Préformation Statistiques

Luc Meyer

28, 29 et 30 août 2023

Content

Un premier exemple

- Un industriel souhaite évaluer la proportion de pièces défectueuses dans un lot.
- Il prélève au hasard 100 pièces et les teste : sur les 100 pièces testées, 10 sont défectueuses.
- Il en déduit la proportion de pièce défectueuse sur l'ensemble des pièces :

$$\hat{p} = \frac{10}{100} = 0.1. \tag{1}$$

Questions statistiques

Différentes questions peuvent se poser :

- Comment modéliser le problème ?
  -> Modélisation d'un problème statistique.
- Existe-t-il d'autres façons d'estimer p ? Si oui, quelle est la meilleure façon ?
  -> Estimation ponctuelle.
- Peut-on être sûr à 90% que la proportion de pièce défectueuses appartient à l'intervalle $[0.05; 0.15]$?
  -> Estimation par intervalles de confiance.

Contenu du cours

- Rappels de probabilités
- Premières définitions : Paramètre, Échantillon, Population, etc.
- Introduction à l'estimation paramétrique
- Estimation paramétrique : méthode des moments
- Estimation paramétrique : méthode du Maximum de Vraisemblance

Content

Pré-requis

Les notions suivantes sont censées être acquises (cf. Pré-formation en Probabilités) :

- Espace probabilisé
- Variables aléatoires réelles (discrètes et continues)
- Calculs de probabilités, moyennes et variance d'une variable aléatoire réelle, connaissant sa loi (discrète ou continu à densité)
- Théorème de transfert
- Différents mode de convergence des variables aléatoires
- Exemples classiques de lois discrètes : uniforme, Bernouilli, Binomiale, Poisson
- Exemples classiques de lois continues : uniforme, normale, exponentielle
- Moments d'ordre k (cf. partie du cours consacrée à l'estimation paramétrique par la méthode des moments)

Variables aléatoires réelles

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, c'est-à-dire (rappel) :

- Ω est un ensemble,
- $\mathcal{A}$ est une tribu de Ω (i.e. un sous-ensemble de $\mathcal{P}(\Omega)$ vérifiant certaines propriétés spécifiques),
- $\mathbb{P}$ est une probabilité (i.e. une mesure dont la masse totale vaut 1) sur $\mathcal{A}$.

On appelle *événement* les éléments de $\mathcal{A}$.

Definition (Rappel)

Une variable aléatoire réelle X est une application mesurable de $(\Omega, \mathcal{A}, \mathbb{P})$ dans $\mathbb{R}$.

Definition

La loi d'une variable aléatoire réelle X est la mesure image de $\mathbb{P}$ par X, c'est-à-dire, la mesure $\mathbb{P}_X$ sur $\mathbb{R}$ définie par :

$$\forall A \in \mathcal{P}(\mathbb{R}), \qquad \mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\omega \in \Omega : X(\omega) \in A) = \mathbb{P}(X \in A). \quad (2)$$

Fonction de répartition

Lemma (Rappel)

Pour identifier une mesure de probabilités $\mathbb{P}$ *sur* $\mathbb{R}$, *il faut et il suffit que l'on connaisse* $\mathbb{P}(]-\infty, x])$ *pour tout réel* x.

Definition

Soit X une variable aléatoire réelle (v.a.r.), la fonction :

$$\begin{array}{cccc} F_X : & \mathbb{R} & \to & [0;1] \\ & x & \mapsto & \mathbb{P}_X(]-\infty, x[) = \mathbb{P}(X \leq x) \end{array} \qquad (3)$$

s'appelle la *fonction de répartition* de X.

Fonction de répartition (cont.)

- Propriétés des fonctions de répartition

Theorem

Pour toute fonction de répartition F_X, on a :

- $\lim_{x\to-\infty} F_X(x) = 0$,
- $\lim_{x\to+\infty} F_X(x) = 1$,
- *F_X est croissante, continue à droite, i.e.,* $\lim_{y\to x, y>x} F_X(y) = F_X(x)$.

On pose

$$F_X(x^-) = \lim_{y\to x, y>x} F_X(y) = P(X < x), \quad (4)$$

et

$$\Delta F_X(x) = F_X(x) - F_X(x^-) = \mathbb{P}(X = x). \quad (5)$$

Densité d'une variable aléatoire

- Théorème-définition sur la densité de probabilité d'une variable aléatoire

Theorem

Lorsqu'elle exite, on définit la densité de la loi de X la fonction f_X telle que :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du. \tag{6}$$

(i.e. F_X est une fonction absolument continue). On dit alors que X est une variable aléatoire à densité.

Espérance, variance et moments

- Espérance

Definition

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire à valeurs dans $\mathbb{R}$. On appelle espérance de X, notée $E[X]$ la quantité :

$$E[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega). \tag{7}$$

Si X, est une variable aléatoire discrète, l'espérance s'écrit :
$E[X] = \sum_{x \in \Omega} xP(X = x)$.
Si X est une variable aléatoire réelle à densité, l'espérance s'écrit :
$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$.

Remarque : l'espérance est une application linéaire.

Espérance, variance et moments (cont.)

- Moment d'ordre p

Definition

Plus généralement, on définit le moment d'ordre $p \in \mathbb{N}$, s'il existe, par $E[X^p]$.

- Variance et écart-type

Definition

- La variance d'une variable aléatoire X est définie par :

$$Var(X) = E[(X - E[X])^2]. \tag{8}$$

- L'écart-type d'une variable aléatoire X est définie comme la racine carrée de la variance :

$$\sigma(X) = \sqrt{Var(X)}. \tag{9}$$

Remarque. On montre facilement que $Var(X) = E[X^2] - E[X]^2$, et que :

$$\forall a \in \mathbb{R}, \qquad Var(aX) = a^2 Var(X), \quad et \quad \sigma(aX) = a\sigma(X). \tag{10}$$

Théorème de transfert

- Théorème de transfert

Theorem

Soit X une variable aléatoire réelle (abrégée v.a.r dans ce cours), et $Y = \phi(X)$ une autre v.a.r. fonction de X, avec $\phi : \mathbb{R} \mapsto \mathbb{R}$. Alors :

$$E[Y] = E[\phi(X)] = \int_{\mathbb{R}} \phi(x) dP_X(x), \tag{11}$$

sous réserve d'existence de l'intégrale.
Dans le cas particulier où X et Y sont des v.a.r. continues à densité f_X, on a :

$$E[Y] = \int_{\mathbb{R}} y f_Y(y) dy = \int_{\mathbb{R}} \phi(x) f_X(x) dx. \tag{12}$$

Théorème de transfert (cont.)

- Application aux calculs des moments

Corollary

Si X est une variable aléatoire réelle à densité, son moment d'ordre $p \in \mathbb{N}$ est égal à :

$$E[X^p] = \int_{-\infty}^{+\infty} x^p f_X(x) dx. \quad (13)$$

Exemple. Soit X une v.a.r. uniforme sur $[0;1]$ et $Y = \phi(X) = X^3$. Alors :

$$E[Y] = E[X^3] = \int_{\mathbb{R}} \phi(x) \frac{1}{1-0} \mathbb{1}_{[0;1]} dx = \int_0^1 \phi(x) dx = \int_0^1 x^3 dx = \frac{1}{4}. \quad (14)$$

Probabilités conditionnelles et théorème de Bayes (Rappels)

Definition

Soient A et B deux événements avec B de probabilité non nulle. La probabilité conditionnelle de A sachant que B est réalisé est :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{15}$$

Theorem (Formule de Bayes)

Soient A et B deux événements avec B de probabilité non nulle. Alors :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}. \tag{16}$$

Probabilités conditionnelles et théorème de Bayes (Rappels) (cont.)

Theorem (Formule des probabilités totales)

Soit A un événement, et $(B_i)_{i\in I}$ un système complet d'événements (fini ou dénombrable), tous de probabilité non nulle. Alors :

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \cap B_i) = \sum_{i\in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i). \tag{17}$$

Content

Premières définitions

Definition

- Paramètre : caractéristique du modèle
- Population : ensemble d'individus
- Echantillon : sous-ensemble extrait d'une population
- Variables : caractéristiques de la population régies par un modèle partiellement connu.
- *n*-échantillon : vecteur $X = (X_1, ..., X_n)$ de n variables aléatoires (v.a.) indépendantes et identiquement distribuées (i.i.d.). En particulier tous les X_i suivent la même loi.
- Observation : réalisation d'une variable aléatoire. On notera x_i l'observation issue de la variable aléatoire X_i.
  Remarque : X_i est une variable aléatoire et x_i est un réel (la réalisation de X_i).

Cadre de l'estimation paramétrique

Hypothèses de l'estimation paramétrique :

- Les variables étudiées sur la population sont des variables aléatoires réelles (v.a.r.).
- Les n variables sont supposées issues de variables aléatoires réelles indépendantes et identiquement distribuées (i.i.d., i.e. distribuées suivant la même loi de probabilité).
- Dans le cadre de l'estimation paramétrique, on suppose que les variables suivent un modèle partiellement connu (cf. cours de probabilités) : la loi est supposée connue, mais un (ou plusieurs) paramètre(s) sont supposés inconnus.

Remark

Une autre approche (estimation non paramétrique, hors cadre de ce cours) consiste à considérer comme inconnue la loi suivie par les variables aléatoires.

Ainsi, un modèle paramétrique est un ensemble $\{\mathbb{P}_\theta ; \theta \in \Theta \subset \mathbb{R}^p\}$, où $\mathbb{P}_\theta$ est la loi de probabilité, θ est le paramètre, et Θ est l'ensemble des valeurs admissibles pour θ.

Exemples :

- $\{\mathbb{P}_\theta ; \theta \in \Theta\} = \{\mathcal{P}(\theta) ; \theta \in \mathbb{R}_+^*\}$, les observations sont issues de loi de Poisson,

Cadre de l'estimation paramétrique (cont.)

- $\{\mathbb{P}_\theta; \theta \in \Theta\} = \{\mathcal{U}_{[-\theta;\theta]}; \theta \in \mathbb{R}_+^*\}$, les observations sont issues d'une loi uniforme,
- $\{\mathbb{P}_\theta; \theta \in \Theta\} = \{\mathcal{N}(\theta_1, \theta_2); \theta = (\theta_1, \theta_2), \theta_1 \in \mathbb{R}, \theta_2 \in \mathbb{R}_+\}$, les observations sont issues d'une loi normale.
- $\{\mathbb{P}_\theta; \theta \in \Theta\} = \{\mathcal{B}(\theta); \theta \in [0, 1]\}$, les observations sont issues d'une loi de Bernoilli.
  Exemple : probabilité de tirer au sort une pièce défectueuse dans un lot.

L'objectif de l'estimation est donc de rechercher la valeur de $\theta \in \Theta$ qui correspond au mieux aux observations $(x_1, ..., x_n)$ données.

L'estimation paramétrique est elle même divisée en deux branches :

- Estimation ponctuelle : on cherche une estimation du paramètre.
  Exemple : $\hat{p} = 1.27$.
  -$>$ Cadre de ce cours.
- Estimation par intervalles de confiance : on cherche un intervalle dans lequel appartient le paramètre avec une probabilité donnée.
  Exemple : $p \in [1.25; 1.29]$ avec un risque d'erreur de $0.05 = 5\%$.
  -$>$ Non traité dans le cadre de ce cours.

Définition d'un estimateur

Définition

Soit $(X_1, ...X_n)$ un n-échantillon suivant une loi *connue* de paramètre $\theta \in \mathbb{R}$ *inconnu*.

- Un estimateur du paramètre θ est une variable aléatoire réelle (v.a.r.) $\hat{\theta}_n$ fonction du n-échantillon, i.e. il existe une fonction $F : \mathbb{R}^n \to \mathbb{R}$ avec $\hat{\theta}_n = F(X_1, ..., X_n)$.
  Exemple : $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (estimateur de la moyenne empirique du n-échantillon).
- Une estimation de θ est un réel fonction des réalisations $x_1, ..., x_n$ des v.a. $X_1, ..., X_n$. C'est donc une réalisation de $\hat{\theta}_n$.
  Exemple : $real(\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$ (estimation de la moyenne empirique du n-échantillon).

Définition d'un estimateur (cont.)

Exemples d'estimateurs pour la loi normale

Soit $(X_1, ..., X_n)$ un n-échantillon de loi parente $\mathcal{N}(\theta, 1)$, de moyenne θ inconnue et de variance connue égale à 1.

- $\hat{\theta}_n^{(1)} = X_1$,
- $\hat{\theta}_n^{(2)} = -4.7$,
- $\hat{\theta}_n^{(3)} = \frac{1}{n} \sum_{i=1}^{n} X_i$,
- $\hat{\theta}_n^{(4)} = \frac{1}{n} \sum_{i=1}^{n} X_i + \frac{1}{n^2}$.

Remarque

$\hat{\theta}_n = F(X_1, ..., X_n)$, où les X_1, ..., X_n sont les variables aléatoires réelles.
Ainsi, un estimateur $\hat{\theta}_n$ est une *variable aléatoire réelle*.

Objectif de l'estimation

Choisir la fonction F de façon à estimer *correctement* θ.

Consistance d'un estimateur

Définition

Un estimateur $\hat{\theta}_n$ est dit *consistant* pour la valeur θ (ou *convergent* vers θ) si $\hat{\theta}_n$ tend vers θ presque sûrement (par rapport à $\mathbb{P}_\theta(X_1, ..., X_n)$) quand n tend vers l'infini.

Remarque

La consistance est une propriété majeure d'un estimateur ponctuel. C'est elle qui donne du sens à l'approximation de θ par $\hat{\theta}_n$ pour n grand.

Biais et variance

Rappel : un estimateur est une variable aléatoire.

Ainsi, les définitions et propriétés classiques des variables aléatoires sont applicables.

Notations

Soit X une variable aléatoire. On note $\mathrm{E}[X]$ son espérance mathématique, et $Var(X)$ sa variance. Pour rappel, $Var(X) = \mathrm{E}[(\mathrm{E}[X] - X)^2] = \mathrm{E}[X^2] - \mathrm{E}[X]^2$.

Définition

Le *biais* d'un estimateur $\hat{\theta}_n$ de θ est :
$B(\hat{\theta}_n, \theta) = \mathrm{E}[\hat{\theta}_n] - \theta$.
Un estimateur $\hat{\theta}_n$ est dit *sans biais* si $B(\hat{\theta}_n, \theta) = 0$. Sinon, il est dit *biaisé*.

Propriété

Un estimateur sans biais et de variance asymptotiquement nulle est convergent (consistant).

Exemples

Reprenons les exemples précédents d'estimateurs pour la loi $\mathcal{N}(\theta, 1)$:

- $\hat{\theta}^{(1)} = X_1$,
  $\mathrm{E}[\hat{\theta}^{(1)}] = \mathrm{E}[X_1] = \theta \Rightarrow B(\hat{\theta}^{(1)}, \theta) = 0 \Rightarrow$ Estimateur sans biais.
  $Var(\hat{\theta}^{(1)}) = Var(X_1) = 1 \Rightarrow \lim\limits_{n\to\infty} Var(\hat{\theta}^{(1)}) = 1 \neq 0 \Rightarrow$ Variance asymptotiquement non nulle.
- $\hat{\theta}^{(2)} = -4.7$,
  $\mathrm{E}[\hat{\theta}^{(2)}] = -4.7 \Rightarrow B(\hat{\theta}^{(2)}, \theta) = -4.7 - \theta \Rightarrow$ Estimateur biaisé (sauf dans le cas particulier où $\theta = -4.7$).
  $Var(\hat{\theta}^{(2)}) = 0 \Rightarrow \lim\limits_{n\to\infty} Var(\hat{\theta}^{(2)}) = 0 \Rightarrow$ Variance asymptotiquement nulle.
- $\hat{\theta}^{(3)} = \frac{1}{n}\sum_{i=1}^{n} X_i$,
  $\mathrm{E}[\hat{\theta}^{(3)}] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n} \theta = \theta \Rightarrow B(\hat{\theta}^{(3)}, \theta) = 0 \Rightarrow$ Estimateur sans biais.
  $Var(\hat{\theta}^{(3)}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} 1 = \frac{1}{n} \Rightarrow \lim\limits_{n\to\infty} Var(\hat{\theta}^{(3)}) = 0 \Rightarrow$ Variance asymptotiquement nulle.

Exemples (cont.)

- $\hat{\theta}^{(4)} = \frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n^2}$,
  $\mathrm{E}[\hat{\theta}^{(4)}] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] + \frac{1}{n^2} = \frac{1}{n}\sum_{i=1}^{n}\theta + \frac{1}{n^2} = \theta + \frac{1}{n^2} => B(\hat{\theta}^{(4)}, \theta) = \frac{1}{n^2} =>$ Estimateur biaisé.
  $Var(\hat{\theta}^{(4)}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} 1 = \frac{1}{n} => \lim_{n\to\infty} Var(\hat{\theta}^{(4)}) = 0 =>$ Variance asymptotiquement nulle.

Estimateurs sans biais à minimum de variance

Définition

Un estimateur $\hat{\theta}_n$ est dit *sans biais à minimum de variance* si :

- $E[\hat{\theta}_n] = \theta, \qquad \forall \theta \in \Theta$,
- pour tout estimateur $\tilde{\theta}_n$ tel que "$E[\tilde{\theta}_n] = \theta, \forall \theta \in \Theta$", on a :

$$Var(\tilde{\theta}_n) \geq Var(\hat{\theta}_n). \tag{18}$$

Propriété

Si un tel estimateur existe, alors il est unique (presque sûrement).

Définition

Quand il existe, un tel estimateur est appelé estimateur *efficace*.

Risque Quadratique Moyen

Définition

Soit $\hat{\theta}_n$ un estimateur de $\theta \in \mathbb{R}$. On appelle *erreur quadratique moyenne* (ou *risque quadratique moyen*) la valeur : $EQM(\hat{\theta}_n, \theta) = \mathrm{E}[(\hat{\theta}_n - \theta)^2]$.

Propriété

$$EQM(\hat{\theta}_n, \theta) = Var(\hat{\theta}_n) + B(\hat{\theta}_n, \theta)^2 \tag{19}$$

Ainsi, pour un estimateur sans biais : $EQM(\hat{\theta}_n, \theta) = Var(\hat{\theta}_n)$.

Exercice : calculer le risque quadratique moyen des estimateurs précédemment proposés de la loi $\mathcal{N}(\theta, 1)$.

Risque Quadratique Moyen (cont.)

Définition

- Un estimateur $\hat{\theta}_n$ est dit *(quadratiquement) préférable* à un estimateur $\tilde{\theta}_n$ pour la valeur θ si $EQM(\hat{\theta}_n, \theta) \leq EQM(\tilde{\theta}_n, \theta)$.
- Un estimateur $\hat{\theta}_n$ est dit *(quadratiquement) uniformément préférable* à un estimateur $\tilde{\theta}_n$ s'il est quadratiquement préférable pour toute valeur de $\theta \in \Theta$.
- Un estimateur est dit admissible s'il n'existe aucun estimateur qui lui est préférable. Il est dit inadmissible dans le cas contraire.

Meilleurs estimateurs ?

Quel(s) critère(s) pour une estimation optimale ?

- (Rappel) Un bon estimateur est nécessairement consistant.
- Deux approches possibles (selon le critère à optimiser) :
  - si cela est possible, on cherche en priorité un estimateur (possiblement biaisé) à minimum d'erreur quadratique moyenne $=>$ idéal mais pas toujours possible,
  - estimateur sans biais à minimum de variance $=>$ plus facile à mettre en place en pratique.

Remarque

Il existe des estimateurs biaisés meilleurs (au sens de l'EQM) que des estimateur sans biais (cf. exemple suivant).

Meilleurs estimateurs ? (cont.)

Exemple

Soit un n-échantillon $(X_1, ..., X_n)$ suivant une loi de Bernouilli $\mathcal{B}(\theta)$. Pour $\theta = 1/2$, l'estimateur $\hat{\theta}_n^{(1)} = \frac{1}{n}\sum_{i=1}^n X_i$ est sans biais et son erreur quadratique est :

$$EQM(\hat{\theta}_n^{(1)}) = \frac{1}{4n}. \tag{20}$$

Or, l'estimateur $\hat{\theta}_n^{(2)} = \frac{n\hat{\theta}_n^{(1)} + \sqrt{n/4}}{n + \sqrt{n}}$ a pour risque quadratique :

$$EQM(\hat{\theta}_n^{(2)}) = \frac{n}{4(n + \sqrt{n})^2} < \frac{1}{4n}. \tag{21}$$

Fin de la séance 1

Dans la séance 2, nous verrons différentes méthodes de construction d'estimateurs :

1. Estimateurs empiriques de la moyenne et de la variance
2. Méthode des moments pour la construction d'estimateurs
3. Méthode du maximum de vraisemblance pour la construction d'estimateurs

Content

Estimateur de la moyenne empirique

Considérons un n-échantillon $(X_1, ..., X_n)$ suivant une loi de probabilité de moyenne m et de variance σ^2.

Définition

On appelle *estimateur de la moyenne empirique* du n-échantillon $(X_1, ..., X_n)$ la variable aléatoire :

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{22}$$

Propriété

L'estimateur $\overline{X}$ est un estimateur sans biais et consistant de la moyenne m de la loi de probabilité suivi par le n-échantillon.

Démonstration traitée en exercice.

Estimateur de la variance empirique

Considérons un n-échantillon $(X_1, ..., X_n)$ suivant une loi de probabilité de moyenne m et de variance σ^2.

Définition

On appelle *estimateur de la variance empirique* du n-échantillon $(X_1, ..., X_n)$ la variable aléatoire :

$$V_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j)^2. \tag{23}$$

Propriété

- L'espérance de l'estimateur V_n est $E[V_n] = \frac{n-1}{n}\sigma^2$. Ainsi l'estimateur V_n est un estimateur biaisé de σ^2.
- La variable aléatoire $\tilde{V}_n = \frac{n}{n-1}V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j)^2$

est un estimateur sans biais de la variance σ^2 du n-échantillon. Il est appelé *estimateur corrigé de la variance*.

Démonstration traitée en exercice.

Content

Estimateur des moments

Principe

Soit un n-échantillon $(X_1, ..., X_n)$ suivant une loi $\mathbb{P}_\theta$ de paramètre $\theta \in \Theta$ inconnu. L'estimateur des moments d'ordre $k \in \mathbb{N}^*$ $\hat{\theta}_n^{(k)}$ s'obtient en résolvant l'équation :

$$m_k(\hat{\theta}_n^{(k)}) = \hat{m}_k, \tag{24}$$

avec pour tout $k \geq 0$:

- $m_k(\theta) = E_\theta[X_i^k]$, $\forall i \in \{1, ..., n\}$ (v.a. i.i.d.), le moment théorique et
- $\hat{m}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$, le moment empirique.

Exemple

Exemple : soit un n-échantillon $(X_1, ..., X_n)$ suivant une loi de poisson $\mathcal{P}(\theta)$ de paramètre θ.

Calculons l'estimateur des moments de θ :

- A l'ordre 1 :

$$E_{\hat{\theta}_n^{(1)}}[X] = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{25}$$

  Or : $E_\theta[X] = \theta$
  Ainsi : $\hat{\theta}_n^{(1)} = \frac{1}{n}\sum_{i=1}^{n} X_i$ est un estimateur des moments de θ.

Propriété

Questions :

- Choix des moments à utiliser !
- Existence (et calcul) des solutions

Propriété

Si $E[|X|^k] < \infty$, et si $m_k(\theta) = f(\theta)$, avec f une fonction inversible continue, alors l'estimateur $\hat{\theta}_n = f^{-1}(\hat{m}_k)$, est un estimateur consistant de θ.

Extension

Soit $g : \mathbb{R} \to \mathbb{R}$, une fonction continue. Alors la méthode des moments se généralise par la résolution en θ de l'équation :

$$\mu(\theta) = \hat{\mu}, \tag{26}$$

avec :

- $\mu(\theta) = \mathrm{E}_\theta[g(X_1)]$,
- $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} g(X_i)$.

Propriété

Si $\mathrm{E}[|g(X)|] < \infty$, et si $\mu(\theta) = f(\theta)$, avec f une fonction inversible continue, alors l'estimateur $\hat{\theta}_n = f^{-1}(\hat{\mu})$, est un estimateur consistant de θ.

Content

Définition de la vraisemblance

Définition

Dans un modèle paramétrique dominé, la vraisemblance est une fonction de la variable θ définie pour toute réalisation $(x_1, ..., x_n)$ du n-échantillon $(X_1, ..., X_n)$ et qui associe à $\theta \in \Theta$ la valeur $f_\theta(x_1, ..., x_n)$:

$$\theta \mapsto L_n(\theta; x_1, ..., x_n) = f_\theta(x_1, ..., x_n) = \prod_{i=1}^{n} f_\theta(x_i). \tag{27}$$

Dans le cas discret, on a :

$$\theta \mapsto L_n(\theta; x_1, ..., x_n) = \mathbb{P}_\theta(x_1, ..., x_n) = \prod_{i=1}^{n} \mathbb{P}_\theta(X = x_i). \tag{28}$$

Les dernières égalités de (27) et de (28) viennent de l'indépendance des v.a. X_1, ..., X_n. La vraisemblance est donc une fonction du paramètre θ, alors que la densité est une fonction de la variable x.

Définition de la vraisemblance (cont.)

Pour des questions pratiques, on utilisera souvent la log-vraisemblance :

$$l_n(\theta; x) = \log(L_n(\theta; x_1, ..., x_n)) = \log(f_\theta(x_1, ..., x_n)) = \sum_{i=1}^{n} \log f_\theta(x_i). \tag{29}$$

Recherche d'estimateur

Intuition : Choisir comme estimateur la valeur du paramètre qui rend l'observation la plus probable.

Définition

Soit un n-échantillon $(X_1, ..., X_n)$. On appelle estimateur du maximum de vraisemblance la v.a. $\hat{\theta}(X_1, ..., X_n)$ telle que :

$$\hat{\theta}(x_1, ..., x_n) = argmax_{\theta \in \Theta} L_n(\theta; x_1, ..., x_n) = argmax_{\theta \in \Theta} l_n(\theta; x_1, ..., x_n). \tag{30}$$

Exemple 1

Supposons que $X_i \sim \mathcal{B}(p)$ avec $p \in [0;1]$ (loi de Bernouilli).

On cherche l'estimateur du maximum de vraisemblance pour le paramètre p.

On a, pour tout $i \in \{1, ..., n\}$:

$$f_p(x_i) = \mathbb{P}_p(X = x_i) = p^{x_i}(1-p)^{(1-x_i)} \tag{31}$$

La log-vraisemblance s'écrit :

$$\begin{aligned} l_n(p; x_1, ..., x_n) &= \textstyle\sum_{i=1}^n \log f_p(x_i) \\ &= \textstyle\sum_{i=1}^n \big(x_i \log p + (1-x_i)\log(1-p)\big) \\ &= (\textstyle\sum_{i=1}^n x_i)\log p + (n - (\textstyle\sum_{i=1}^n x_i))\log(1-p) \end{aligned} \tag{32}$$

On recherche ensuite p tel que :

$$\frac{\partial l_n(p; x_1, ..., x_n)}{\partial p} = 0, \tag{33}$$

et on trouve finalement :

$$\hat{p} = \frac{1}{n}\sum_{i=1}^n X_i. \tag{34}$$

Exemple 2

Supposons que $X_i \sim \mathcal{E}(\lambda)$ avec $\lambda > 0$ (loi exponentielle).

On cherche l'estimateur du maximum de vraisemblance pour le paramètre λ.

On a, pour tout $i \in \{1, ..., n\}$:

$$f_\lambda(x_i) = \lambda \exp^{-\lambda x_i} \tag{35}$$

La log-vraisemblance s'écrit :

$$\begin{aligned} &l_n(\lambda; x_1, ..., x_n) = \sum_{i=1}^n \log f_\lambda(x_i) \\ &= \sum_{i=1}^n \left(\log \lambda - \lambda x_i\right) \\ &= n \log \lambda - \lambda \sum_{i=1}^n x_i \end{aligned} \tag{36}$$

On recherche ensuite λ tel que :

$$\frac{\partial l_n(\lambda; x_1, ..., x_n)}{\partial \lambda} = 0, \tag{37}$$

et on trouve finalement :

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i}. \tag{38}$$

Propriétés

Définition

Un modèle dominé est dit homogène si le support de $f_\theta(.)$ ne dépend pas de θ.

Propriété

Dans un modèle dominé homogène, lorsque la vraie valeur de θ appartient dans l'espace des paramètres Θ, alors l'estimateur du Maximum de Vraisemblance est consistant.

Discussion

- Avantages
  - Méthode universelle
  - L'estimateur prend ses valeurs l'espace des paramètres Θ
  - Bonnes propriétés asymptotiques (cf. résultat de consistance)
- Limites (liées aux difficultés de rechercher un maximum)
  - Ni existence, ni unicité acquises en général
  - Instabilité numériques (lors de recherche numérique)

Content

Bibliographie

- Stephan Morgenthaler, *Introduction à la statistique*, PPUR, 2013.
- Jean-François Delmas, *Introduction aux probabilités et à la statistique*, Les Presses de l'Ensta
- Christophe Pouet, *Probabilités et Statistique*, Les Presses de l'Ecole Centrale de Marseille
- Maxime Ossonce, *Statistique*, Les Presses de l'Ensta