une fonction asymptotiquement pivotale pour θ (cf l'exemple de la loi de Poisson; Section 5.1). Donnons un deuxième exemple.

Intervalle de confiance pour une proportion θ

X suit une loi de Bernoulli de paramètre θ . D'après le théorème central limite on a:

$$\frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1) \quad \text{quand } n \longrightarrow +\infty$$

puisque $E(X) = \theta$ et $\text{Var}(X) = \theta(1 - \theta)$. Donc

$$\frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}$$

est asymptotiquement pivotale pour θ Partant de:

$$P\left(\left[-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \leq z_{1-\frac{\alpha}{2}}\right]\right) \longrightarrow 1 - \alpha \quad \text{quand } n \longrightarrow +\infty$$

on obtient un intervalle de confiance de niveau de confiance asymptotique $1 - \alpha$ en résolvant (en θ) la double inéquation:

$$-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \leq z_{1-\frac{\alpha}{2}}$$

où $z_{1-\frac{\alpha}{2}}$ désigne le quantile d'ordre $1 - \alpha/2$ de la loi normale $\mathcal{N}(0, 1)$ (cf TD).

Une autre façon (plus rapide) de procéder consiste à remarquer que:

$$\frac{\bar{X}_n - \theta}{\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1) \quad \text{quand } n \longrightarrow +\infty.$$

On en déduit l'intervalle de confiance de niveau de confiance asymptotique $1 - \alpha$

$$\left[\bar{X}_n - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \right].$$

Remarque. Ces intervalles de confiance de niveau de confiance asymptotique $1 - \alpha$ sont en pratique utilisés dès que $n \geq 100$. Sinon on utilise les tables statistiques qui fournissent des intervalles de confiance reposant sur la statistique exhaustive $\sum_{i=1}^n X_i$ (cf TD).

L'utilisation de la loi limite de l'estimateur du maximum de vraisemblance.

Si l'estimateur du maximum de vraisemblance est noté $\hat{\theta}_n$, rappelons que:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\text{loi}} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$$

Donnons un exemple. On suppose que X qui suit une loi de Poisson de paramètre θ . On montre alors facilement que $\hat{\theta}_n = \bar{X}_n$ et que $I(\theta) = 1/\theta$. On a donc:

$$\sqrt{n} \frac{\bar{X}_n - \theta}{\sqrt{\theta}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1).$$

Puis on construit l'intervalle de confiance asymptotique comme dans l'exemple précédent. A noter que, au cas particulier, les utilisations du théorème central limite et de la loi limite de l'estimateur du maximum de vraisemblance fournissent le même intervalle de confiance asymptotique.

Chapitre 6. LES TESTS

6.1 Exemples introductifs

Dans le chapitre 1 du cours, différentes situations ont déjà été considérées pour introduire la notion de test (cf exemples 2 et 3). Donnons ici un autre exemple.

- Dans une assemblée de 100 personnes on demande à chacun de donner un chiffre au hasard compris entre 0 et 9. On note $x_i \in \{0, \dots, 9\}$ le chiffre donné par l'individu i et n_j le nombre d'individus ayant donné le chiffre j . Les résultats (c'est à dire l'ensemble des (j, n_j) où $j = 0, \dots, 9$) sont les suivants:

$$(0, 10), (1, 8), (2, 9), (3, 14), (4, 8), (5, 9), (6, 11), (7, 9), (8, 12), (9, 10).$$

Peut-on considérer que ces chiffres ont été effectivement donnés au hasard, au sens où les x_i sont des réalisations de variables aléatoires *i.i.d.* distribuées selon une loi uniforme sur $\{0, \dots, 9\}$?

- Dans les exemples 2 et 3 du chapitre 1, on cherche à tester une valeur particulière du paramètre θ ; la loi de X étant supposée connue (seul θ est inconnu). On parle alors de test paramétrique. Ainsi dans l'exemple 2 du chapitre 1, X suit une loi de Bernoulli de paramètre θ , et la question d'intérêt est: la pièce de monnaie est-elle équilibrée ou pas ?

On cherche donc à tester la valeur $\theta = 1/2$.

- Par contre, dans ce nouvel exemple, la loi de X n'est pas supposée connue. On parle alors de test non paramétrique. Plus précisément, il s'agit dans cet exemple d'un test d'adéquation (ou d'ajustement) à une loi donnée.

- Les sections 6.1 à 6.5 seront consacrées aux tests paramétriques, et les dernières sections aux tests non paramétriques.

6.2 la problématique d'un test

On se place dans le cadre du modèle d'échantillonnage où X suit la loi P_θ . On dispose donc d'un échantillon $X_1, \dots, X_i, \dots, X_n$ où les X_i sont *i.i.d.* et de loi P_θ .

Le contexte est le suivant. On se donne une partition de Θ en Θ_0 et Θ_1 . On doit, au vu des x_i , choisir entre deux hypothèses, $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_1$. L'hypothèse H_0 s'appelle l'hypothèse nulle, et l'hypothèse H_1 s'appelle l'hypothèse alternative (ou plus simplement alternative). On dit qu'on teste H_0 contre H_1 . On écrit:

$$H_0 : \theta \in \Theta_0 \quad \text{contre} \quad H_1 : \theta \in \Theta_1.$$

- Le terme d'hypothèse nulle vient du fait que l'hypothèse H_0 correspond souvent en pratique à tester l'absence d'effet d'un facteur (ou d'une covariable) sur la grandeur étudiée, par exemple, l'absence d'effet d'un traitement, l'absence d'effet du facteur âge, du facteur sexe, etc; cette absence d'effet du facteur correspondant à une hypothèse du type $h(\theta) = 0$, où $h(\theta)$ représente une fonction affine de θ (qui désigne ici un paramètre multidimensionnel); d'où la terminologie adoptée.

- Si une hypothèse (H_0 ou H_1) est réduite à un singleton, on parle d'hypothèse simple, sinon on parle d'hypothèse composée. Exemple. On souhaite tester:

$$H_0 : \theta = \frac{1}{2} \quad \text{contre} \quad H_1 : \theta \neq \frac{1}{2}$$

comme dans l'exemple sur la pièce de monnaie. Ici H_0 est simple et H_1 est composée.

- Les hypothèses H_0 et H_1 sont exclusives l'une de l'autre. Soit c'est H_0 qui est vraie, soit c'est H_1 qui est vraie. A l'issue d'un test on doit prendre une décision sur la base des x_i ; c'est à dire choisir entre H_0 et H_1 . L'espace des décisions possibles est noté \mathcal{D} et est réduit à deux éléments 0 et 1 (avec des notations évidentes).

Définition 1. H_0 et H_1 étant spécifiées, on appelle test (ou fonction de test) une application mesurable ϕ de \mathcal{X}^n dans \mathcal{D} .

remarque.

ϕ induit une partition de \mathcal{X}^n en deux sous-ensembles: $\phi^{-1}\{1\}$ et $\phi^{-1}\{0\}$.

Définition 2. $W = \phi^{-1}\{1\} = \{\mathbf{x} \in \mathcal{X}^n \mid \phi(\mathbf{x}) = 1\}$ s'appelle la région critique du test ϕ .

remarques.

- W est donc l'ensemble des \mathbf{x} pour lesquels on décide de choisir H_1 .
- Si $\mathbf{x} \in W$ on dit qu'on rejette H_0 (pour accepter H_1). Si $\mathbf{x} \in \overline{W}$ on dit qu'on garde (ou qu'on conserve) H_0 . Cette dissymétrie "en faveur" de H_0 au niveau de la terminologie sera justifiée ultérieurement.
- W est parfois noté R_c .
- W s'appelle aussi la région de rejet (sous-entendu de l'hypothèse nulle) du test ϕ .
- En pratique, on identifie ϕ et W .

Rappelons qu'à l'issue d'un test on doit prendre une décision au vu des x_i ; c'est à dire choisir entre H_0 et H_1 . Il y a donc deux mauvaises décisions possibles. Soit on choisit H_1 alors que c'est H_0 qui est vraie, et on parle alors d'erreur de première espèce; soit on choisit H_0 alors que c'est H_1 qui est vraie, et on parle alors d'erreur de seconde espèce.

Définition 3. Soit ϕ un test et W sa région critique. On appelle: risque de première espèce du test ϕ , la probabilité notée $\alpha_\phi(\theta)$ de rejeter à tort H_0 ; risque de seconde espèce du test ϕ , la probabilité notée $\beta_\phi(\theta)$ de garder à tort H_0 ; et enfin, puissance du test ϕ , la probabilité notée $\pi_\phi(\theta)$ de rejeter à juste titre H_0 . Autrement dit:

$$\alpha_\phi(\theta) = P(W|H_0) \quad \beta_\phi(\theta) = P(\overline{W}|H_1) \quad \pi_\phi(\theta) = P(W|H_1).$$

remarques.

- Il est important de signaler que α_ϕ , β_ϕ et π_ϕ sont en fait des fonctions de θ , et qu'à ce titre on parle parfois de fonctions de risque et de fonction puissance.

- Il est clair que $\pi_\phi = 1 - \beta_\phi$.
- On utilise fréquemment la notation suivante: $P_{H_0}(W)$ au lieu de $P(W|H_0)$. Même remarque pour β et π . Si Θ_0 est réduit au singleton θ_0 , on écrira: $P_{\theta_0}(W)$ pour indiquer que la probabilité est calculée sous (l'hypothèse) $\theta = \theta_0$. Idem pour β et π .
- Quand l'hypothèse nulle correspond à l'absence d'effet d'un facteur (ou d'une covariable) la puissance π représente la probabilité de mettre en évidence l'effet du facteur (c.a.d. accepter H_1) quand celui-ci existe effectivement (c.a.d. sous H_1).

Définition 4. On appelle niveau d'un test ϕ la quantité:

$$\alpha_\phi = \sup_{\theta \in \Theta_0} \alpha_\phi(\theta)$$

Cette quantité est souvent notée plus simplement α s'il n'y a pas d'ambiguïté.

6.3 L'approche de Neyman-Pearson

H_0 et H_1 étant spécifiées, la question se pose de savoir quel test ϕ choisir et comment départager deux tests ϕ_1 et ϕ_2 . Il est naturel de considérer que ϕ_2 est meilleur que ϕ_1 si:

$$\forall \theta \in \Theta_0, \quad \alpha_{\phi_2}(\theta) \leq \alpha_{\phi_1}(\theta) \quad \text{et} \quad \forall \theta \in \Theta_1, \quad \beta_{\phi_2}(\theta) \leq \beta_{\phi_1}(\theta)$$

Malheureusement, il n'existe pas (en général) de test ϕ optimal au sens de cette relation de préférence. Pour sortir de cette impasse, nous adopterons l'approche de Neyman-Pearson qui consiste à introduire une dissymétrie entre H_0 et H_1 . Cette approche préconise en effet de se fixer $\alpha \in]0, 1[$, puis de faire la recherche d'un test optimal au sein des tests de niveau α . Cette idée est précisée dans les définitions 5 et 6 ci-dessous.

Définition 5. Les hypothèses H_0 et H_1 étant spécifiées et α étant fixé, on se donne deux tests ϕ_1 et ϕ_2 de niveau α . On dira que ϕ_2 est meilleur que ϕ_1 si:

$$\forall \theta \in \Theta_1, \quad \beta_{\phi_2}(\theta) \leq \beta_{\phi_1}(\theta)$$

ou ce qui est équivalent:

$$\forall \theta \in \Theta_1, \quad \pi_{\phi_2}(\theta) \geq \pi_{\phi_1}(\theta)$$

Définition 6. Soit $\alpha \in]0, 1[$. On note $\mathcal{T}(\alpha)$ l'ensemble des tests de niveau α . H_0 et H_1 étant spécifiées, un test $\phi \in \mathcal{T}(\alpha)$ est dit uniformément plus puissant (en abrégé U.P.P.) s'il est optimal dans l'ensemble des tests $\mathcal{T}(\alpha)$, au sens de la relation de préférence définie ci-dessus. Autrement dit $\phi \in \mathcal{T}(\alpha)$ est U.P.P. si:

$$\forall \phi' \in \mathcal{T}(\alpha), \forall \theta \in \Theta_1, \pi_{\phi}(\theta) \geq \pi_{\phi'}(\theta)$$

Remarques.

- Un test ϕ de niveau α est donc U.P.P. s'il est le plus puissant des tests de niveau α .
- L'approche de Neyman-Pearson privilégie donc H_0 au sens où elle permet de contrôler le risque de première espèce, en fixant α ; typiquement $\alpha = 5\%$.

Un des objectifs de la suite du cours est la recherche de tests U.P.P. dans les cas usuels. Pour des raisons pédagogiques, on commencera par examiner le test qui met en jeu deux hypothèses simples.

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta = \theta_1 .$$

Puis on donnera quelques résultats concernant les autres tests usuels, à savoir:

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta > \theta_0 \quad (1)$$

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta < \theta_0 \quad (2)$$

$$H_0 : \theta = \theta_0 \quad \text{contre} \quad H_1 : \theta \neq \theta_0 \quad (3)$$

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \quad \text{contre} \quad H_1 : \theta < \theta_1 \quad \text{ou} \quad \theta > \theta_2 \quad (4)$$

$$H_0 : \theta \leq \theta_0 \quad \text{contre} \quad H_1 : \theta > \theta_0 \quad (5)$$

$$H_0 : \theta \geq \theta_0 \quad \text{contre} \quad H_1 : \theta < \theta_0 \quad (6)$$

$$H_0 : \theta \leq \theta_1 \quad \text{ou} \quad \theta \geq \theta_2 \quad \text{contre} \quad H_1 : \theta_1 < \theta < \theta_2 \quad (7).$$

6.4 Recherche du test uniformément plus puissant

6.4.1 Test du type: $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$.

Le théorème ci-dessous assure, dans un tel contexte, l'existence d'un test U.P.P.

Théorème de Neyman-Pearson. Soit X est une variable aléatoire continue de loi P_θ .

On se fixe $\alpha \in]0, 1[$. Le test ϕ de niveau α et de région critique:

$$W = \left\{ (x_1, \dots, x_i, \dots, x_n) \in \mathcal{X}^n ; \frac{L(x_1, \dots, x_i, \dots, x_n; \theta_0)}{L(x_1, \dots, x_i, \dots, x_n; \theta_1)} \leq k \right\}$$

est uniformément plus puissant (dans l'ensemble des tests de niveau α).

remarques.

- Puisque Θ_0 est réduit à θ_0 et Θ_1 à θ_1 on a : $\alpha = P_{\theta_0}(W)$ et $\pi = P_{\theta_1}(W)$.
- Le test U.P.P. ci-dessus est unique (α étant fixé).
- A noter que le théorème de Neyman-Pearson ne donne que la forme de la région critique. C'est la donnée de α qui permet de déterminer la constante k et donc d'obtenir W de façon complètement explicite.
- Si X est une variable aléatoire discrète l'existence d'un test ϕ U.P.P. de niveau α fixé n'est plus assurée (cf TD).

A titre d'exemple, on donne la région critique de test UPP dans le cas où la loi de X est normale.

Proposition. $X \sim \mathcal{N}(\mu, \sigma^2)$ où σ^2 est supposée connue. On considère le test:

$$H_0 : \mu = \mu_0 \quad \text{contre} \quad H_1 : \mu = \mu_1 .$$

On note z_a le fractile d'ordre a de la loi $\mathcal{N}(0, 1)$. La région critique du test ϕ uniformément plus puissant (de niveau α) est:

$$W = \left\{ (x_1, \dots, x_i, \dots, x_n) \in \mathcal{X}^n \mid \bar{x} \geq \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

si $\mu_1 > \mu_0$; la région critique de ϕ est:

$$W = \left\{ (x_1, \dots, x_i, \dots, x_n) \in \mathcal{X}^n \mid \bar{x} \leq \mu_0 - z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

si $\mu_1 < \mu_0$.

Proposition. $X \sim \mathcal{N}(\mu, \sigma^2)$ où μ est supposée connue. On note $\chi_a^2(n)$ le fractile d'ordre a de la loi du $\chi^2(n)$. On considère le test:

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{contre} \quad H_1 : \sigma^2 = \sigma_1^2.$$

La région critique du test ϕ uniformément plus puissant (de niveau α) est:

$$W = \left\{ (x_1, \dots, x_i, \dots, x_n) \in \mathcal{X}^n \mid \sum_{i=1}^n (x_i - \mu)^2 \geq \sigma_0^2 \chi_{1-\alpha}^2(n) \right\}$$

si $\sigma_1 > \sigma_0$; la région critique de ϕ est:

$$W = \left\{ (x_1, \dots, x_i, \dots, x_n) \in \mathcal{X}^n \mid \sum_{i=1}^n (x_i - \mu)^2 \leq \sigma_0^2 \chi_{\alpha}^2(n) \right\}$$

si $\sigma_1 < \sigma_0$.

remarques.

- Lorsque le test porte sur la moyenne μ , la région critique W est donc de la forme $\bar{x} > K$ (quand $\mu_1 > \mu_0$); cela n'est pas 'surprenant' compte tenu de l'alternative ($\mu_1 > \mu_0$) et du fait que \bar{X} est l'estimateur 'naturel' de $\mu = E[X]$.

- Lorsque le test porte sur σ^2 la région critique pourrait évidemment aussi s'écrire en faisant intervenir l'estimateur 'naturel' de σ^2 à savoir

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

Un point de vocabulaire.

La statistique sur laquelle repose un test s'appelle la *statistique de test*. Exemple: dans les deux dernières propositions, la statistique de test utilisée est \bar{X} (s'agissant du test portant sur μ), et $\sum_{i=1}^n (X_i - \mu)^2$ (s'agissant du test portant sur σ^2).

6.4.2 Tests du type (5), (6) et (7).

Définition 7. Soit X une variable aléatoire de loi P_θ . On considère le rapport

$$R = \frac{L(x_1, \dots, x_i, \dots, x_n; \theta')}{L(x_1, \dots, x_i, \dots, x_n; \theta)}$$

dans lequel on suppose que $\theta' > \theta$. La loi P_θ est dite à rapport de vraisemblance croissante (resp. décroissante) s'il existe une statistique T (c.a.d. une fonction des x_i) telle que R soit une fonction croissante (resp. décroissante) de $t = T(x_1, \dots, x_i, \dots, x_n)$.

Exemple.

On suppose que $X \sim \mathcal{N}(\theta, \sigma^2)$ où σ^2 est supposée connue. Il est immédiat de vérifier que:

$$R(t) = \exp \left[-\frac{1}{2\sigma^2} [2n(\theta - \theta')t + n(\theta'^2 - \theta^2)] \right]$$

est une fonction croissante de $t = T(x_1, \dots, x_i, \dots, x_n) = \bar{x}$ quand $\theta' > \theta$. Donc la loi de X est à rapport de vraisemblance croissante.

Théorème. Soit X une v.a.r. continue de loi P_θ à rapport de vraisemblance croissante.

On veut tester $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$. On se fixe $\alpha \in]0, 1[$. Le test ϕ de niveau α et de région critique:

$$W = \{(x_1, \dots, x_i, \dots, x_n) \in \mathcal{X}^n ; T(x_1, \dots, x_i, \dots, x_n) \geq k\},$$

où k est déterminé par $\alpha = P_{\theta_0}(W)$, est uniformément plus puissant (dans l'ensemble des tests de niveau α).

remarques.

- On a des résultats similaires si la loi P_θ est à rapport de vraisemblance décroissante. Par exemple, si on veut tester $H_0 : \theta \leq \theta_0$ contre $H_1 : \theta > \theta_0$, la région de rejet W est $T(x_1, \dots, x_i, \dots, x_n) \leq k$.

- On a aussi des résultats similaires si l'on veut tester: $H_0 : \theta \geq \theta_0$ contre $H_1 : \theta < \theta_0$. Si la loi P_θ est à rapport de vraisemblance croissante, la région de rejet W est $T(x_1, \dots, x_i, \dots, x_n) \leq k$.

Théorème. Soit X une v.a.r. continue de loi P_θ à rapport de vraisemblance croissante. On veut tester $H_0 : \theta \leq \theta_1$ ou $\theta \geq \theta_2$ contre $H_1 : \theta_1 < \theta < \theta_2$ (7). On se fixe $\alpha \in]0, 1[$. Le test ϕ de niveau α et de région critique:

$$W = \{(x_1, \dots, x_i, \dots, x_n) \in \mathcal{X}^n ; k_1 \leq T(x_1, \dots, x_i, \dots, x_n) \leq k_2\},$$

où k_1 est déterminé par $\alpha = P_{\theta_1}(W)$, et k_2 par $\alpha = P_{\theta_2}(W)$, est uniformément plus puissant (dans l'ensemble des tests de niveau α).

6.4.3 Tests du type (1), (2), (3) et (4).

- S'agissant des tests du type (1) et (2), il peut exister des tests UPP. C'est le cas si, quand on éprouve $H_0^* : \theta = \theta_0$ contre $H_1^* : \theta = \theta_1$, où θ_1 est fixé et appartient à $]\theta_0, +\infty[$ ou à $] - \infty, \theta_0[$ (selon le cas), la région critique (notée W^*) du test UPP (associée aux hypothèses H_0^* et H_1^*), ne dépend pas de θ_1 (on rappelle que W^* est donnée par le théorème de Neyman-Pearson). Il est facile de vérifier que, dans ce cas, le test qui a pour région critique W^* est UPP pour éprouver $H_0 : \theta = \theta_0$ contre $H_1 : \theta \in \Theta_1$, où Θ_1 désigne soit $]\theta_0, +\infty[$, soit $] - \infty, \theta_0[$.

Donnons un exemple. Supposons que $X \sim \mathcal{N}(\mu, \sigma^2)$ où σ^2 est supposée connue. On considère les deux hypothèses suivantes:

$$H_0 : \mu = \mu_0 \quad \text{et} \quad H_1 : \mu > \mu_0 .$$

D'après le résultat ci-dessus, le test qui a pour région critique:

$$W^* = \left\{ (x_1, \dots, x_i, \dots, x_n) \in \mathcal{X}^n \mid \bar{x} \geq \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

est UPP, puisque W^* ne dépend pas de μ_1 (le niveau α du test étant fixé).

- Il n'existe pas de tests U.P.P. pour les tests du type (3) et (4).

Compléments.

Quand il n'existe pas de test UPP, une stratégie possible consiste à réduire l'espace $\mathcal{T}(\alpha)$ de telle sorte que, dans ce nouvel espace, il existe des tests optimaux.

Définition. Un test ϕ de région critique W est dit sans biais au niveau α si:

- 1) $P_\theta(W) \leq \alpha$ pour tout $\theta \in \Theta_0$
- 2) $\alpha \leq P_\theta(W)$ pour tout $\theta \in \Theta_1$.

Autrement dit ϕ est sans biais si la probabilité de rejeter H_0 à tort est toujours inférieure à la probabilité de rejeter H_0 à juste titre (la première condition signifiant simplement que ϕ est de niveau α).

Pour chaque type de test (1), (2), (3) ou (4), on montre qu'il existe un test U.P.P. dans l'ensemble des tests sans biais de niveau α fixé. Un tel test est dit uniformément plus puissant sans biais; en abrégé U.P.P.S.B.

6.5 La pratique des tests.

Comme indiqué dans la section précédente, quand il n'existe pas de tests UPP dans l'ensemble des tests de niveau α fixé, on peut se restreindre aux tests sans biais. En pratique, on renonce le plus souvent à une quelconque optimalité et on base la région de rejet sur une statistique de test judicieusement choisie. L'objet de ce qui suit est de préciser ce dernier point et de décrire la procédure habituellement mise en place en pratique pour

construire un test paramétrique.

On se place comme d'habitude dans le cadre du modèle d'échantillonnage; on dispose donc d'un échantillon $X_1 \dots, X_i, \dots, X_n$ de loi P_θ ; le test porte sur θ . On suppose pour simplifier la présentation que l'hypothèse nulle H_0 est du type $\theta = \theta_0$. La mise en oeuvre d'un test comporte typiquement six étapes.

Etape 1. Expliciter mathématiquement les hypothèses H_0 et H_1 .

Etape 2. Se fixer le niveau α du test.

Etape 3. Choisir une statistique de test sur laquelle sera basée la région critique. La statistique de test, notée par la suite T_n , doit être liée de façon naturelle à θ ; typiquement on choisit pour T_n un estimateur raisonnable de θ (par exemple, l'e.m.v. de θ , ou l'estimateur de θ par la méthode des moments). Ainsi, si le test porte sur $E(X)$, un choix naturel pour T_n est \bar{X}_n . Si le test porte sur $\sigma^2 = Var(X)$, un choix naturel pour T_n est $\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ si $\mu = E(X)$ est connu. Il est important de réaliser que pour mettre en oeuvre un test de façon effective il faut connaître la loi de T_n sous H_0 , et si possible aussi sous H_1 (l'idéal étant que la loi de T_n soit tabulée). Ce dernier point est commenté en détails plus loin.

Etape 4. Déterminer la forme de la région critique W en utilisant la statistique de test T_n et en considérant également H_1 .

Etape 5. Expliciter complètement W en utilisant la définition de α . Puis calculer éventuellement la (fonction) puissance.

Etape 6. Conclure au vu de $\mathbf{x} = (x_1 \dots, x_i, \dots, x_n)$; c'est à dire rejeter H_0 si $\mathbf{x} \in W$ ou conserver H_0 dans le cas contraire.

Commentaires

1) Le fait de devoir connaître la loi de la statistique de test T_n sous H_0 (et si possible aussi sous H_1) fait qu'il faille parfois modifier le choix initial de T_n . On cherche une fonction

$h(\cdot)$ de T_n telle que la loi de $h(T_n)$ soit connue (sous H_0). On considère en pratique des fonctions h simples: par exemple linéaire ou affine. Donnons un exemple. Supposons que le test porte sur la variance σ^2 d'une loi normale $\mathcal{N}(\mu, \sigma^2)$ où μ est supposée connue, on prendra

$$n \frac{\widehat{\sigma^2}}{\sigma_0^2} = \sum_{i=1}^n \left[\frac{X_i - \mu}{\sigma_0} \right]^2$$

comme statistique de test, au lieu de $\widehat{\sigma^2}$. En effet, sous l'hypothèse nulle $H_0 : \sigma^2 = \sigma_0^2$, la loi de $n\widehat{\sigma^2}/\sigma_0^2$ est connue et tabulée: il s'agit d'une loi du chi-deux à n degrés de liberté.

2) Si n est grand, il est parfois intéressant de travailler avec la loi limite de T_n , pourvu que celle-ci soit connue (et tabulée). Introduisons à ce propos la notion de niveau asymptotique d'un test.

Définition.

On suppose que Θ_0 est réduit au singleton θ_0 . On dit qu'un test ϕ de région critique W est de niveau asymptotique α^* fixé si $P_{\theta_0}(W) \rightarrow \alpha^*$ quand $n \rightarrow +\infty$.

Donnons un exemple. Supposons que le test porte sur le paramètre θ de la loi de Bernoulli. Quand n est grand, on peut prendre comme statistique de test

$$T_n = \frac{\overline{X}_n - \theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$$

qui, sous l'hypothèse $H_0 : \theta = \theta_0$, converge en loi vers une loi $\mathcal{N}(0, 1)$ quand $n \rightarrow +\infty$.

3) Quand on met en oeuvre un test non paramétrique, on retrouve également ces 6 étapes; la seule différence est que la référence au paramètre θ (dans l'étape 3) n'a pas lieu d'être.

L'objet de la section 6.5 est de mettre en oeuvre cette procédure dans des situations classiques où le paramètre sur lequel porte le test est soit la moyenne, soit la variance.

ANNEXE

Dans cette annexe, on rappelle quelques notions de base du cours de probabilités; on y donne également quelques résultats de probabilités utiles en statistique inférentielle.

1. Définitions de base

Tribu

Soit Ω un espace non vide. On note $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω . Un sous-ensemble \mathcal{A} de $\mathcal{P}(\Omega)$ est une tribu sur Ω si:

- $\Omega \in \mathcal{A}$
- \mathcal{A} est stable par passage au complémentaire,
- \mathcal{A} est stable pour la réunion dénombrable.

Donnons un exemple de tribu. On prend $\Omega = \mathbb{R}$. La tribu engendrée par les intervalles ouverts de \mathbb{R} s'appelle la tribu des boréliens de \mathbb{R} . Elle est notée $\mathcal{B}(\mathbb{R})$.

Espace mesurable

Un espace Ω sur lequel on a défini une tribu \mathcal{A} s'appelle un espace mesurable.

Probabilité

Soit (Ω, \mathcal{A}) un espace mesurable. On appelle probabilité toute application P de \mathcal{A} vers $[0, 1]$ vérifiant: $P(\Omega) = 1$ et:

$$P\left(\bigcup_{k=1}^{+\infty} A_k\right) = \sum_{k=1}^{+\infty} P(A_k)$$

où les A_k sont dans \mathcal{A} et sont deux à deux disjoints.

Le triplet (Ω, \mathcal{A}, P) s'appelle un espace probabilisé.

Application mesurable

Soient (Ω, \mathcal{A}) et (E, \mathcal{B}) deux espaces mesurables. Une application f de Ω vers E est mesurable si: $f^{-1}(B) \in \mathcal{A}$ pour tout $B \in \mathcal{B}$. Exemple: toute application continue de \mathbb{R} vers \mathbb{R} muni de la tribu $\mathcal{B}(\mathbb{R})$ est mesurable.

Variable aléatoire.

Soient (Ω, \mathcal{A}, P) un espace probabilisé et (E, \mathcal{B}) un espace mesurable. Une application mesurable de Ω vers E s'appelle une variable aléatoire.

Soit (Ω, \mathcal{A}, P) un espace probabilisé. Une variable aléatoire définie sur Ω et à valeurs dans \mathbb{R} (muni de la tribu $\mathcal{B}(\mathbb{R})$) s'appelle une variable aléatoire réelle.

Loi d'une variable aléatoire.

Soient (Ω, \mathcal{A}, P) un espace probabilisé, (E, \mathcal{B}) un espace mesurable, et X une variable aléatoire de Ω dans E . La loi de probabilité de X est notée P_X . Elle est définie par: $P_X(B) = P(X^{-1}(B))$ où $B \in \mathcal{E}$. A noter que $X^{-1}(B) \in \mathcal{A}$ car X est une variable aléatoire et que $P_X(B)$ est donc bien défini.

La loi P_X s'appelle aussi la loi image de P par X .

2. Propriétés portant sur les moments d'une variable aléatoire réelle

X et Y désignent dans cette section deux variables aléatoires réelles définies sur le même espace probabilisé (Ω, \mathcal{A}, P) .

Espérance ou moyenne

On emploie aussi le terme *moyenne* pour parler de l'espérance de X .

L'espérance d'une variable aléatoire réelle X discrète à valeurs dans $E = X(\Omega)$ est:

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

L'espérance d'une variable aléatoire réelle X continue est:

$$E(X) = \int_{\mathbb{R}} x f(x) dx.$$

où $f(\cdot)$ désigne la densité de X .

A noter que:

$$E[h(X)] = \int_{\mathbb{R}} h(x) f(x) dx.$$

$$E(aX + bY) = aE(X) + bE(Y),$$

$E(a) = a$ si a désigne une constante réelle.

X et Y indépendantes $\implies E(XY) = E(X)E(Y)$ (la réciproque est fausse).

Variance et écart-type

$$\text{Var}(X) = E[(X - E(X))^2] \text{ et } \sigma(X) = \sqrt{\text{Var}(X)}.$$

$$\text{Var}(X) = E[X^2] - (E[X])^2.$$

$\text{Var}(a) = 0$ et $\text{Var}(aX) = a^2\text{Var}(X)$ (où a désigne une constante réelle).

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y).$$

X et Y indépendantes $\implies \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$. La réciproque est vraie si (X, Y) est un vecteur gaussien (cf Section 4).

Covariance

$$\text{Cov}(X, Y) = E[X - E(X)]E[Y - E(Y)]$$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y),$$

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)$$

X et Y indépendantes $\implies \text{Cov}(X, Y) = 0$. La réciproque est vraie si (X, Y) est un vecteur gaussien (cf Section 4).

Coefficient de corrélation

On appelle coefficient de corrélation linéaire de X et Y la quantité:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

on a les propriétés suivantes:

$$-1 \leq \rho(X, Y) \leq 1.$$

$|\rho(X, Y)| = 1 \iff Y$ est une fonction affine de X .

X et Y indépendantes $\implies \rho(X, Y) = 0$ (la réciproque est fausse en général). Si (X, Y) est un vecteur Gaussien alors la réciproque est vraie (cf Section 4).

3. Lois de probabilité utiles en statistique

Loi du χ^2

Soit $Z_1, \dots, Z_i, \dots, Z_n$ un échantillon *i.i.d.* de loi $\mathcal{N}(0, 1)$ alors:

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

suit par définition une loi du χ^2 à n degré de liberté. On note $X \sim \chi^2(n)$. La loi du χ^2 est tabulée. On montre que $E(Y) = n$ et que $Var(Y) = 2n$

Loi de Student

Soient Y et Z deux variables aléatoires indépendantes telles $Y \sim \chi^2(p)$ et $Z \sim \mathcal{N}(0, 1)$.

On appelle loi de de Student à n degrés de liberté la loi suivie par la variable aléatoire réelle

$$X = \frac{Z}{\sqrt{Y/n}};$$

on note $X \sim T(n)$. La loi de Student est tabulée.

Notation: $t_a(n)$ désigne le quantile d'ordre a d'une loi de de Student à n degrés de liberté.

- La densité de la loi de Student est une fonction symétrique; on a donc $t_{1-\alpha}(n) = -t_\alpha(n)$.

- On montre que: $T(n)$ converge en loi vers une $\mathcal{N}(0, 1)$ quand $n \rightarrow +\infty$.

Loi de Fisher-Snedecor

Soient Y_1 et Y_1 deux variables aléatoires réelles indépendantes telles que: $Y_1 \sim \chi^2(n_1)$ et

$Y_2 \sim \chi^2(n_2)$. On appelle loi de Fisher-Snedecor à n_1 et n_2 degrés de liberté la loi suivie

par la variable aléatoire réelle

$$X = \frac{Y_1/n_1}{Y_2/n_2};$$

on note $X \sim F(n_1, n_2)$. La loi de Fisher-Snedecor est tabulée.

- Si $X \sim T_n$ alors $X^2 \sim F(1, n)$.

- Notation: $f_a(p, q)$ désigne le quantile d'ordre a d'une loi $F_{p,q}$. On vérifie facilement que

$f_{1-\alpha}(p, q) = 1/f_\alpha(q, p)$.

4. Vecteurs gaussiens de \mathbb{R}^2

Définition

Un vecteur aléatoire de \mathbb{R}^2 est dit gaussien (ou normal) si toute combinaison linéaire de ces composantes suit une loi normale.

Propriétés

- On montre qu'un vecteur gaussien de \mathbb{R}^2 est entièrement déterminé par sa moyenne M et sa matrice de variance-covariance $\Gamma = [\gamma_{i,j}; i = 1, 2; j = 1, 2]$ où $\gamma_{i,j} = \text{Cov}(X_i, X_j)$.

On note $X \sim N_2(M, \Gamma)$. La loi de X s'appelle la loi normale bi-dimensionnelle.

- Soit X un vecteur gaussien de moyenne M et de matrice de variance-covariance Γ (qu'on suppose régulière). On montre que:

1) le vecteur aléatoire X a pour densité:

$$f(x_1, x_2) = \frac{1}{2\pi(\det\Gamma)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x - M) \Gamma^{-1} (x - M)'\right]$$

où $(x - M)'$ désigne la transposée de $(x - M)$.

2) la variable aléatoire réelle X_k suit une loi $\mathcal{N}(m_k, \gamma_{k,k})$ où X_k désigne la k -ième composante du vecteur X , et m_k la k -ième composante du vecteur M .

3) X_1 et X_2 indépendantes $\iff \text{Cov}(X_1, X_2) = 0$. A noter que seul l'implication \implies est toujours vraie (cf section 2).

5. Convergence d'une suite de variables aléatoires réelles

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles définies sur un espace probabilisé (Ω, \mathcal{A}, P) et X une variable aléatoire réelle définie sur le même espace.

Convergence en loi

La suite $(X_n)_{n \geq 1}$ converge en loi vers X si $F_n(x) \longrightarrow F(x)$ quand $n \rightarrow \infty$ (et ce en tout point x où F est continue). On écrit: $X_n \xrightarrow{\text{loi}} X$

Convergence en probabilité

La suite $(X_n)_{n \geq 1}$ converge en probabilité vers X si: $\forall \epsilon > 0, P(|X_n - X| < \epsilon) \rightarrow 1$ quand $n \rightarrow +\infty$. On écrit: $X_n \xrightarrow{P} X$

Quelques propriétés de convergence

- $X_n \xrightarrow{P} X \implies X_n \xrightarrow{\text{loi}} a$
- $X_n \xrightarrow{\text{loi}} a \implies X_n \xrightarrow{P} a \quad (a \in \mathbb{R})$
- Si $X_n \xrightarrow{\text{loi}} X$ et $Y_n \xrightarrow{P} a \quad (a \in \mathbb{R})$ alors: $X_n + Y_n \xrightarrow{\text{loi}} X + a, X_n Y_n \xrightarrow{\text{loi}} aX$, et $\frac{X_n}{Y_n} \xrightarrow{\text{loi}} \frac{X}{a}$ (avec $a \neq 0$).
- Soit une fonction continue de \mathbb{R} vers \mathbb{R} . Si $X_n \xrightarrow{P} X$ alors: $f(X_n) \xrightarrow{P} f(X)$. Si $X_n \xrightarrow{\text{loi}} X$ alors: $f(X_n) \xrightarrow{\text{loi}} f(X)$.
- Si $E(X_n) \rightarrow a \in \mathbb{R}$ et $Var(X_n) \rightarrow 0$ quand $n \rightarrow +\infty$ alors $X_n \xrightarrow{P} a$.

Théorème central limite et loi des grands nombres

- Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires réelles *i.i.d.* de moyenne μ et de variance σ^2 (finie), alors:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1) \quad \text{quand } n \rightarrow +\infty$$

Ce premier résultat s'appelle le théorème central limite.

- Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires réelles *i.i.d.* de moyenne μ (finie), alors:

$$\bar{X}_n \xrightarrow{P} E(X) \quad \text{quand } n \rightarrow +\infty$$

Ce second résultat s'appelle la loi (faible) des grands nombres.

6. Moments et lois de deux statistiques d'échantillonnage

Soit X une variable aléatoire réelle. $X_1, \dots, X_i, \dots, X_n$ désignent n variables aléatoires réelles indépendantes ayant pour loi celle de X . On se place donc dans le cadre du modèle

statistique d'échantillonnage . On suppose que les moments centrés d'ordre $k \leq 4$ de X existent. On note $\mu = E(X)$; $\sigma^2 = Var(X)$, et μ_k le moment centré d'ordre k (dès que $k \geq 3$). On introduit les statistiques d'échantillonnage suivantes:

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

La loi de X est quelconque.

$$E(\bar{X}_n) = \mu \text{ et } Var(\bar{X}_n) = \sigma^2/n.$$

$$E(S_n^2) = \sigma^2; \quad Var(S_n^2) = \frac{\mu_4}{n} - \frac{n-3}{n(n-1)}\sigma^4;$$

$$Cov(\bar{X}_n, S_n^2) = \frac{\mu_3}{n}$$

La loi de X est normale.

- Concernant les lois des statistiques \bar{X}_n et S_n^2 on a le résultat suivant:

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{et} \quad (n-1)\frac{S_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

- Les statistiques \bar{X}_n et S_n^2 sont indépendantes; par conséquent $Cov(\bar{X}_n, S_n^2) = 0$.

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